FSA Integration Partner

United States Department of Education

Federal Student Aid

Data Strategy Enterprise-Wide Data Framework Team 123.1.5 Quality Assurance Strategy and Implementation Plan

Task Order #123

Version 1.0

November 17, 2003

Executive Summary

The Federal Student Aid (FSA) program annually encompasses over eight million students, six thousand schools, four thousand lenders, thirty-six guaranty agencies, and twelve hundred internal FSA employees.[1] In 2002 the Department guaranteed and disbursed approximately $65 billion and managed a $267 billion loan portfolio for FSA programs.[2] To support these programs, FSA currently has seventeen major internal systems with dozens of external and internal interfaces through which they exchange substantial volumes of data. With the data being processed and stored in various systems, there is a high probability of data becoming out of sync, incorrect data being entered into a system, or data not being exchanged at all. Additionally, the environment in which FSA operates is constantly adapting in response to new legislation from Congress, new technologies such as distance learning, a more knowledgeable and sophisticated student population, and greater demands from its trading partners as competition between institutions increases.

These issues have contributed to data quality problems for FSA systems and business processes. The following sample from the Data Quality Mad Dog Report (deliverable 123.1.3) demonstrates the range of quality issues that FSA faces and the need for an enterprise wide solution.

- Mad Dog Issue 23: The FFEL data reported to NSLDS comes from multiple Guaranty Agencies at varying times (monthly/quarterly) with varying close dates. This makes the picture of FFEL data inconsistent and hard to analyze.
- Mad Dog Issue 24: There is currently no SSA or INS verification for PLUS borrowers' SSN and citizenship. This can result in the lending of Title IV aid to ineligible borrowers.
- Mad Dog Issue 20: GAPS and FMS receive school id status and eligibility data updates at different intervals. This can result in errors being returned from GAPS to multiple systems that send it updates (COD, FMS etc.). This impacts the schools ability to draw money in GAPS.
- Mad Dog Issue 10: Manual data entry can result in records being submitted to FSA systems with incorrect identifiers which results in the creation of an invalid student/recipient.
- Mad Dog Issue 6: It is possible for a FAA to make changes to an applicant's EFC, or other aid influencing data, due to extenuating circumstances that cannot be communicated on the FAFSA. These changes are made at the school and except for Pell Grant changes; they are not reported to FSA.

[1] Based on statistics from FSA's "Front to Back FSA on the Web" located at http://extranet.sfa.ed.gov/sfa_university/training/f2botw/business/players/operatingpartners.html.
[2] Based on statistics from "Statement of John P Higgins Jr. Inspector General Department of Education Before the House Budget Committee United States House of Representatives July 9,2003".

The Price of Data Quality

FSA systems act as a central repository for all Title IV aid and trading partner information in order to effectively manage the disbursement of aid to students. The efficiency of this organization rests upon the quality of the data contained within it. As demonstrated above, if the quality of the data degrades so does the financial health of the organization as students may not receive timely disbursements, schools may misreport program funding, and outside trading partners may take advantage of faulty business processes. Gartner mentions that creating an executive awareness of data quality issues alone can provide gains: "…providing executives with visibility into operation process metrics can reduce the cost of quality by as much as 30%"[3].

To address these potential deficiencies, this document proposes a Data Quality Assurance Strategy to be managed by a Quality Steering Committee (initially formed by the FSA Business Integration Group - BIG). This strategy outlines a methodology for the Business Owners to identify, track, and fix quality issues throughout the enterprise. By driving the methodology through the Business Owners, buy-in by FSA personnel is facilitated, quality issues that impact the business are more easily identified, and resistance to change is lowered. The Data Quality Assurance Methodology presents the tools which may be leveraged to discover new data quality problems and to ensure the integrity of the data through cross program analytics and audits.

Industry Trends and Quality Methodologies

Until the 1980's, Total Quality Management (TQM) was seen as the primary quality assurance methodology in the market place. TQM created a paradigm shift for management by shifting focus from business profits to the overall quality of how the business operates, which defines those profits. As this methodology spread to other industries (manufacturing and service industries), TQM evolved into business specific implementations such as Six Sigma, ISO 9000, and the Capability Maturity Model (CMM).

Traditionally, ISO 9000 and Six Sigma have been applied to the manufacturing business but recently have been expanded to include services based industries. The Six Sigma methodology is used to quantify defects and variation among delivered products and services. This leads to an in-depth root cause analysis to correct the source of the error in order to improve and maintain quality. ISO 9000 is a Quality Management System with multiple versions that serves as a general quality standard. This requires continuous improvement processes if an organization wants to fulfill its criteria. In contrast, CMM provides a means for an organization to measure its processes against a standard maturity level and understand what improvement steps must be taken to mature to the next level.

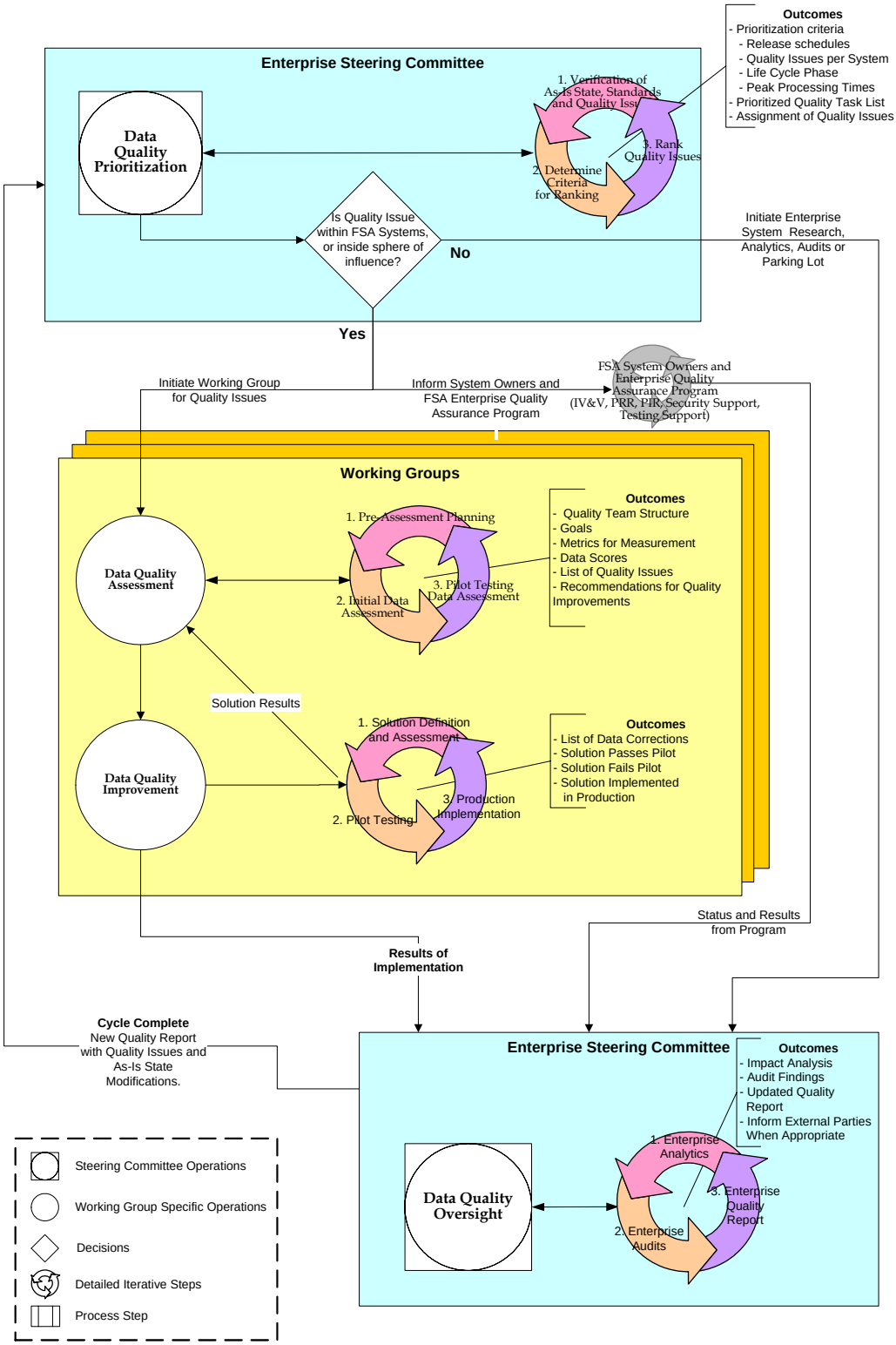
In contrast to these private industry methodologies, FSA has developed an Enterprise Quality Assurance program to address the quality issues within its enterprise and to obtain compliance

[3] © Copyright Gartner, Inc. Source: "Six Sigma Strategy for Quality and Customer Satisfaction." July 22, 2003.

with Office of Management and Budget (OMB) mandates. This program ensures quality of new development projects through its use of an Independent Validation and Verification process (based on the CMM methodology) to verify that all documentation and security audits are complete. The Production Readiness Review and Post Implementation Review programs further guarantee that application requirements set initially within the development life cycle are met fully before and after implementation. Security compliance to government standards for new and existing projects within FSA is handled by Security Support Services which performs independent risk assessments. In addition to these programs, the Enterprise Quality Assurance Program also offers Testing Support services for any new or existing application that requires additional testing assistance.

Creation of a FSA Enterprise Quality Assurance Strategy

Accenture has reviewed these methodologies with additional insight from data quality industry experts and determined that no one methodology is applicable to the unique way FSA does its business. Concepts and processes extracted from each methodology were leveraged and integrated with FSA's ongoing Quality Assurance efforts, including Testing Support, Independent Verification and Validation, Production Readiness Review, Post Implementation Review, Security Support, and other program/system specific efforts to formulate a single strategy for FSA. This strategy delivers the organizational and procedural recommendations which will allow FSA to meet its data quality assurance needs. As part of the recommendation, the high level components for establishing a permanent FSA Data Quality Steering Committee and provisional Working Groups are defined. The procedures and methodologies these groups perform are organized into four phases as depicted in the following diagram, Data Quality Methodology: Overview.



Data Quality Methodology: Overview

- **Data Quality Prioritization Phase:** The Steering Committee determines current data quality levels in order to establish a target state vision. As data quality issues are identified, the committee ranks the issues based on the impact to the enterprise. Assignment is then determined on whether the issue is solved through new development efforts (the issue is deployed to the FSA Enterprise Quality Assurance Program group), further research is required (the issue is deployed to the Oversight Phase for additional information), or a customized Working Group of system and Business Owners is formed for issue resolution.
- **Data Quality Assessment Phase**: If the Steering Committee deems that an issue specific Working Group must be formed for resolution, an Assessment Phase is initiated whereby cross system business and technical representatives are enlisted to analyze and solve the data quality issue at hand. In this phase, data is inspected, defects are measured, and sample data is determined for solution testing.
- **Data Quality Improvement Phase:** Based on data quality findings in the Assessment Phase, the Working Group determines solution criteria for ranking such as deployment speed, Return on Investment (ROI), and impact on the enterprise. Brainstorming by the Working Group (with assistance from additional system specific resources if needed) commences to propose different solutions to the data quality issue. The solutions are then measured to the criteria and the best solution is picked. After successful pilot testing of the solution, implementation in a production environment is scheduled.
- **Data Quality Oversight Phase:** Data is fed into the Oversight Phase from requests for additional information from the Prioritization Phase, status updates from the FSA Enterprise Quality Assurance Group on new development efforts, and Working Group results from individual solution implementations. The Steering Committee gathers this information and updates the Data Quality Report with issue results and status. In this phase, regularly scheduled maintenance and ad hoc requests are made for cross program analytics and audits to measure the quality of the enterprise's data and ensure the data meets agreed upon quality standards.

The Quality Assurance Methodology provides a means of addressing a number of the FSA Business Objectives. These objectives are part of the initial To-Be visioning effort of the Data Strategy project and help ensure that the methodology is crafted to the unique business model and technical needs of FSA. Validation of this Data Quality Methodology is achieved by simulating examples of data quality issues, obtained from the Data Quality Mad Dog Report (Deliverable 123.1.3), throughout the different phases of the Data Quality Methodology.

To successfully implement the Data Quality Strategy, a phased approach is recommended. These phases include Steering Committee formation, Tools analysis, and Development of a Quality Assurance Training Program. Each of these stages can begin in unison and are managed by the Data Quality Administrator with assistance from the BIG:

- **Steering Committee Preparation:** The Data Quality Administrator, with assistance from the BIG, will select Steering Committee members who have cross channel businesses knowledge and technical architecture skills. The Steering Committee's initial activities will revolve

around reviewing the As-Is state of FSA business and technologies, creating a high-level data quality standards and requirements for the Data Quality Report, and compiling a list of data quality issues for the Quality Report. Other activities will be supervision of the tools analysis team and training program work group. At the end of this plan, the Steering Committee will be ready to assume the mantle of managing the Quality Assurance Strategy for FSA.

- **Tools Analysis:** A team will be created with technical and business resources to evaluate the As-Is State of Data Quality tools within FSA. This team will then identify COTS tools in the market place in order to create a To-Be vision for an Enterprise Data Quality Tool. As a final phase, a GAP analysis and ranking will be done by the team to determine the best solution (e.g., a new customized solution or customized COTS tool) for an Enterprise Quality Tool. With a tool selected, future work will be required for implementation and management of the tool within FSA.
- **Training Program:** This team will be responsible for creation of a Quality Assurance Strategy training program that will inform FSA personnel the goals, objectives, and vision of the Quality Assurance Strategy. This work group will implement a solution life cycle of gathering high level requirements, designing a detailed methodology training application, and the build out with testing of the training program. A disbursement plan for the training program will be designed during the solution life cycle stage that will be implemented once the training program is finalized.

This document concludes with overall results achieved and a highlight of topics that have not been addressed by this strategy as future considerations for FSA.

Bigger Picture and Next Steps

It is important to re-iterate that the Data Quality Assurance Strategy assists FSA in achieving not only its business objectives of reducing data quality issues within the enterprise but will lay a foundation of repeatable processes that FSA can leverage, and collaborate with outside parties and agencies, to progress toward the high level organizational goals of removal of the Federal Student Aid program from the General Accounting Office's (GAO) high-risk list.

In the 2002 Department of Education report to the OMB, the financial impact that data and process quality issues has translated to over $400 million that has being identified as erroneous or improper payments.[4] The following is a breakdown of these issues:

- An audit of Pell Grant disbursements in 2001 found that over $336 million was improperly disbursed due to applicants falsely providing understated income or dependency status on their Free Application for Federal Student Aid (FASFA). This resulted in aid recipients receiving funds they were not qualified to receive.

[4] Based on statistics from "Statement of John P Higgins Jr. Inspector General Department of Education Before the House Budget Committee United States House of Representatives July 9, 2003".

- Out of the 36 Guaranty Agencies (GAs) working with FSA, audits of nine of these in 2002 and 2003 have identified approximately $164 million in unrecoverable aid.
- There are a rising number of financial aid consultants and businesses supplying false FAFSAs and tax returns containing false income amounts (enabling ineligible students to receive financial aid). This has resulted in over $4.8 million fraudulent claims that the government has to collect upon.
- Audits and internal analytics run by the Debt Collection System (DCS) has revealed one collection agency submitting false claims for payments of consolidated defaulted loans resulting in a $6.4 million settlement. In a second case, it was discovered that a collection agency used client trust funds for personal and operating expenses. This resulted in the collection agency owing more then $800,000 to 177 schools.[5]

Some of these issues are larger than data quality within the FSA organization, perhaps even larger than FSA programs themselves. However, despite this lofty task, the implementation of the Data Quality Assurance Strategy Methodology can help reduce these types of misuse of funds through proactive identification of data issues and anomalies throughout the enterprise; even if their resolution is larger than the reach of FSA and its actions. It is the goal of this strategy to impart tighter oversight of trading partner data and an overall reduction in data quality issues within FSA's data and systems. In comparison, the larger Data Strategy Target Vision provides the To-Be vision for FSA business, data, and technical architectures. Through the two pronged approach of adoption of the Data Quality Assurance Strategy and movement toward that Data Strategy Target Vision, FSA will move closer to its end goal of providing the right data to the right people at the right time.

[5] "Review of Collection Activities at Unger and Associates", ED-OIG/ACN: A06-90011; February 2000

Amendment History

DATE	SECTION/ PAGE	DESCRIPTION	REQUESTED BY	MADE BY
11/17/03	All	Document submitted for FSA-wide review.	N/A	T. Helwig

Table of Contents

1 INTRODUCTION 3

1.1 SCOPE 3
1.2 ASSUMPTIONS 3

2 INDUSTRY TRENDS AND PROVEN PRACTICES 3

2.1 TOTAL QUALITY MANAGEMENT 3
2.2 SIX SIGMA 3
2.2.1 Lean Six Sigma 3
2.3 CAPABILITY MATURITY MODEL 3
2.3.1 Capability Maturity Model – Integration 3
2.4 CURRENT FSA QUALITY INITIATIVES 3
2.4.1 The Independent Verification & Validation Process 3
2.4.2 The Production Readiness Review Process 3
2.4.3 The Post-Implementation Review Process 3
2.4.4 Security Support Services 3
2.4.5 Testing Support Services 3
2.4.6 Other Quality Initiatives at FSA 3
2.5 ISO 9000 3
2.6 METHODOLOGY COMPARISON 3
2.6.1 Cost of Poor Quality (Return on Investment) 3
2.6.2 Data Centric 3
2.6.3 Timeliness 3
2.6.4 Methodology Rankings 3

3 DATA QUALITY ASSURANCE STRATEGY 3

3.1 DATA QUALITY STEERING COMMITTEE 3
3.2 DATA QUALITY WORKING GROUP 3
3.3 DATA QUALITY METHODOLOGY 3
3.3.1 Overview 3
3.3.2 Data Quality Prioritization Phase 3
3.3.3 Data Quality Assessment Phase 3
3.3.4 Data Quality Improvement Phase 3
3.3.5 Data Quality Oversight Phase 3

4 RESULTS ACHIEVED 3

4.1 BUSINESS OBJECTIVES 3
4.2 DATA QUALITY EXAMPLES 3
4.2.1 Data Quality Prioritization Phase 3
4.2.2 Data Quality Assessment Phase 3
4.2.3 Data Quality Improvement Phase 3
4.2.4 Data Quality Assessment Phase 3
4.2.5 Data Quality Improvement Phase 3
4.2.6 Data Quality Oversight Phase 3

5 IMPLEMENTATION PLAN 3

5.1 GROUP 1: STEERING COMMITTEE PREPARATION 3
5.1.1 Phase 1 3
5.1.2 Phase 2 3
5.1.3 Phase 3 3

5.1.4 Phase 4 ..... *3*
5.1.5 Implementation *3*
5.2 GROUP 2: TOOLS SELECTION 3
5.2.1 Phase 1 ..... *3*
5.2.2 Phase 2 ..... *3*
5.2.3 Phase 3 ..... *3*
5.2.4 Phase 4 ..... *3*
5.2.5 Implementation *3*
5.3 GROUP 3: TRAINING PROGRAM DEVELOPMENT 3
5.3.1 Phase 1 ..... *3*
5.3.2 Phase 2 ..... *3*
5.3.3 Phase 3 ..... *3*
5.3.4 Phase 4 ..... *3*
5.3.5 Implementation *3*

6 FUTURE CONSIDERATIONS 3

APPENDIX A: ACRONYMS 3

APPENDIX B: BUSINESS OBJECTIVES – RAW 3

APPENDIX C: BUSINESS OBJECTIVES – DATA QUALITY SUBSET 3

APPENDIX D: DATA QUALITY TOOLS – DIAGRAMS 3

APPENDIX E: DATA QUALITY TOOLS – TABLES 3

APPENDIX F: DATA QUALITY TOOLS – COTS 3

APPENDIX G: DATA QUALITY METHODOLOGY OVERVIEW 3

List of Figures

Figure 1 . Data Quality Methodologies: Overview 3
Figure 2 . Data Quality Methodologies: TQM 3
Figure 3 . Data Quality Methodologies: Six Sigma 3
Figure 4 . Data Quality Methodologies: CMM 3
Figure 5 . Data Quality Methodologies: FSA Quality Assurance 3
Figure 6 . Data Quality Methodologies: ISO 9000 3
Figure 7 . Data Quality Methodologies: Model Comparison 3
Figure 8 . Data Quality Methodologies: Data Quality Assurance Strategy Methodology 3
Figure 9 . Quality Assurance Strategy Methodology: Overview 3
Figure 10 . Business Process Tools Mapping: Methodology Overview 3
Figure 11 . Data Quality Methodology: Prioritization Phase 3
Figure 12 . Business Process Tools Mapping: Prioritization Phase 3
Figure 13 . Prioritization Phase: Verification of As-Is and Quality Issues Stage 3
Figure 14 . Prioritization Phase: Determine Criteria for Ranking Stage 3
Figure 15 . Prioritization Phase: Rank Quality Issues Stage 3
Figure 16 . Data Quality Methodology: Assessment Phase 3
Figure 17 . Business Process Tools Mapping: Assessment Phase 3
Figure 18 . Assessment Phase: Pre-Assessment Planning (As-Is) Stage 3
Figure 19 . VOC Process Steps 3
Figure 20 . CTQ Tree 3
Figure 21 . Assessment Phase: Initial Data Assessment Stage 3
Figure 22 . Template of SIPOC Diagram 3
Figure 23 . Example A and B of SIPOC Diagrams 3
Figure 24 . Fishbone Diagram Template 3
Figure 25 . Fishbone Example Diagram 3
Figure 26 . Pareto Chart 3
Figure 27 . Assessment Phase: Pilot Testing Data Assessment Stage 3
Figure 28 . Data Quality Methodology: Improvement Phase 3
Figure 29 . Business Process Tools Mapping: Improvement Phase 3
Figure 30 . Improvement Phase: Solution Definition and Assessment Stage 3
Figure 31 . Improvement Phase: Pilot Testing Stage 3
Figure 32 . Improvement Phase: Production Implementation Stage 3
Figure 33 . Improvement Phase: Change Leadership Process 3
Figure 34 . Data Quality Methodology: Oversight Phase 3
Figure 35 . Business Process Tools Mapping: Oversight Phase 3
Figure 36 . Oversight Phase: Enterprise Analytics Stage 3
Figure 37 . Oversight Phase: Enterprise Audits Stage 3
Figure 38 . Oversight Phase: Enterprise Quality Report Stage 3
Figure 39 . Business Example Issue A: VOC Tool 3
Figure 40 . Business Example Issue A: CTQ Tree 1 3
Figure 41 . Business Example Issue B: VOC Tool 3
Figure 42 . Business Example Issue B: CTQ Tree 3
Figure 43 . Business Example Issue A: As-Is State SIPOC Diagram 3

Figure 44 . Business Example Issue A: Fishbone Diagram 3
Figure 45 . Business Example Issue A: Pareto Chart 3
Figure 46 . Business Example Issue B: SIPOC Diagram 3
Figure 47 . Business Example Issue A: Fishbone Diagram 3
Figure 48 . Business Example Issue A: Pareto Chart 3
Figure 49 . Data Quality Assurance Strategy Implementation Plan 3

List of Tables

Table 1 FSA FY2003 Detailed High Risk Plan Mapping 3
Table 2 Security Assessment Questionnaire Sample 3
Table 3 FMEA Analysis 3
Table 4 FMEA SSIM Example 3
Table 5 Data Quality Scoring Table 3
Table 6 Decision Matrix 3
Table 7 Improvement Phase: Control and Response Plan 3
Table 8 High Level Business Objectives – Sample 1 3
Table 9 High Level Business Objectives – Sample 2 3
Table 10 Business Example Issue A: FMEA Process 3
Table 11 Business Example Issue B: FMEA Process 3
Table 12 Business Example Issue C: FMEA Process 3
Table 13 Business Example Issue A: Sample Data Criteria 3
Table 14 Business Example Issue A: Sample Data Criteria 3
Table 15 Business Example Issue A: Solution Criteria Matrix 3
Table 16 Business Example Issue A: Solution Criteria Assessment Matrix 3
Table 17 Business Example Issue B: Sample Data Criteria 3
Table 18 Business Example Issue B: Solution Criteria Assessment Matrix 3
Table 19 Business Example Issue A: Pilot Testing Data Sample Results 3

1 Introduction

The goal of the FSA Enterprise Data Strategy is to define FSA enterprise data vision and strategy for how it will handle its enterprise data needs. Senior FSA leadership has distilled this strategy into the following program-wide business objectives:

- Reduce redundant data storage
- Improve customer service
- Increase accuracy of analytics
- Increase efficiency in data handling
- Reduce costs
- Remove FSA from the GAO high-risk list
- Maintain a clean audit

Further, senior FSA leadership has created an action plan to specifically define the tasks and scope of the activities needed to meet these objectives. Those tasks that pertain to data quality and are addressed in this document and have been mapped in the following table:

FSA'S FY 2003 DETAILED HIGH RISK PLAN (Current as of 02/04/2003)			TO 123 Data Strategy Deliverable Mapping		
FSA ID No.	Action Item		Del. Num	Name	Description
16.3	Determine current data quality and establish target state plan and quality assurance process.	→	123.1.3 123.1.5	Data Quality Mad Dog Report Quality Assurance Strategy and Implementation Plan	Documents current-state of data quality and potential areas for data quality improvement. This strategy will define a quality methodology for FSA and make recommendations for an on going data quality oversight team. Additionally, this deliverable will provide an Implementation Plan that will evaluate key data quality issues based on business needs, cost, value, and urgency.
16.3.1	Evaluate and identify data quality problems and identify correct source of data.	→	123.1.3 123.1.5	Data Quality Mad Dog Report Quality Assurance Strategy and Implementation Plan	Documents current-state of data quality and potential areas for data quality improvement. This strategy will define a quality methodology for FSA and make recommendations for an on going data quality oversight team. Additionally, this deliverable will provide an Implementation Plan that will evaluate key data quality issues based on business needs, cost, value, and urgency.
16.3.2	Develop a data cleansing plan based on cost, value, and urgency.	→	123.1.5	Quality Assurance Strategy and Implementation Plan	This strategy will define a quality methodology for FSA and make recommendations for an on going data quality oversight team. Additionally, this deliverable will provide an Implementation Plan that will evaluate key data quality issues based on business needs, cost, value, and urgency.
16.3.3	Develop and administer a data quality process which ensures the FSA projects implement enterprise data standards and quality standards.	→	123.1.5	Quality Assurance Strategy and Implementation Plan	This strategy will define a quality methodology for FSA and make recommendations for an on going data quality oversight team. Additionally, this deliverable will provide an Implementation Plan that will evaluate key data quality issues based on business needs, cost, value, and urgency. *Although the action item requires the development and administration of a data quality process, this deliverable only develops the process, it does not administer it.*
16.3.4	Establish a FSA data quality committee comprised of both business and technical subject matter experts to ensure that enterprise data standards are addressed within each project.	→	123.1.3 123.1.5	Data Quality Mad Dog Report Quality Assurance Strategy and Implementation Plan	Documents current-state of data quality and potential areas for data quality improvement. This strategy will define a quality methodology for FSA and make recommendations for an on going data quality oversight team. *It should be noted that this deliverable only defines the structure, tasks, and process for forming a committee. It does not create nor estabilsh the committe within FSA.*

Table 1 FSA FY2003 Detailed High Risk Plan Mapping

The purpose of this document is to provide a strategy for the improvement of data quality and the maintenance and refinement of data across the enterprise. This strategy establishes a Data Quality Methodology built from industry best practices, Business Owner input, external quality subject matter experts, and with the assistance of the FSA Enterprise Quality Assurance Program group. The methodology contains high-level components to establish a FSA Data Quality Steering Committee to determine current data quality levels and to establish a target state vision. As data quality issues are identified, the Steering Committee will organize Working Groups to create and execute detailed solutions.

This document also demonstrates FSA's emphasis on data quality through its business objectives as defined in the initial phases of the Data Strategy project and how the methodology meets those requirements through the use of two data quality issues from the Data Quality Mad Dog Report (Deliverable 123.1.3). By implementing the Data Strategy (as defined by its deliverables), FSA establishes an enterprise-wide data quality initiative through a consolidation of business processes, data storage, and new technology. In the final sections of this document, an Implementation Plan prioritizes tasks and addresses which phases of the methodology should be performed first. Future considerations, which will require additional research, are mentioned for those items not covered by the methodology.

1.1 Scope

This document provides a data quality framework that can be applied to the individual systems within FSA and the data feeds of those systems. It does not attempt to apply this strategy to any external trading partner system. Rather, when implemented, a data quality initiative would inform the trading partner of potential data issues. The external entity would then be responsible for fixing any data discrepancies.

As part of a data quality blueprint, the strategy demonstrates the process for creating data quality rules/standards within systems and the structure of the teams needed to perform this work. However, this blueprint does not determine the actual data quality rules nor does it include a tools analysis for commercial off the shelf (COTS) applications that may enhance the Data Quality Methodology and simplify a quality oversight process.

The Implementation Plan presented in this deliverable is an initial approach for the initiation of the process steps contained within the Data Quality Methodology and an outline for addressing the quality issues from the Data Quality Mad Dog Report (Deliverable 123.1.3). This information is input that the Steering Committee would consider when first performing the data quality Prioritization Phase of the methodology.

1.2 Assumptions

The following is a list of assumptions for Quality Assurance Strategy and Implementation Plan (Deliverable 123.1.5):

- This deliverable defines the processes for a Quality Assurance Plan. It does not initiate the plan nor does it form an FSA Data Quality Steering Committee to enforce data quality initiatives within FSA.
- The Quality Assurance Plan is for FSA systems only and the data that feeds into those systems. It does not include trading partners such as lenders, GAs, or state agencies. When a feed from an external system is found to have errors, documentation on the error will be gathered and the external system vendor will be responsible for supplying the correct data. If the external vendor would like to dispute the error, the Steering Committee will resolve all conflicts and, when applicable, form Working Groups to provide further examination.
- While the current FSA Enterprise Quality Assurance effort addresses issues of quality within projects, an overall data quality plan for FSA has not been developed. The Quality Assurance Strategy and Implementation Plan (Deliverable 123.1.5) attempts to address these gaps by communicating the necessary steps, at the enterprise level, to ensure data quality within FSA.
- The success of the proposed Data Quality Assurance Strategy is dependent on the appropriate commitment and buy-in from resources that are identified to participate in the Steering Committee and Working Groups.

2 Industry Trends and Proven Practices

With regard to enterprise-wide quality, the Department of Education has developed general Information Quality Guidelines as a result of Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001. This act directed the OMB to issue government-wide guidelines that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies." [6] These guidelines provide a standard to assess information quality by using factors such as utility, objectivity and integrity and are intended to ensure that information the Department disseminates is useful, accurate, reliable, unbiased and secure.

These guidelines, however, do not govern all information within the Department of Education, nor do they cover all information distributed by the Department. Specifically, documents intended only for inter-agency or intra-agency communications are exempt from these recommendations. The FSA Data Quality Assurance Strategy provides an opportunity to develop the methodology for controlling data quality issues internal to FSA that are not covered by the established guidelines. Currently, FSA does not have an enterprise-wide data quality initiative that systematically identifies business process issues, performs root cause analysis, resolves issues, and provides enterprise-wide analytics and/or audits to ensure the integrity of the data.

The following section of the document explores different key quality methodologies, their history, and their applicability to FSA. Through this exploration, the best components of each methodology are used to form an enterprise-wide Quality Assurance Strategy and Implementation Plan that focuses on business processes and the supporting data. Figure 1 Data Quality Methodologies: Overview outlines the relationship between each methodology discussed in FSA's new enterprise-wide Quality Assurance Strategy and Implementation Plan:

[6] http://www.ed.gov/policy/gen/guid/infoqualguide.html

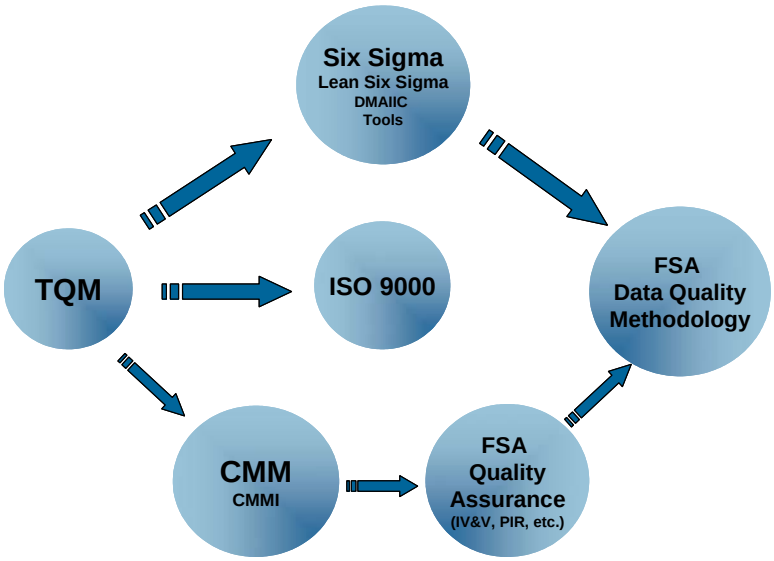


Figure 1 . Data Quality Methodologies: Overview

2.1 Total Quality Management

Total Quality Management (TQM) was developed by Dr. W. Edwards Deming who urged a new style of management that shifted the focus from profits to quality. Deming's innovative ideas centered on employees learning how to monitor, control, and continually improve their work processes and systems within the application of a scientific approach. With the collective attention of people on their work processes, products are produced that better meet customer expectations. TQM focuses products and services on improving *how* the work is done (the methods) instead of *what* is done (the results). Through this TQM philosophy, management is customer-oriented with all members striving to systematically manage the improvement of the organization through the ongoing participation of all employees in problem solving efforts.

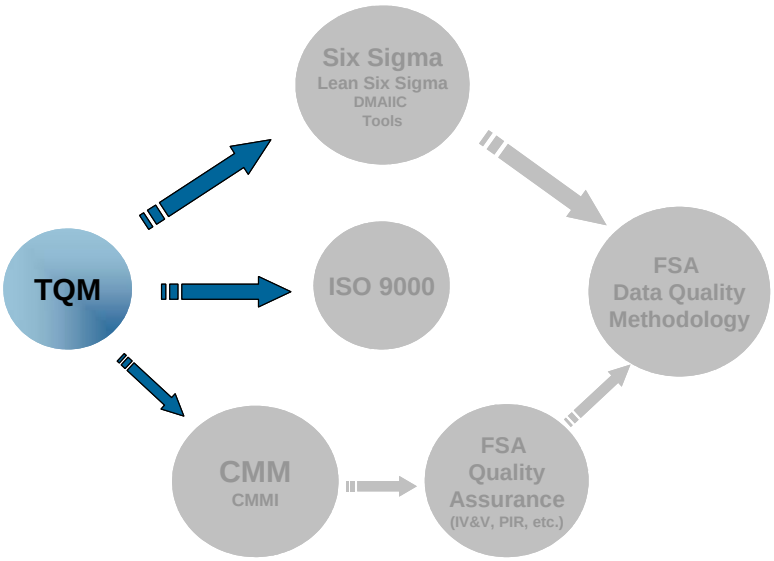


Figure 2 . Data Quality Methodologies: TQM

TQM incorporates the concepts of product quality, process control, quality assurance, and quality improvement. Pushing problem-solving and decision-making down from upper management to lower levels in the organization allows people who do the work to both measure quality and take corrective action in order to deliver a product or service that meets the needs of their customer.

With the implementation of this Quality Assurance Strategy and Implementation Plan, a Data Quality Prioritization Phase is designed to obtain subject matter expert input to identify quality issues within the enterprise. Much like the TQM methodology, the Prioritization Phase then pushes the problem solving of these issues down to the organization level where an Assessment Phase forms a Working Group composed of systems personnel and business analysts to address the issues.

The TQM methodology acts not only as the foundation of the new Quality Assurance Strategy but many quality methodologies are founded upon it. For example, a quality methodology known as Six Sigma evolved from TQM due to the necessity for constantly monitoring product quality to drive cross functional business processes. Six Sigma also distinguished itself from TQM by providing executive ownership, verifiable return on investment, a focus on statistical and quality training, and a more business results oriented methodology.

2.2 *Six Sigma*

Six Sigma is an enterprise methodology for businesses and corporations to help develop and deliver near-perfect products and services while efficiently controlling costs. First used by Mikel Harry and Bill Smith at Motorola in the mid-1980s, Six Sigma symbolizes 3.4 mistakes per

million opportunities with an "opportunity" defined as a chance for not meeting the required specifications. [7] "Sigma" is a statistical term that measures how far a given process deviates from perfection. The central idea behind Six Sigma is that if you can measure how many "defects" you have in a process, you can systematically figure out how to eliminate them and get as close to "zero defects" as possible.[8] According to Gartner, most service enterprises currently function at a two or three sigma level; this roughly translates to poor quality customer experiences occurring between 7 and 31 times for every 100 opportunities.[9]

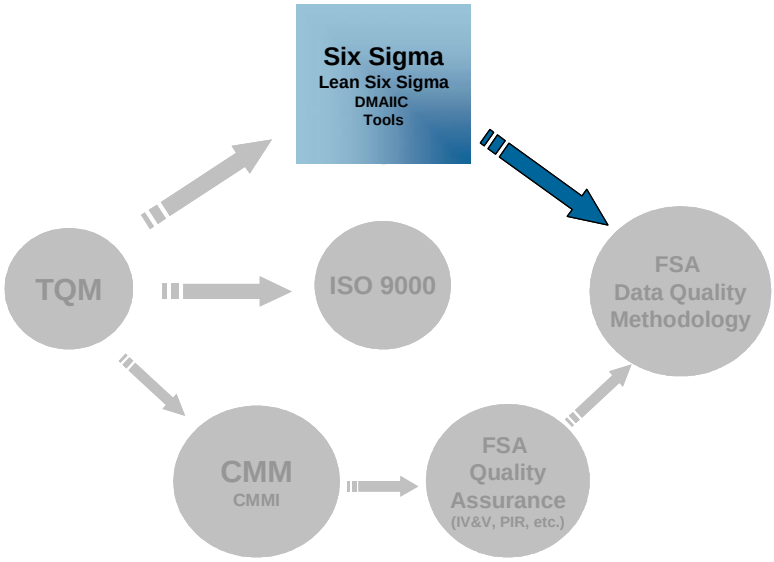


Figure 3 . Data Quality Methodologies: Six Sigma

Motorola originally created the Sigma process to eliminate assembly-line defects. However as the exceptional return on investments from the quality initiatives were made public, other major corporations soon adopted the methodology. For example:

- General Electric incorporated Six Sigma into its manufacturing, service, and financial enterprises.
- J.P. Morgan Chase applied Six Sigma to enhance its customer service and created efficient business operations.

[7] http://www.usatoday.com/money/companies/management/2002-10-30-sixsigma_x.htm by Del Jones, USA TODAY Posted 10/30/2002 4:12 PM
[8] A "Defect" is interpreted as failure of adhering to customer needs or business process and not to a technical problem.
[9] © Copyright Gartner, Inc. Source: "Six Sigma Strategy for Quality and Customer Satisfaction." July 22, 2003.

- Dow Chemicals used Six Sigma to determine the relative priority of user needs to assist in defining services levels with business value.[10]

Six Sigma is not only private sector focused. Recently, the Department of Homeland Security considered utilizing Six Sigma in their continuing efforts to identify terrorist threats. Since the government must sift through large quantities of data looking for terrorist related information (emails, intercepted phone calls, etc), it is crucial to identify and discard bad information. Six Sigma helps translate this problem into a 99.9% chance that all decision points made in a day are correct, which could result in making the USA [theoretically] 1,800 times safer then it is today.[11]

Six Sigma stresses designing a process that meets customer needs and business process capability. Specifically, it is a systematic method to improve business processes through the use of various tools in six critical phases, collectively known as the DMAIIC Method (Define, Measure, Analyze, Improve, Implement, and Control). The definitions for these phases are as follows:

- **Define**: Define the customer needs and problem, then identify where the current process falls short. In this step, a high-level process map, or As-Is diagram is created to identify the current state of the process and where defects occur. This phase includes developing a business case to address the quality issue and a communication plan for stakeholders. This step also includes laying the groundwork for how a team will work together to prioritize where focus should be.
- **Measure**: Metrics must be established and measured for the given process in order to quantify the extent and/or severity of defects. Current process performance is measured in terms of sigma (variation in products and percentage of defects) or other useful metrics. It is also important to measure customer requirements and issues, which involves gathering Voice of Client (VOC), Voice of Employee (VOE) and Voice of Business (VOB) data. From this data, issues that are Critical to Quality (CTQs) can be identified.
- **Analyze**: This phase is vital to understanding the key factors that drive process performance. During this phase, steps need to be taken to understand the current process performance and use problem solving tools to drive down to the root cause level. A successful Six Sigma effort is contingent upon understanding the fundamental causes of quality variation, rather then the expressions of it.
- **Improve**: This phase focuses on creating a new design and implementation plan with broad buy-in from all stakeholders. Activities in this phase range from brainstorming sessions with subject matter experts to mapping business decisions in a matrix for prioritization.
- **Implement**: Involves executing the process improvements that have been developed after any risk assessment and through possible test piloting. An implementation plan is used to

[10] © Copyright Gartner, Inc. Source: "Six Sigma Strategy for Quality and Customer Satisfaction." July 22, 2003.
[11] http://www.usatoday.com/money/companies/management/2002-10-30-sixsigma_x.htm by Del Jones, USA TODAY Posted 10/30/2002 4:12 PM

establish and communicate the solution and ensure that all involved have a common understanding of how the plan will be implemented.

- **Control**: Serves to institutionalize the process changes and ensure that the gains are sustainable over time. This is done by implementing continuous measurement systems while documenting and identifying continuous improvement opportunities.

2.2.1 Lean Six Sigma

Similar in methodology to the Six Sigma approach, Lean Six Sigma (LSS) focuses on eliminating non value-added and unnecessary tasks, and is bottom-line focused. The original ideas behind this methodology are based upon concepts that grew from Toyota's production system and seek the elimination of waste and the synchronization of every aspect of a company's operation.[12] LSS can be applied to manufacturing, software, operational, transactional, and service processes. It is widely used among industry leaders such as General Electric Co., Sony Corporation, J.P. Morgan Chase & Co, and Citibank. It is also used among many government contractors such as Lockheed Martin, Raytheon, Boeing, TRW and Northrop Grumman.

Six Sigma was developed to drive all defects in a process to zero. However, there is no direct initiative in the original methodology that addresses the elimination of unnecessary tasks. LSS follows a phased methodology similar to DMAIIC with the exception that the Define Phase is modified and splits out a new Vision stage:

- **Vision:** This stage focuses in on the nonvalue-added processes of an organization. Critical-to-customer factors, teams and key stakeholders are identified, and business process impacts are analyzed in order to gain insight into which processes should be improved and which should be re-examined.
- **Define:** This stage is now focused on highlighting those processes that, while understood by experience personnel, are not actually documented. The theory is that by asking people to sit down and document a procedure it may inherently improve it.

The Data Strategy Initiative within FSA is focused on the quality assurance of enterprise-wide data. Six Sigma is an enterprise-wide business initiative and is not solely data specific. While it does address some of the components of a data quality assurance strategy, the cost, complexity, and private sector focus of the Six Sigma's DMAIIC process does not quite fit FSA. Six Sigma is ideal for identifying defects within a manufacturing process and only recently has it been applied to the services industry. In order for it to be highly effective in reducing defects within data and interface transfers, the Six Sigma process needs to be molded to fit the FSA Business model and technical architectures. FSA's business model is unique in that it deals with the complexities of disbursing and managing student aid bound by governmental laws and regulations. Additionally, FSA has many business partners ranging from private institutions (lenders, schools, third party services) to government entities (Department of Defense, Department of Justice, Internal Revenue Service, etc), which, in some cases, would require a

[12] SAE Internal Web Site: http://www.sae.org/servlets/index

Congressional mandate to modify technical or business processes. Therefore, a thorough Six Sigma analysis will not be cost or time effective for FSA. Instead, FSA should leverage process improvement tools utilized in Six Sigma's DMAIIC methodology to evaluate and identify root causes of data errors and implement a unique Quality Assurance strategy customized to the government's structure.

2.3 Capability Maturity Model

The Capability Maturity Model (CMM) was created in the early 1990's by the federally funded Software Engineering Institute at Carnegie Mellon University, and serves as a framework for improving application development and maintenance process. This effort was initiated in response to the request of the U.S. Government to provide a method for assessing the capability of its contractors. CMM was mandated by the Department of Defense for guidance in selecting process improvement strategies by determining current process maturity and identifying critical issues to software quality and process improvement. CMM was written exclusively for software developers and rarely involves the work of other process groups, e.g., quality management and product/customer focus.

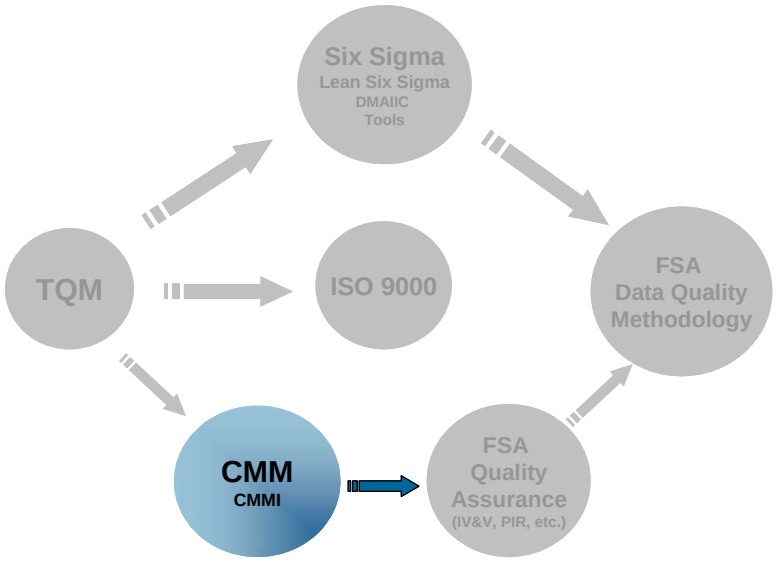


Figure 4 . Data Quality Methodologies: CMM

CMM utilizes a maturity scale to assess an organizations performance and to define objective criteria to be used for progressing to higher maturity levels. The model also provides the relevant process and process outcomes associated with each level. The maturity process has been designed so that as organizations move up the five levels, they will experience an increase

in predictability, effectiveness, and control of the software processes. The five levels of maturity are:

- Level 1: Initial – The organization is not capable of consistently meeting with commitments and may even seem occasionally chaotic. Weak project management exists across the enterprise, where individual developers have individualized methodologies for performing assignments.
- Level 2: Repeatable – Applications and projects are delivered on schedule without any additional effort, such as overtime, needed. The basic project management process is established to track cost, schedule, and functionality. Thus an environment is established to facilitate the repetition of successful practices with the planning of achievable commitments with baseline requirements being met.
- Level 3: Defined – After establishing successful practices, best practices are identified and integrated into a standardized organization-wide process.
- Level 4: Managed – After establishing a common process, statistical capability baselines are developed and detailed measures of the process and product quality are collected. Quantitative management of the process benefits project teams and management with increased empowerment and predictability of results.
- Level 5: Optimizing – At the highest level, there is an establishment of an infrastructure to support continuous change management. With management established, there is now a continuous evaluation of process capability in order to identify areas of improvement. Evaluations occur via feedback from the process and from piloting innovative ideas and technologies.

CMM essentially identifies the current status and maturity level of an organization and provides a vision of where an improved state may be. It is thus a descriptive and not a prescriptive model. According to Gartner, "CMM provides a measure for the organization's state and defines a reasonable strategy for process improvement." CMM does not, however, provide explicit methodology of achieving the vision state. FSA could leverage the maturity state definitions as verification of quality levels within FSA new development efforts and utilize components of Six Sigma as a roadmap for data quality improvement.

2.3.1 Capability Maturity Model – Integration

In late 2001, the Capability Maturity Model – Integration (CMMI) was released which integrated all associated CMM models that may have been formed since the original CMM implementation. The new CMMI essentially includes basic structural changes to CMM, making the model more effective and more broadly applicable to companies that might have found the original models too rigid.

CMMI is used to help make business management decisions about how software is developed. For example, there is Project Management functionality in CMMI not provided by CMM. The traditional CMM model is essentially a software development model meant for programmers to code, test, deploy, and build on their software; CMMI serves as a business management model

used to gain insight into and control over the development, in order to predict and adjust project activities towards success.

Today, CMM/CMMI is the de facto standards for software management throughout the federal government and is internationally recognized as very powerful business tools and competitive differentiators. CMM/CMMI has been leveraged within the FSA Enterprise Quality Assurance plan which implements an Independent Validation & Verification Phase (IV&V) for new development projects. IV&V evaluates new software development efforts to a series of management and quality levels (see Section 2.4.1 - The Independent Verification & Validation Process description below). It is recommended that FSA continue to utilize the FSA Quality Assurance plan for all new development projects within FSA.

2.4 Current FSA Quality Initiatives

As mandated by OMB, FSA has created an Enterprise Quality Assurance Program comprised of an Independent Verification and Validation process, Production Readiness Review, Post-Implementation Review, Security Support Services, and Testing Support Services. These processes help accomplish adherence to federal guidelines for new systems development and security assessments while providing early problem detection. Additionally, by increasing the visibility into new development projects, FSA reduces risks and improves system maintainability, reliability and integrity.

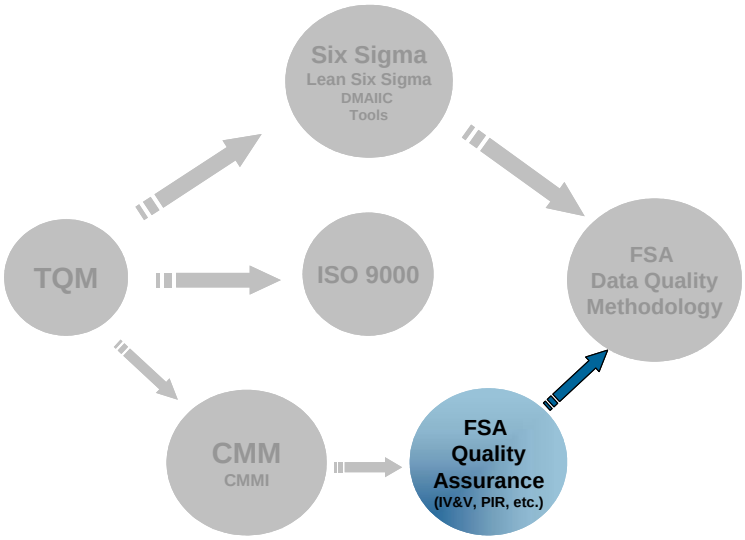


Figure 5 . Data Quality Methodologies: FSA Quality Assurance

Of the existing Enterprise Quality Assurance Program initiatives, the IV&V process is mainly applied to new development efforts and can therefore be utilized at the discretion of the system owners or Steering Committee when a new development effort or new release is needed for data quality issue resolution. In contrast, the PRR is a required process for all new applications before they are released in to production. The PIR is at the discretion of the Chief Information Officer (CIO) management and is used to ensure requirements have been satisfied after production release while the Security and Testing Support Services can be utilized for any system update or maintenance activity. For a complete review of these processes, please refer to the FSA Enterprise Quality Assurance Program documentation.

2.4.1 The Independent Verification & Validation Process

For the Independent Verification and Validation (IV&V) process, a systems engineering approach was taken to review the quality of new software development initiatives within FSA. When initiated, an IV&V team is formed independent of the development effort to generate a project plan that will identify key activities and deliverables for the development effort. This ensures proper application creation while enhancing software quality and reliability through the use of early error detection controls and quick assessments of proposed changes and consequences. The overall result is a reduction in risks and costs with an increase in customer satisfaction.

IV&V is heavily based on CMM and uses it as a standard for assessing and recommending improvements for the developer's software processes. As discussed above, CMM can be utilized to effectively identify key practices that are required to increase the maturity of the developer's software processes. As such, IV&V provides the following benefits to an organization's business processes:

- By creating baselines, change controls, and issue tracking for IV&V usage, a "watchdog effect" is created that fosters better software development, as developers are more careful and conscientious in their work with code reviews and adherence to development standards.
- Increased accuracy of system documentation for improved maintainability and customization when needed.
- Less cost as a greater number of errors and issues are resolved in the development phase rather then further down the system life cycle (testing or even production support).
- Greater business user participation as the Solution Acquisition Plan (SAP) and the Statement of Objectives (SOO) are verified to correctly map to FSA needs.

IV&V also has similarities to the Six Sigma methodology as the review, analysis, and testing techniques utilized to determine system compliance to functional and performance capabilities are comparable to the Six Sigma Critical to Quality (CTQ) phase. Under this process, CTQ key items are identified as those attributes that serve the user's need for a product that are fully capable of meeting a set of objectives. CTQ is then leveraged by the Six Sigma methodology as a baseline for a particular process. Thus the CTQ process identifies and verifies user attributes

much the same way that IV&V ensure requirements are met in the design / build phases of an applications development.

IV&V is utilized for new applications or during major version or point releases within FSA. In regards to the new Quality Assurance Strategy, the IV&V process will be utilized by a Steering Committee for those quality issues that will be fixed or addressed by the development of a new application. For instance, if there was a quality issue regarding the Post Secondary Education Participation System (PEPS), eZ-Audit, or eCB systems, the issue might be fixed by incorporating a new requirement into the new Trading Partner Management System (TPM) application. While some issues may be addressed by new application development within FSA, it is expected that the majority of data quality issues will require cross system Working Groups to be formed. It will be up to the Steering Committee's discretion to decide which issues required IV&V assistance and which do not.

2.4.2 The Production Readiness Review Process

For all applications approaching the deployment phase of their development life cycle or for major system releases, the Production Readiness Review (PRR) process verifies that the tasks, activities, and operations support for the release in production are prepared. The PRR process is divided into Preparation, Collaboration, and Presentation phases:

- **Preparation:** Internal documentation is prepared and reviewed while the application status is reviewed within the project team. The team also performs preparation work towards the collaboration and presentation phase, by ensuring proper documentation is available.
- **Collaboration:** The project team must review, with all project stakeholders and participants, PRR documentation. This ensures that the application and any associated functions and processes surrounding its support (such as Help Desk support) are functioning properly or available.
- **Presentation:** An executive overview slide presentation of the Production Readiness Review is created. This gives an overview of the development approach, testing summary, collaboration summary, quality assurance methodology, and risk assessment. Upper management approval and signoff must occur to complete the PRR process.

A PRR checklist is prepared, reviewed, and must be satisfied by the project in question. The following example is an excerpt from a PRR Checklist to demonstrate some of the detailed documents and steps a project must go through before an authorization is given:

- **Client:** Escalation procedures for client notification of problems are properly defined and change control procedures are in place.
- **General:** Documentation has been completed for the task order, scope, approach, and business drivers.
- **Service Operations/Recovery/Contingency:** The application hosting center has properly documented the responsibilities matrix, escalation list, SLA, installation guidelines, and

troubleshooting procedures. As part of this stage, it is important to understand the number of application users, ramp up plans, and help desk readiness.

- **Configuration Management/Source Control:** A source code library is in place in conjunction with version control procedures and a configuration management plan.
- **Technical Architecture:** All environments have detailed specifications (Development, Test, and Production). The development standards are properly defined in the context of the overall development life cycle.

The benefit from implementing the PRR processes is to make the production implementation more efficient through a reduction in the duplication of efforts. PRR also ensures upper management is aware of any open issues and that a mitigation strategy is in place. The introduction of controls such as a checklist of tasks for the hosting facility, security requirements, and development teams lowers the risks and any necessary mitigation plans. This is similar to the outcomes from the IV&V processes. The PRR guidelines are updated quarterly based on FSA needs and are currently conducted for all project and major system releases.

2.4.3 The Post-Implementation Review Process

A Post-Implementation Review (PIR) process is conducted in order to ensure FSA development projects are meeting their original investment goals and objectives after their implementation. This process was implemented by FSA/CIO in July 2002 to establish repeatable processes, and to make certain new applications are functioning effectively and satisfying customer needs; all of which are required for an application to be compliant with the OMB, the Clinger-Cohen Act of 1996, and IT Investment Management and Planning legislation. A PIR is usually conducted three to twelve months after implementation and consists of the following phases:

- Assess mission needs and determine project goals
- Collect and analyze data
- Provide major findings and issues
- Provide feedback & incorporate lessons learned.

PIR provides a documented record of a project's processes and allows for actual versus planned costs, returns, and risk analysis. It also provides insight into system completeness and functionality so that any areas in need of correction may be identified and addressed. Theoretically, FSA can use PIR effectively when implementing new system processes and testing their functionality with regards to data quality.

2.4.4 Security Support Services

As part of Federal Government requirements, all government agencies are instructed to protect information from unauthorized access, disclosure, modification, manipulation, or destruction.[13] The main components of achieving this objective are information classification, security risk assessment, residual risk determination, and mitigation strategy. This section provides an overview of Security Support Services as these components contain large amounts of security verification information and documentation. For a more detailed examination, please refer to the Independent Verification & Validation Handbook Section 4 - Security Assessment Standards and Procedures.

Information classification provides a conceptual security framework for data by defining each piece of information as either public, proprietary, sensitive or classified. At the least secure level, public information is data that has been authorized for public use, whereas proprietary information is data that is not classified, public, or sensitive (for example, a contractor's in-house applications and information are considered proprietary). At a higher level of security, sensitive information is data that has been mandated by federal law requirements to have special handling or restrictive access criteria. This data usually resides in a company's financial department and contains information such as SSN, income, etc. At the highest security level, data that is designated as the National Classification Information by the Presidential Executive Order 12598 is classified according to the criteria set in that executive order.

A security risk assessment is a comprehensive detailed analysis of a system's processes, documentation, risks and vulnerabilities. One of the major components of a risk assessment is the interviewing process. A cross sectional sample of system and Business Owners (database administrators, configuration manager, operations staff, business analysts, etc.) review a detailed questionnaire compiled from a National Institute of Standards and Technology (NIST) security assessment questionnaire, security best practices, and experience from FSA security reviews. The following is an excerpt from this questionnaire (for a complete questionnaire please refer to the Independent Verification & Validation Handbook Appendix F: Security Assessment Questionnaire):

[13] Department of Education Federal Student Aid Chief Information Office Electronic Commerce Applications Development. QUALITY ASSURANCE HANDBOOK. Section4, Security Assessment Standards & Procedures March 4, 2002

WBS	Specific Control Objectives and Techniques	Policy	Procedures	Implemented	Tested	Integrated	COMMENT	Notes
Management Controls								
	Risk Management *OMB Circular A-130, III*							
1.1	**Critical Element:** Is risk periodically							
1.1.1	Is the current system configuration do *NIST SP 800-18*							
1.1.2	Are risk assessments performed and documented on a regular basis or whenever the system, facilities, or other conditions change? *FISCAM SP-1*							
1.1.3	Has data sensitivity and integrity of th *FISCAM SP-1*							
1.1.4	Have threat sources, both natural and *FISCAM SP-1*							
1.1.5	Has a list of known system vulnerabilities, system flaws, or weaknesses that could be exploited by the threat sources been developed and maintained current? *NIST SP 800-30*							
1.1.6	Has an analysis been conducted that determines whether the security requirements in place adequately mitigate vulnerabilities? *NIST SP 800-30*							

Table 2 Security Assessment Questionnaire Sample

Other components of the risk assessment include:

- Documentation review: Each system must have a system security plan, system security authorization agreement, disaster recovery plan, and other pieces of information as defined by FSA's Independent Verification & Validation Handbook section 4 - Security Assessment Standards and Procedures.
- Standardized criteria are defined: Evaluation measurements must be established for comparison in the security assessment. Items include probability and severity of security vulnerabilities, potential and residual risks, mitigating countermeasures, information categorization, level of proof calculations for objective evidence.
- Corrective Action Plans: This component tracks the performance of the security assessment and mitigation factors through the use of Gantt charts and diagrams.

After the security risk for a system has been determined and measures have been applied to reduce or eliminate the risk, what is left is the residual risk. By managing the system to the residual risks, mitigation strategies can be designed to further lower these risks in order to maintain a highly secure system. Residual risk is a calculated field composed of the proposed countermeasures, mitigation strategies, and input from a security engineer's experience. There are three rating levels.

- **HIGH:** severe impact to US interests with current countermeasures inadequate.
- **MEDIUM:** serious impact to US interests with countermeasures that are effective but not enough to eliminate the risk.
- **LOW:** moderate impact to US interests with counter measures effective in lowering the risk.

Through the use of security services risk management for a system is effectively achieved through the assessment of risks, development and implementation of mitigation strategies, and lowering of risks to acceptable thresholds. Similar to PIR and PRR, Security Support Services can be utilized within FSA Data Quality initiative for those issues that involve security violations. The Steering Committee would evaluate quality issues, determine which are security related, and then inform the Enterprise Quality Assurance group that security services are required to help reduce or eliminate the risks created by the quality issue.

2.4.5 Testing Support Services

Testing Support Services provides independent testing support and assessment to project teams and management, which would normally be outside of IV&V procedures. This process provides FSA project teams with access to additional testing support and resources that may be outside the scope of the development Task Order. In addition, it provides testing standards for integration project teams. Testing Support Services also improves product quality and reduces resource burdens on the stakeholder team. Independent testing support is also provided to Web application teams as they can use a Rational Site Check 2000 tool for recurring application testing support at no additional cost.

For the new Quality Assurance Strategy, the Steering Committee could leverage Testing Support Services to support the efforts of the Working Groups as independent testers. Additionally, as new quality issues are discovered through the use of the methodology, Testing Support Services could be considered for independent verification of results.

2.4.6 Other Quality Initiatives at FSA

Within FSA there are a variety of other quality initiatives such as the Quality Assurance for Schools program and the Effective Practice Database. While these programs are not within the scope of this strategy, they do have quality components that could further enrich this new Quality Assurance Strategy. Future examination should be done into these tools to determine what, if any components could be tied into the methodology. While these other quality initiatives demonstrate the breadth and depth of FSA quality programs, they also demonstrate a lack of enterprise-wide data quality standards and highlight the need for a comprehensive assurance program to tie them all together.

2.5 ISO 9000

The history behind the ISO 9000 standards begins with the International Standardizing Associations (ISA), which was set up in 1926 to act as a governance board for mechanical engineers. ISA's activities came to an end in 1942 and in 1946, delegates from 25 countries met

in London and decided to create a new international organization, the International Organization for Standards, of which the object would be "to facilitate the international coordination and unification of industrial standards".[14] The new organization, ISO, officially began operations on February 23rd, 1947 and coined "ISO" for the standards based on the Greek word "isos" which translates to "equal" in English.

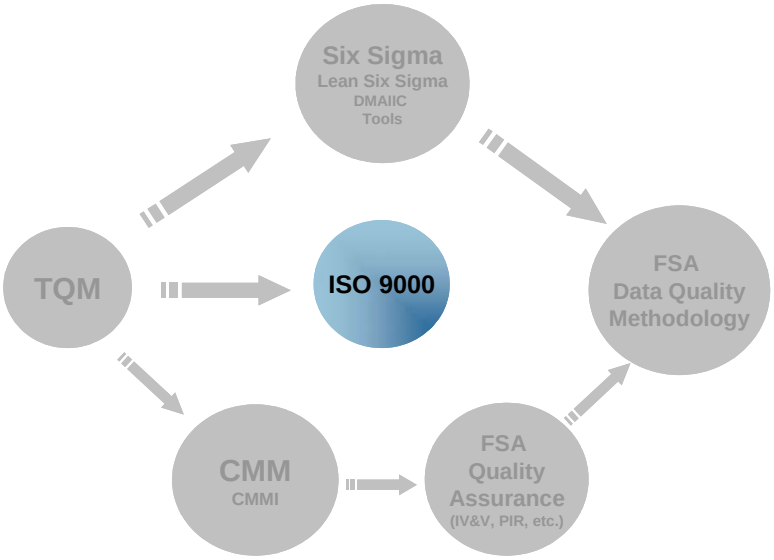


Figure 6 . Data Quality Methodologies: ISO 9000

Today ISO administers over 11,000 standards covering 97 categories. It has been adopted by 60 countries, including the United States, Canada, Japan, and the members of the European Union. ISO 9000, which was first published in 1987, is a series of quality standards that organizations use to audit their systems in order to provide external organizations and their clients an assurance of effective system operations. The ISO 9000 family of models include ISO 9001, ISO 9002, and ISO 9003. The implementation of each model depends on whether an organization is based on design control, process control, or servicing and testing (an organization can be audited on all three as well).

The new ISO 9000:2000 standards are based on eight quality management principles. ISO chose these principles because they can be used to improve organizational performance and achieve success. These principles are listed below:

- Organizations must focus on understanding customer needs, meeting customer requirements, and exceeding customer expectations.

[14] International Organization for Standardization web site at http://www.iso.ch/iso/en/aboutiso/introduction/index.html#two

- Organizations must rely on strong leadership. Leaders need to establish a unity of purpose, set direction, and create an environment that encourages people to achieve the organization's objectives.
- Organizations must rely on their people and thus encourage the involvement of people at all levels while helping to develop their abilities.
- Organizations must use a process approach to manage activities and related resources.
- Organizations must use a system approach to identify and manage interrelated processes.
- Organizations must make a permanent commitment to continually improve overall performance.
- Organizations must base decisions on analysis of factual information and data.
- Organizations depend on suppliers to create value; therefore they must maintain a mutually beneficial relationship with them.

Critics of ISO 9000 view it as too general and too centered on manufacturing processes. Additionally, there are no universal standards for ISO 9000 auditors and no data quality levels defined as there are with CMM. Instead "you must document the processes that you follow, and you must follow those processes."[15] Thus, even bad processes may be documented despite the fact that the organization is ISO 9000 certified. This translates into lack of control around quantitative measurements not meeting base line requirements, failure to consistently establish performance standards based on peer group input, and lack of documentation on best-practice performance levels. Because ISO 9000 does not support any process improvement methodology or even guarantee acceptable performance levels, it is not the right methodology to ensure improved data quality within FSA.

2.6 Methodology Comparison

A well defined FSA Data Quality Methodology is vital to improving and maintaining overall data quality. Data quality is a facet of overall quality across the enterprise, and refers to identifying and solving one-time or reoccurring issues that are adversely affecting stored and transferred data. From the methodologies outlined above, it becomes apparent that each utilizes different tools, standards, and processes to analyze the cost of quality for an enterprise and how each issue affects the overall return on investment. Extracting the most relevant pieces of industry and government practices that are data centric, a data quality assurance methodology can be built and integrated into a new FSA Quality Assurance Strategy.

2.6.1 Cost of Poor Quality (Return on Investment)

A fundamental driver of any data quality methodology is its cost of poor quality. The term "Cost of Quality" refers to the economic impact and lost revenue associated with not conforming to customer requirements or continuing a process with unmitigated service problems or process defects. For instance, the Data Quality Mad Dog Report (Deliverable 123.1.3) issue 20 details that Grant Administrative and Payment System (GAPS) and Financial

[15] © Copyright Ovum, Inc. Source: "The Software Quality Maze" April 2003

Management System (FMS) receive school ID status and eligibility data at different intervals. The Cost of Quality here is that schools are not able to draw down money from GAPS causing a series of manual processes to be invoked. This costs the Department customer satisfaction, awarding aid, and schools revenue as multiple FSA staff members along with school representatives must coordinate the manual draw down procedures.

When implementing a data quality strategy, the Cost of Quality is important to keep in mind. In private industry, Dennis Sester, Senior VP & Motorola Director of Quality has commented on this Cost of Quality and ROI with regards to the Six Sigma methodology: "In the early phases, the Cost of Quality can be as high as 20% of revenues... in order to make major gains, the product, process or system usually must be designed. So, in many cases, several iterations of product or system upgrading may be required." [16]

The Cost of Quality process is an effective tool to measure the impact a quality issue can have on an enterprise. In comparison, neither ISO 9000, CMM, nor any of the FSA Quality Initiatives have a quantitative measurement process on how to determine the impact of a data quality issue on the enterprise.

2.6.2 Data Centric

Implementing a Quality Strategy that is specific to an organization's data, how it is stored, shared, transferred, and interfaced is a unique process that is not addressed by a single methodology. Six Sigma utilizes statistical calculations and analysis when assessing variances within a process. The overall goal of any Six Sigma program is to reduce defects resulting from a process. CMM, in contrast does not specifically utilize metrics relying on conformance to standards. It focuses on improving the quality of a product via improving the quality of the process used to build it. FSA's Quality Assurance program has components that manage new application conformance to a set of quality standards while ensuring the proper security levels. None of these methodologies specifically examines the data of an enterprise, the standards that exist, and different tools to discover and fix quality issues. Additionally, these methodologies do not operate in the overall context of the future vision for the enterprise. The new Quality Assurance Strategy along with the Data Framework Specification (Deliverable 123.1.4) and Technology Vision and Strategic Plan (Deliverable 123.1.12) provide a comprehensive plan for FSA to reach stated business objectives through a re-organization of data with the enablement of new technologies.

2.6.3 Timeliness

Six Sigma projects are considered for short term, three to six months, and applied to optimize iterative processes within manufacturing and services. CMM is applied across the long term for an enterprise to improve the success of product development or maintenance projects by implementing processes that have created proven results in the past. For ISO 9000 initiatives, timeliness is not as applicable since the methodology is a set of standards and not a process for

[16] http://www2.cio.com/ask/expert.cfm?EXPERT=26

improving data quality. Timeliness depends on the amount of effort it takes to map enterprise processes to standards defined by ISO9000.

By contrast, FSA's Enterprise Quality Assurance program varies in length depending on the type of service being utilized. For new application development, the IV&V program will continue for the life of the application effort and result in PRR and PIR reviews. The security and testing components of the program provide support for new or existing applications depending on the size of the application or amount of work requested.

The new Quality Assurance Strategy and Implementation Plan is a continual quality plan for FSA's data. Its timeliness depends on the phase of the methodology being implemented. For instance, the Steering Committee is a long term group that periodically reviews and manages data quality issues within FSA. It never disbands and continually decides which data quality issues are addressed through Working Groups, FSA's own Enterprise Quality Assurance program, or through further research in an oversight phase. The timeliness of these activities depend on the effort needed to solve the data quality issue identified, the new application being built, or the amount of time it takes for cross program analytics and/or audits.

2.6.4 Methodology Rankings

According to Gartner's Process Model Selection Framework pictured below, various quality methodologies are charted relative to each other based on IT specificity (or relevance to the Information Systems (IS) organization) and their level of abstraction (how detailed they are in implementation). Gartner further mentions that their clients have successfully combined models such as ITIL, CMM and Six Sigma. [17]

[17] © Copyright Gartner, Inc. Source: "IS Process Improvement: Making Sense of Available Methods." July 11, 2003.

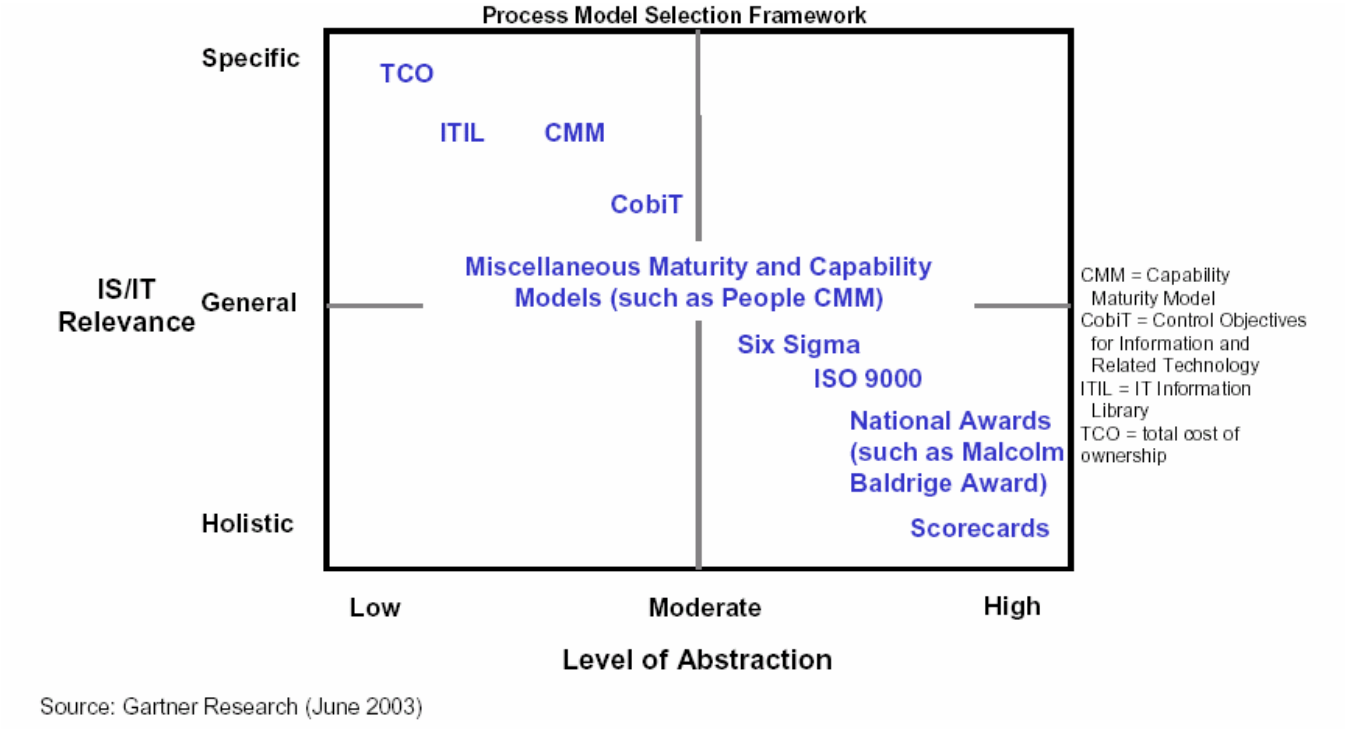


Figure 7 . Data Quality Methodologies: Model Comparison

Gartner mentions, "No 'best practice bundle' of approaches will work for every enterprise or IS organization. Cultural predisposition and the goals and objectives of the process or quality initiative must be considered when choosing a model."[18] Essentially, the process model selection framework can be used as a position marker for each model based on abstraction and IT relevance, but ultimately specific tools and concepts from each model will need to be hybridized into an FSA specific Data Quality Strategy.

[18] © Copyright Gartner, Inc. Source: "IS Process Improvement: Making Sense of Available Methods." July 11, 2003.

3 Data Quality Assurance Strategy

Utilizing concepts from the key industry quality methodologies described in the previous section and input from FSA subject matter experts, a Data Quality Assurance Strategy has been created to address FSA's unique business needs. The diagram below outlines the quality methodology evolution, showing the discussed methodologies as they combine to form the FSA Data Quality Methodology. This Methodology is the core of the new FSA Data Quality Assurance Strategy. This section details the new FSA Data Quality Assurance Strategy by first identifying the various groups involved with the Strategy, followed by a discussion of the Data Quality Methodology's phases and stages.

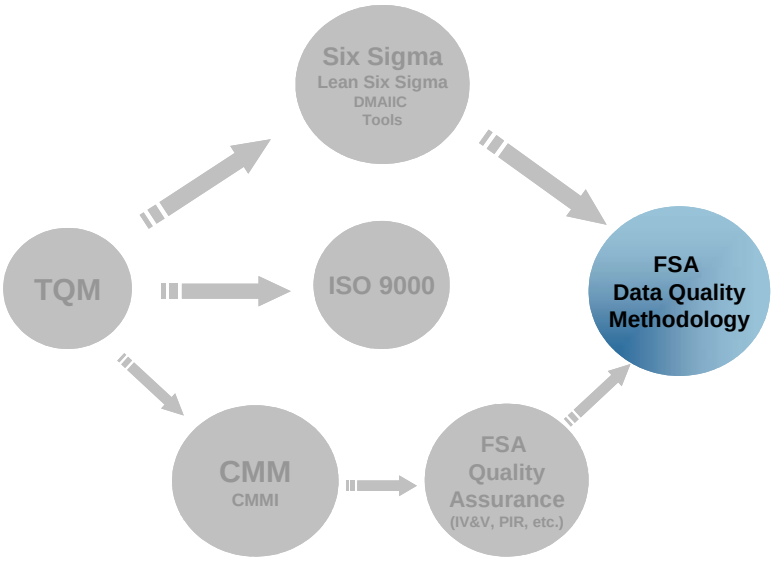


Figure 8 . Data Quality Methodologies: Data Quality Assurance Strategy Methodology

3.1 Data Quality Steering Committee

Central to the new Data Quality Strategy is the formation of a high-level Data Quality Steering Committee that ensures the continuity of the data quality initiative throughout the enterprise. The Data Quality Steering Committee will act as the program management team for all data quality initiatives within the enterprise. It will enforce the standards and procedures of the Data Quality Methodology while monitoring enterprise-wide quality issues. The Steering Committee will receive reports from individuals working on quality issues, obtain status from the Enterprise Quality Assurance program on new project development, and obtain analytics/audit findings from quality analysts. Key responsibilities of the Steering Committee will include:

- **Identify Enterprise-wide Data Quality Issues:** Maintaining data quality is an ongoing initiative. Periodic reviews are needed to ensure data quality levels remain within accepted set standards. The Steering Committee will utilize As-Is state information from an Enterprise Quality Report (created and maintained in the Oversight Phase) to obtain an enterprise-wide and system level view of current FSA data quality issues and performance concerns. The Steering Committee will create enterprise-wide quality criteria to prioritize the issues for assignment.
- **Rank Quality Issues:** Based on the criteria defined by the Steering Committee, each quality issue will be ranked and sequenced in the overall Quality Assurance Strategy. Three outcomes from this phase are possible: (1) assignment to specialized Working Groups who provide and implement a solution for the issue, (2) notification to the Enterprise Quality Assurance program that new development effort will resolve the issue, or (3) assignment to the analysts on the Steering Committee for additional analytics and research.
- **Provide Oversight Functionality:** The enterprise will be continuously monitored and measured for data quality issues within systems, between systems, and external to FSA. Business Intelligence tools (e.g., reporting applications, etc.) can be utilized in this phase to flush out and illuminate data inconsistencies. Data mining can also be utilized to identify trends within the databases. This phase can also be used for verification of quality issues that have been corrected and to identify potential opportunities for improvement in the enterprise.
- **Integrate Data Quality Procedures across the Enterprise:** The Steering Committee will use various quality tools (business case, project objectives, stakeholder analysis, and communication plan) to integrate and ensure awareness of the data quality standards and procedures throughout the enterprise.
- **Create Enterprise-wide Data Quality Education Strategy:** The Steering Committee will need to educate users across the enterprise by developing training materials that illustrate how the new Data Quality Assurance Strategy brings the enterprise in line with FSA's high-level business objectives (see Appendix C for a complete list of High Level Business Objectives). A communication plan in conjunction with training will assist in preparing system owners for data quality implementations as well as increase data quality awareness across the enterprise. Through education, FSA staff will have communication channels to raise quality issues and assist in providing effective solutions.

The composition of the Steering Committee will change depending on the task at hand. For example, in the Data Quality Prioritization Phase the Steering Committee may be comprised of more upper level managers and system owners who determine enterprise-wide sequencing of quality issues. This group may be tasked with presenting the status of the overall quality strategy to FSA management council. The composition of this group may only require one analyst to perform quantitative work of updating the Quality Report and ensuring that meetings are recorded correctly. In contrast, during the Quality Oversight Phase, members may include more technical and business analysts to accurately reflect the cross program input needed to run enterprise analytics and audits (potentially through the use of a quality tool).

At its core, the basic structure of a Quality Steering Committee contains the following personnel:

- **Data Quality Administrator:** The Data Quality Administrator should be a full time position that requires a technical and business background. The main responsibility of the Data Quality Administrator will be to serve as the program manager for all data quality work within FSA. This person will review the Data Quality Strategy and, along with the other members of the Steering Committee, use this methodology as an aid in implementing the Data Quality Plan. This person will focus on several High Level Business Objectives such as; increasing efficiency in data handling, reducing redundant data, maintaining a clean audit, and removing FSA from the GAO high-risk list. Additional responsibilities will include:
  - Create a list of cross program Business Owners and technical staff for formation of a Steering Committee. This list will be initially reviewed and approved by the FSA Business Integration Group (BIG). Modifications to the Steering Committee will be approved by the Data Quality Administrator and FSA's BIG.
  - Initiate the Quality Assurance Strategy. Based on the Implementation Plan and work completed in the Data Quality Mad Dog Report (Deliverable 123.1.3), the Data Quality Administrator will initiate the Prioritization Phase of the current data quality issues within FSA.
  - Provide project management for the Quality Assurance Strategy through the development of enterprise project plan and sequencing plan for system specific implementations. The Data Quality Administrator will have ownership of and be responsible for implementing the FSA quality plan across the enterprise.
  - Coordinate quality maintenance activities and schedule systematic periodic reviews of FSA processes.
  - Coordinate meeting agendas and set meeting frequency and times.
- **Cross System Life Cycle Subject Matter Expert(s) (Application / Delivery / Institution Participation / Servicing):** The Cross System Subject Matter Expert (SME) will need knowledge that spans across Life Cycle Phases and system boundaries. The Cross System SME must have an understanding of how data, and changes to the data in one system, might affect other systems and the enterprise globally. Responsibilities of this team member(s) will include:
  - Provide a deep understanding of the business, processes, and functionality within FSA. This person must understand, from a conceptual standpoint, the detailed data flows and interfaces, both internally and externally, and the major quality issues within the enterprise.
  - Deliver functional interface knowledge. For specific issues involving interfaces, this person must have the functional knowledge of what and when data is needed in order to determine if data is missing or incorrect.
  - Assist in assigning business value to quality problems and in verifying of the importance of each issue.
- **Technical Architecture Specialist(s)**: As problems are discovered in the enterprise, it will be important for the Steering Committee to include a technical resource that understands

databases, different hardware platforms, networks, and software. This resource will allow the Steering Committee to make informed decision based on not only a sound business and technical foundation (This resource could potentially be drawn from the CIO group). Additional activities for this team member will include:

- Participate in the Data Quality Prioritization Phase.
- Provide key FSA technical knowledge for tool selection for data quality applications.
- Provide technical SME input through the Quality Assurance Life Cycle.

- **Analyst(s):** A business or technical analyst will be needed to ensure that project plans are updated, meeting minutes are taken, and that Quality Report updates are performed correctly. This resource may also have responsibility for initially creating prioritization criteria and management of the Enterprise Analytics and Audits needed for verification and discovery of quality issues within FSA.
- **Core Team Representatives**: Core Team Representatives are Project Managers from each Working Group. An extension of their role on the Working Group will be as liaison to the Steering Committee with status updates on data quality implementation and maintenance results (Additional responsibilities are defined in the Working Group Section under Program Manager in Section 3.2 Data Quality Working Group).

Once the data quality Steering Committee is formed, its first task should be a categorization of the known issues into a prioritized list for data quality assessment and improvement. A foundation for discovering data quality issues can be found in the Data Quality Mad Dog Report (Deliverable 123.1.3).

3.2 Data Quality Working Group

Each Data Quality Working Group will assess and fix a specific data quality issue out of the list of issues identified by the Steering Committee. The goal for each Working Group will be to resolve an identified issue in an effort to raise data quality levels to meet standards and overall needs of Business Owners throughout the enterprise. Depending on the issue, the Steering Committee may delegate Working Group formation to a Project Manager, most likely a manager familiar with the issue or systems involved. The Project Manager's initial task will be to determine the Working Group members based on who in the organization would best be able to supply the critical insights and access to the data being studied. Optimally, a Working Group will consist of, but is not limited to, the following individuals:

- **Project Manager:** The Project Manager will be assigned by the Steering Committee to develop a Working Group based on a specific data quality issue. The Project Manager will be the liaison between the Working Group and the Steering Committee and therefore responsible for attending select Steering Committee meetings. The Project Manager will also be responsible for the administrative functions of the group such as allocating resources (e.g., staff and budget), record keeping (e.g., meeting notes), and creating reports (e.g., Data Quality Analytic Reports, Steering Committee updates, etc.) Additional responsibilities will include:

- Coordinate system owners and technical activities – The system owner and technical staff provide different views of the impact of data quality issues on the enterprise. The Project Manager will be tasked with merging these views to ensure a collaborative effort for improving the data quality within the systems.
- Develop assessment plans – The Project Manager will lead the Working Group as they determine the assessment plan for each data quality issue identified. The quality issues will be assessed to determine the cause and impact of the defect.
- Organize and lead solution development – Once the assessment has been completed the Project Manager, with the input of the SME and Database Administrator (DBA), will propose solutions for pilot testing. Once a pilot test is completed successfully, the solution will be scheduled for production implementation. Results from the Improvement Phase will be passed to the Steering Committee in the Oversight Phase.

- **System Owner / Subject Matter Expert(s):** The System Owner/SME will provide functional knowledge in the Assessment and Improvement Phases. This person will be instrumental in determining the scope of the quality issue, providing a data sample that adequately demonstrates the issue, and ensuring that a solution is defined that corresponds with the business objectives of FSA. Other key tasks will include:
  - Review all data samples and solutions for the quality issue.
  - Identify and report any new data quality issues that may result from the Pilot Testing Stage.
  - Ensure compliance with business rules and standards.
- **Database Administrator (optional):** Depending on the quality issue, a Database Administrator (DBA) may be needed to ensure that the correct data is loaded into a system that is affected by the issue. The primary responsibility of the DBA will be to provide technical expertise in specific system databases as they relate to the quality issue. The DBA may also be asked as part of on going maintenance to assess the quality and reliability of existing data, itemize known quality issues, and to offer recommendations for an improvement strategy. Additional responsibilities will be as follows:
  - Set up and maintain the test production environment and provide the test data used in the Assessment Phase.
  - Assist in impact analysis of issues on the technical infrastructure of FSA.
  - Write solution scripts, test scripts, and run scripts as part of a Pilot or Production Improvement Phase.
  - Provide input into the Oversight Phase by performing maintenance activities for a specific database. This will result in the DBA providing reports to the Quality Report that lists known issues and identify new issues.

3.3 Data Quality Methodology

3.3.1 Overview

The Data Quality Strategy was formed through an iterative process that included a foundation formed on industry proven practices, discussions with quality experts in government and private industries, and techniques from previous data quality initiatives. The other internal

Data Strategy teams (XML, Access and Enrollment Management, Technical Strategies, and Common Identifiers) also supplied input ensuring cross team integration in creating a holistic methodology for FSA. Once a draft was completed, key FSA Business Owners reviewed and provided validation. The following diagram shows the various phases and stages of the Data Quality Assurance Strategy:

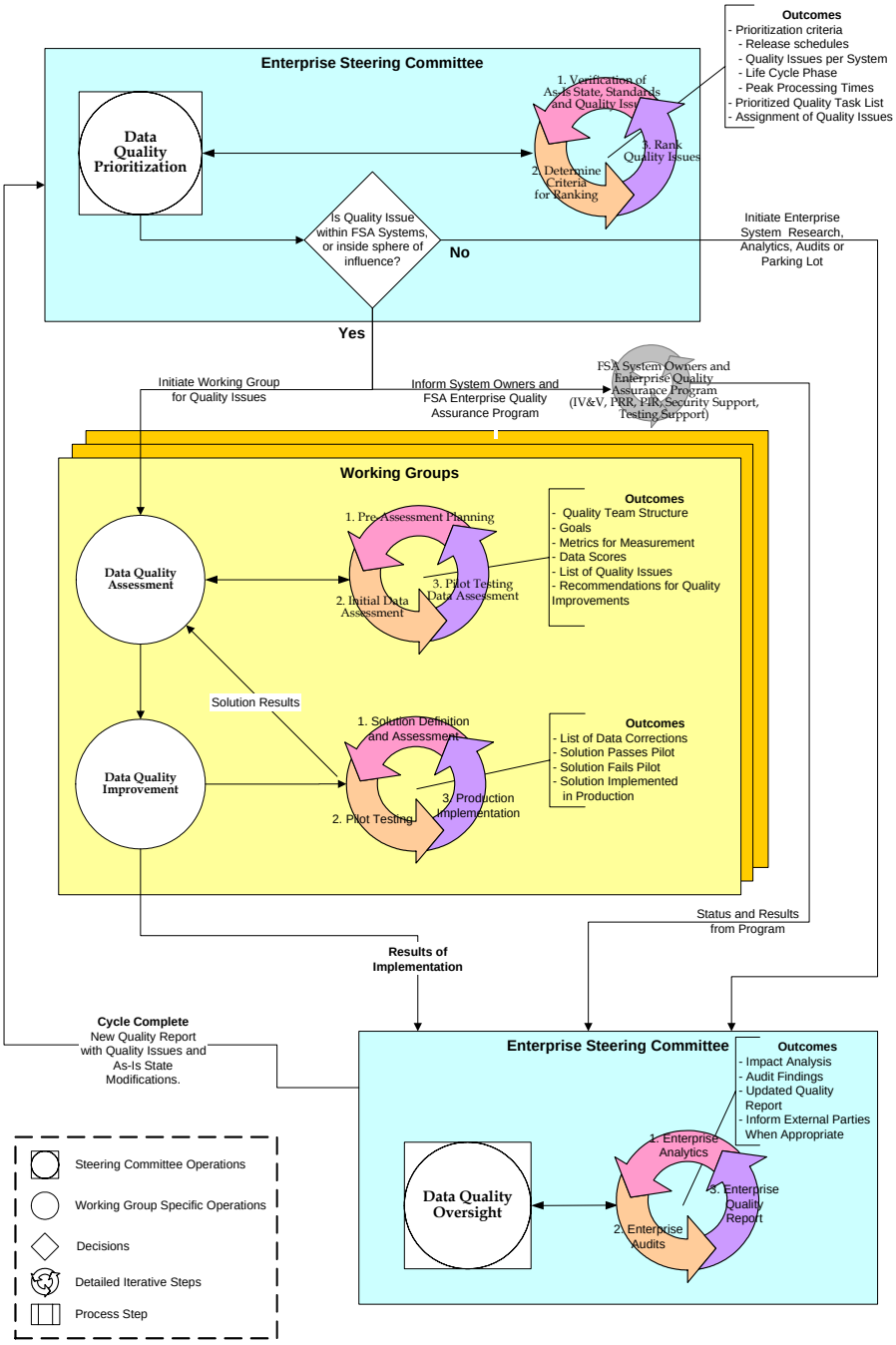


Figure 9 . Quality Assurance Strategy Methodology: Overview

At a high level, the Data Quality Methodology can be broken up into four major phases:

- **Data Quality Prioritization Phase:** The Data Quality Methodology will be initiated with the Data Quality Prioritization Phase conducted by the Steering Committee. This phase will have three stages, the first of which is Verification of As-Is State, Standards, and Quality Issues. As part of this stage, the Committee will utilize a Quality Report to gauge the success of each quality initiative across the enterprise and to prioritize new data quality issues. These data quality issues will be verified against the current As-Is state of the enterprise. In the second stage, Determine Criteria for Ranking, a set of rules will be established. After these rules are created, the Steering Committee will rank the issues and assign Project Managers to develop Working Groups around each issue as part of the last stage, Rank Quality Issues.
- **Data Quality Assessment Phase**: The Steering Committee will form an issue specific Working Group to implement a Data Quality Assessment whereby the data is inspected, the defects are measured, and the cause and impact of the defects are analyzed. The Assessment Phase will be accomplished with three stages: Pre-Assessment, Initial Data Assessment, and Pilot Test Data Assessment.
- **Data Quality Improvement Phase**: Based on quality findings in the Assessment Phase, the Improvement Phase will involve developing and assessing solutions to resolve the issue. This will be conducted in the Solution Definition and Assessment Stage and also involves test piloting a selected solution option. Following successful test pilot results, the Working Group will implement the solution and correct defective data in the Data Quality Implementation Phase through three processes: Cleaning and Scrubbing, Correction, and Reconciliation.
- **Data Quality Oversight Phase**: During this phase, the Steering Committee team will review the results and methodology of each Working Groups issue analysis and resolution. If needed, further analytics and financial audits will be conducted in the Oversight Phase to gauge the enterprise-wide impact of the solution. All results and findings will be updated to the Quality Report, which is in turn utilized by the Steering Committee for monitoring current issues and prioritizing new ones. There will be routine maintenance checks and ad-hoc analytics requested by the Steering Committee. The schedule for maintenance processes will be determined by the Steering Committee and may involve analytics and audits or may simply involve report generation off the Quality Report. The main goal of Oversight will be to ensure data quality standards is constantly maintained throughout the enterprise.

At each quality phase, a group of tools and processes will support the objectives of the Steering Committee and Working Groups. These tools and processes are garnered from industry best practices and from the methodologies described above. Each phase of the Quality Assurance Strategy will discuss the different tools available to enable the processes of that phase.
The following is a high level mapping between the Quality Assurance Strategy phases, process stages, and the tools that sustain them.

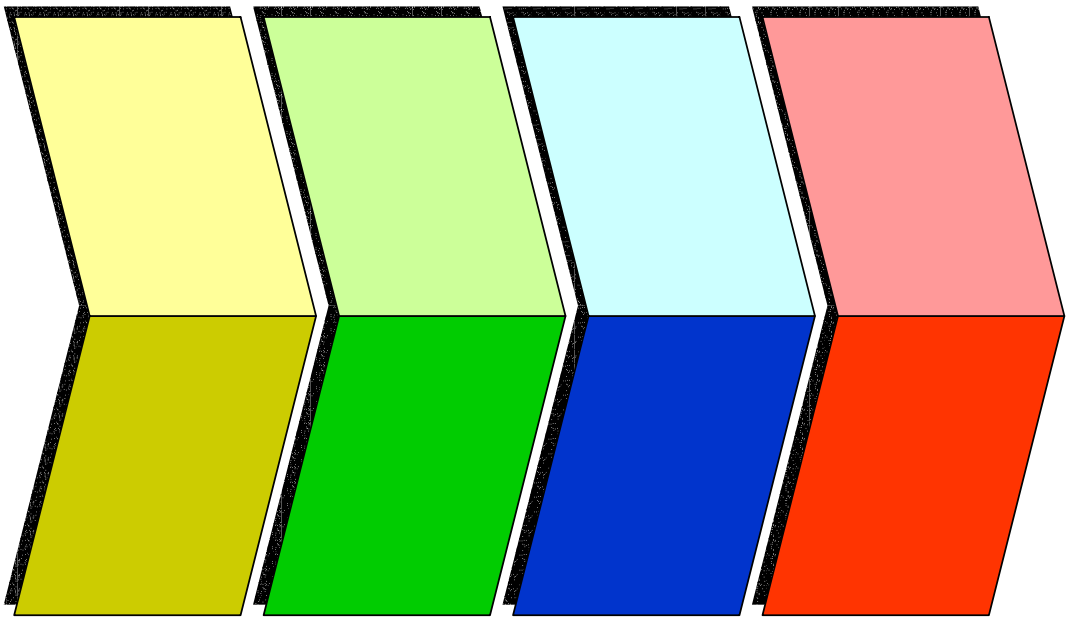


Figure 10 . Business Process Tools Mapping: Methodology Overview

The rest of this section will provide a detailed explanation of each of the phases described above.

3.3.2 Data Quality Prioritization Phase

The sequencing of FSA Data Quality Issues (existing and new) into an Enterprise Data Quality Plan is a multifaceted process requiring different perspectives to determine the criteria necessary for ranking. System owners may concentrate on the number of data quality problems within each system whereas a Data Architect might place greater stock in addressing Quality Issues in systems that are upstream within the organization (e.g., Central Processing System). Below is a process flowchart of the stages and processes in this phase.

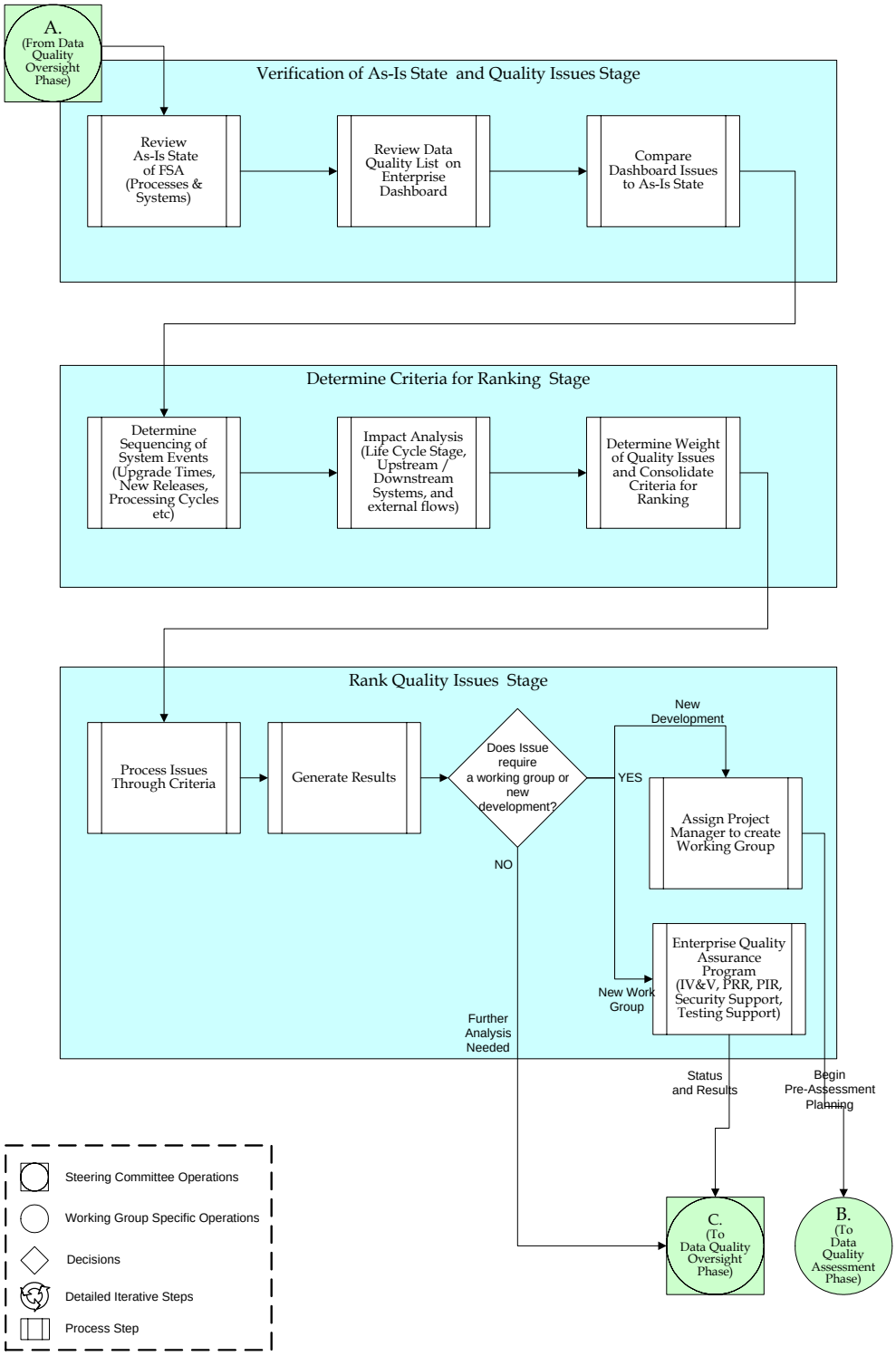


Figure 11 . Data Quality Methodology: Prioritization Phase

As a first step in the Data Quality Prioritization Phase, a snap shot of the enterprise will be taken through the utilization of a Quality Report tool (this tool will be created and updated in the Data Quality Oversight Phase). This tool will record a list of quality issues and major process changes to the enterprise as dictated by subject matter experts, system owners, internal analytics and auditing processes. The Steering Committee will use this list as a starting point to determine enterprise wide criteria to compare quality issues and processes within FSA. A variety of criteria can be used to examine quality issues across the enterprise such as the cycle of system events, the flow of data in the enterprise, and a list of current quality issues. Once the criteria have been determined, quality issues will be ranked and sequenced within the context of FSA business objectives. Steering Committee members should keep in mind the business drivers for the issues resolution such as:

- Will the issue's resolution result in enhanced ease of use by customers?
- Will the data quality enhancement lead to increased business decision efficiency?
- Does the issue require revision of policy standards, both internal and external to the enterprise?
- Will the issue's resolution improve analysis?

These questions map directly to the Data Strategy Objectives (see Appendix C) and should be factors when ranking the issues. A positive answer to most or all of the questions above may indicate that the issue should be given additional weight in comparison to issues that have few positive responses to the criteria, since the issue's resolution correlates with Data Strategy Objectives.

Once the issues have been ranked, the Steering Committee will determine whether the quality issues are solved by new development efforts (notification is then sent to the FSA Quality Assurance team for IV&V examination), additional research is needed (the issue is sent to the Oversight Phase), or a Working Group needs to be formed for the issue.

The tools and analysis techniques that will support this phase include review of the Quality Report, a Cost of Poor Quality analysis, Failure Modes Analysis, and an examination of how each quality issue affects the overall amount of improvement of the business drivers defined above.

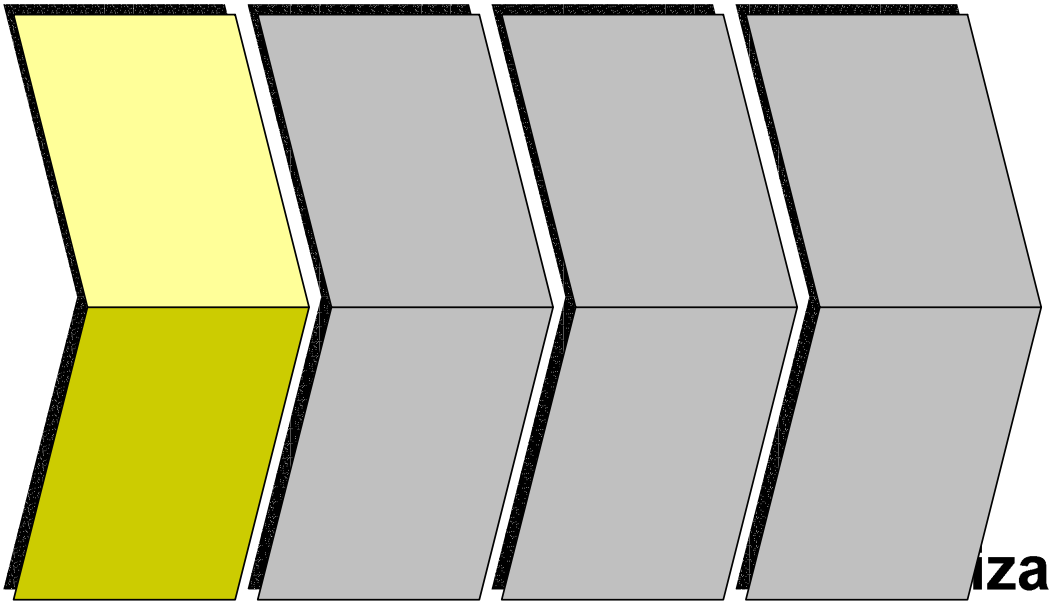


Figure 12 . Business Process Tools Mapping: Prioritization Phase

3.3.2.1 Verification of As-Is State and Quality Issues Stage

FSA is large complex organization with multiple business channels (e.g., student, schools, and financial partners) controlling multiple physical systems (e.g., FAFSA, CPS, PEPS, ezAudit, eCB, etc.). At times, not all organizations in the enterprise are aware of process or system changes. An examination of the As-Is state of the enterprise will level set all Steering Committee members as to the current data flows and quality issues facing FSA. The Enterprise Quality Report (constructed in the Data Quality Oversight Phase) not only will track all quality issues within the enterprise, but can also be utilized to track all major process and system changes that may affect FSA.

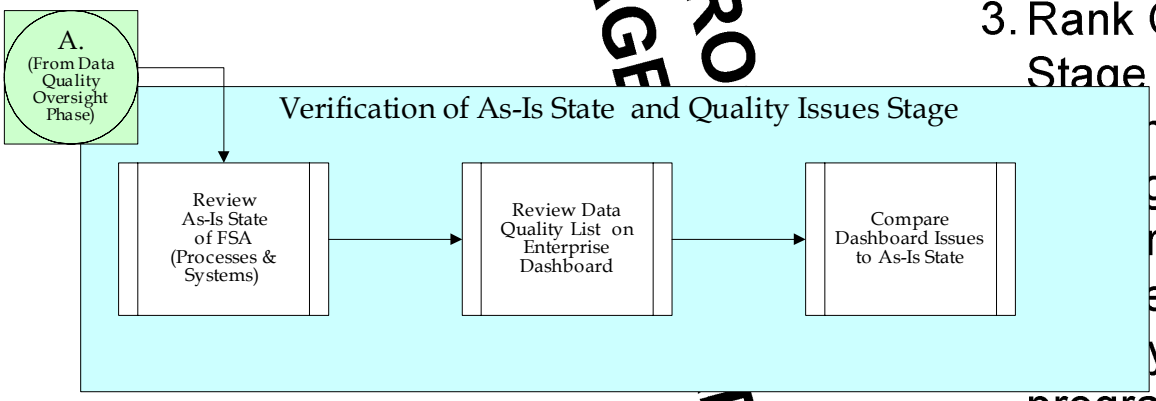


Figure 13 . Prioritization Phase: Verification of As-Is and Quality Issues Stage

The Verification of As-Is State and Quality Issues Stage is outlined in the diagram above and will involve the following processes:

- **Review As-Is State of FSA (Processes & Systems):** The Steering Committee will look at the current systems and business processes within the FSA Enterprise to understand the various systems that may have been created, removed, or modified since the Committee last convened.
- **Review Data Quality List on Enterprise Quality Report:** The Committee will consult the Quality Report to review all reported data quality issues.
- **Compare Quality Report Issues to As-Is State:** From the Quality Report, the Steering Committee members will compare the As-Is state of FSA with the known quality and processes changes within the organization. This will ensure that any reported issues have not already been resolved by recent systems modifications or enhancements.

Additionally, as part of this stage, contact information for each key business and technical owner will be confirmed to ensure a true cross life-cycle group is represented when determining the criteria for issue evaluation. Without proper representation, the selection criteria may not capture a complete set of data quality business processes issues or technical errors.

3.3.2.2 Determine Criteria for System Ranking Stage

Data Quality Issues can be discovered at every phase of the quality assurance strategy. A data issue might reside in a database, a Business Owner might discover a broken process, or an external agency might create a variety of issues (i.e., transfer errors, incorrect information, lack of information, etc.). The first step in determining the measurements for evaluating the issues will be a series of meetings by the Steering Committee with cross-lifecycle representatives. This group will confirm, delete, or update current defects, quality issues, and potential solutions that exist in the Enterprise Quality Report. This work can begin with the Data Quality Mad Dog Report (Deliverable 123.1.3).

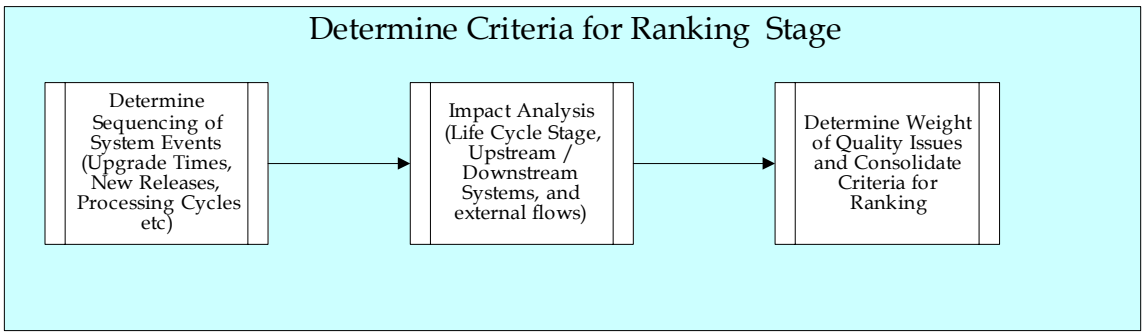


Figure 14 . Prioritization Phase: Determine Criteria for Ranking Stage

The Failure Modes and Effect Analysis (FMEA) will be a useful tool in this stage for identifying and prioritizing gaps in process flows. Potential points of failure in each step associated with a quality issue are identified. These failure modes will be reviewed with SMEs and severity levels will be determined. Next, potential causes for the failures will be explored and listed with their reported frequency and the current controls in the system which will prevent the failure from occurring. Finally, a detectability level will be reviewed and a total score will be calculated by multiplying the severity, occurrence, and detectability scores.

Results of this analysis will be key components of the Rank Quality Issues Stage of the Prioritization Phase of the new Quality Assurance Strategy. The following is a high level description of the different components of the FMEA.

Failure Modes and Effect Analysis (FMEA)

Process Step	Potential Failure Mode	Potential Failure Effects	Severity	Potential Causes	Occurrence	Current Controls	Detectability	Score
What is the Process step under investigation?	In what ways does the Process Step go wrong?	What is the impact on the Key Output Variables or internal requirements?	How severe is the effect to the customer?	What causes the Key Input to go wrong?	How often does cause of Failure Mode occur?	What are the existing controls and procedures (inspections and test) that prevent either the cause or the Failure Mode?	How well can you detect cause or Failure Mode?	Score = Severity * Occurrence * Detectability

Severity:	Occurrence:	Detectability:
1 = None	1 = Almost Never	1 = Excellent
4 = Moderate	4 = Occasionally	2 = Some leaks
7 = High	7 = Frequently	3 = Frequent Leaks
10 = Extended Shutdown	10 = Almost Always	10 = Almost Undetectable

Table 3 FMEA Analysis

In the example below, (see Table 4 - FMEA SSIM Example) Data Quality Mad Dog Issue 10 (i.e., records are accepted in FSA systems with incorrect identifiers) was used to create a hypothetical FMEA. The first column notes that the process step associated with this issue is FSA processing and validating incoming applicant records. There is a failure mode when records with invalid identifiers are allowed to enter FSA.

The effects of this quality issue are that recipient's data may be entered twice into a system, each with a different SSN, creating a duplicate record for the same recipient. Furthermore, updated information for the recipient is not accurately tracked within FSA (i.e., if a recipient updated their information there is no guarantee that all systems within FSA will have the latest information), and there is also potential for errors when disbursing aid to the individuals. These effects likely warrant the Steering Committee and Subject Matter Experts to assign a "high = 7" severity level.

Potential causes are detailed as being lack of Social Security Number (SSN) verification by all systems and the number of occurrences are ranked as "occasionally = 4". Only manual controls were identified and the detectability was ranked "frequent leaks = 3." The final score (i.e., 84 in this example) then provides a tool for the Steering Committee to compare and rank the issue against other quality issues.

Failure Modes and Effect Analysis (FMEA)

Process Step	Mode	Effects	Severity	Potential Causes	Occurrence	Current Controls	Detectability	Score
Applicant records are processed and validated by FSA.	Records with incorrect identifiers are allowed to enter FSA.	Invalid Student borrowers can be created. A student could potentially have multiple SSNs in FSA. This may also cause aid to be disbursed inaccurately	7	Lack of a SSN match in COD, DLCS, DMCS, and NSLDS (systems other than CPS and PIN site do not verify SSN)	4	No automatic verification of SSN. Manually have to look up borrower and fix information.	3	84

Table 4 FMEA SSIM Example

This tool should be used as a repeatable process for each quality issue and can initially be done by a business analyst to save the Steering Committee time when reviewing quality issues. In addition to the score calculated from this tool, other candidates for measuring and ranking issues are as follows:

- **Determine Sequencing of System Events:** The Steering Committee will chronologically list the systems according to their current release schedules, system upgrades, and enhancements. Peak processing times should also be noted to ensure that quality initiatives occur in down cycles. Using information supplied by SMEs, the Committee will identify the flow of data within and external to the enterprise as related to quality issues contained within the Quality Report. Understanding the flow of data will allow the Steering Committee to develop criteria for when to focus quality initiatives on external facing systems (i.e., allowing quality work in those systems that effect downstream processing) or internal systems (i.e., displaying the impact of the quality initiatives on other systems or interfaces).
- **Impact Analysis:** An impact analysis will examine the effect the quality issue has on the business process, aid life cycle stage (e.g., Application, Delivery, Institution Participation, or Servicing), data flows within the Enterprise, and systems or trading partners. This analysis may contain a risk management plan or a Cost of Poor Quality (COPQ) analysis. In respect to a COPQ, this analysis will quantify the economic impact associated with the quality issue and will consist of those costs that are generated as a result of producing defective material. This will include the cost involved in fulfilling the gap between the desired and actual product/service quality. It also will include the cost of lost opportunity due to the loss of resources used in rectifying the defect. Some typical examples are:
  - Material and labor costs for repairing the problem
  - Cost of external trading partners producing defective material (e.g., data, images, incorrect processes)
  - Cost of lost opportunity
  - Loss of sales in aid packages and potential revenue for lenders, GA's ,and schools
  - Lower service level to customers/consumers
- **Determine Weight of Quality Issues and Consolidate Criteria for Ranking:** Once a comprehensive list of data quality criteria has been created, the results should be formulated in such a way that a final consensus meeting can be held so that all experts can determine the level of importance for each criteria. Each criteria should be weighted and all data

quality issues consolidated into a single list to be evaluated against these criteria. This step feeds directly into the ranking of the quality issues.

3.3.2.3 Rank Quality Issues Stage

After the criteria have been developed and approved by the group, the Steering Committee and additional cross system SMEs will rank the issues to determine their implementation sequencing. This ranking should factor in the sequencing of enterprise events, the impact of the effect of a quality initiative on external and internal systems, and the data quality issues most prevalent in the enterprise. This Ranking Stage also will involve formulating next steps for delivering the issue to either the Working Groups or FSA Quality Assurance Program. The process steps involved with this stage are shown in the diagram and described below:

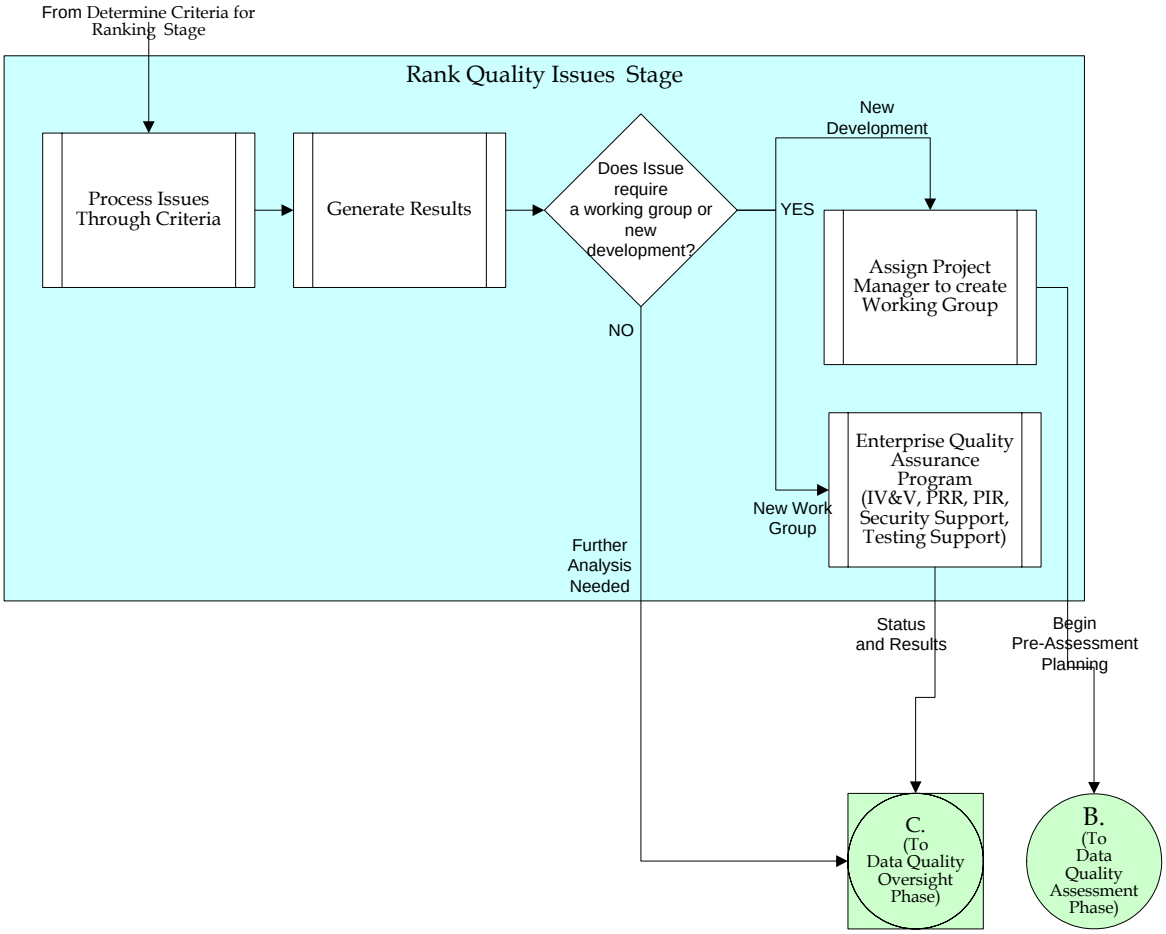


Figure 15 . Prioritization Phase: Rank Quality Issues Stage

- **Process Issues Through Criteria –** Each issue will be assessed with the listed criteria, including following through the FMEA analysis.

- **Generate Results –** The issues, having been evaluated against the established criteria, will be ranked relative to each other based on how critical the issue is. This comes directly from the results of a combination of criteria and FMEA analysis.

At this point, the Steering Committee has generated a ranked list of data quality issues and must determine which issues will be resolved first. They will also need to identify whether the issue will require the formation of a Working Group or whether it will be handled by new development efforts underway. Both avenues are discussed below:

- **Enterprise Quality Assurance Program –** As discussed above, the issue may be handled by new development efforts. In this case, the issue will be sent to the FSA Quality Assurance Program team in addition to the appropriate development team. The issue will be handed off and does not require Working Group resources for resolution. All results from the new development effort will be directly reported to the Data Quality Oversight team.
- **Assign Project Manager to create Working Group –** The Steering Committee may decide that a Working Group around the issue needs to be established for further analysis with existing development efforts. The Committee will assign the issue to a Project Manager who is knowledgeable in the systems involved effected by the issue and will be in charge of Working Group development. Once assigned the issue, the Project Manager will begin the Data Quality Assessment Phase.

3.3.3 Data Quality Assessment Phase

As part of Data Quality Prioritization, the Steering Committee will determine whether an issue will be addressed by future development work or will require additional data quality efforts. For new development work, FSA will utilize the established Enterprise Quality Assurance Program (IV&V, PRR, PIR, etc.). However, for those issues not addressed by future development work, the Steering Committee can either assign a Data Quality Working Group or request further research be performed in the Oversight Phase.

For those issues that require a specialized Data Quality Working Group, the core team members will consist of a Project Manager, System Owner, and a Technical representative (e.g., database administrator, interface analyst, or system administrator, depending on the quality issue). This group will be tasked with completing the Data Quality Assessment and Data Quality Improvement Phases of the Data Quality Methodology. The first of these phases, Data Quality Assessment, will allow the Working Group to develop the criteria for identifying the severity of each issue and its variation from established data quality standards. As the Working Group develops these criteria, additional issues and affected areas may be highlighted. There will be three stages in the Data Quality Assessment Phase: Pre-Assessment Planning, Initial Data Assessment, and the Pilot Testing Data Assessment, with each stage containing a number of process steps.

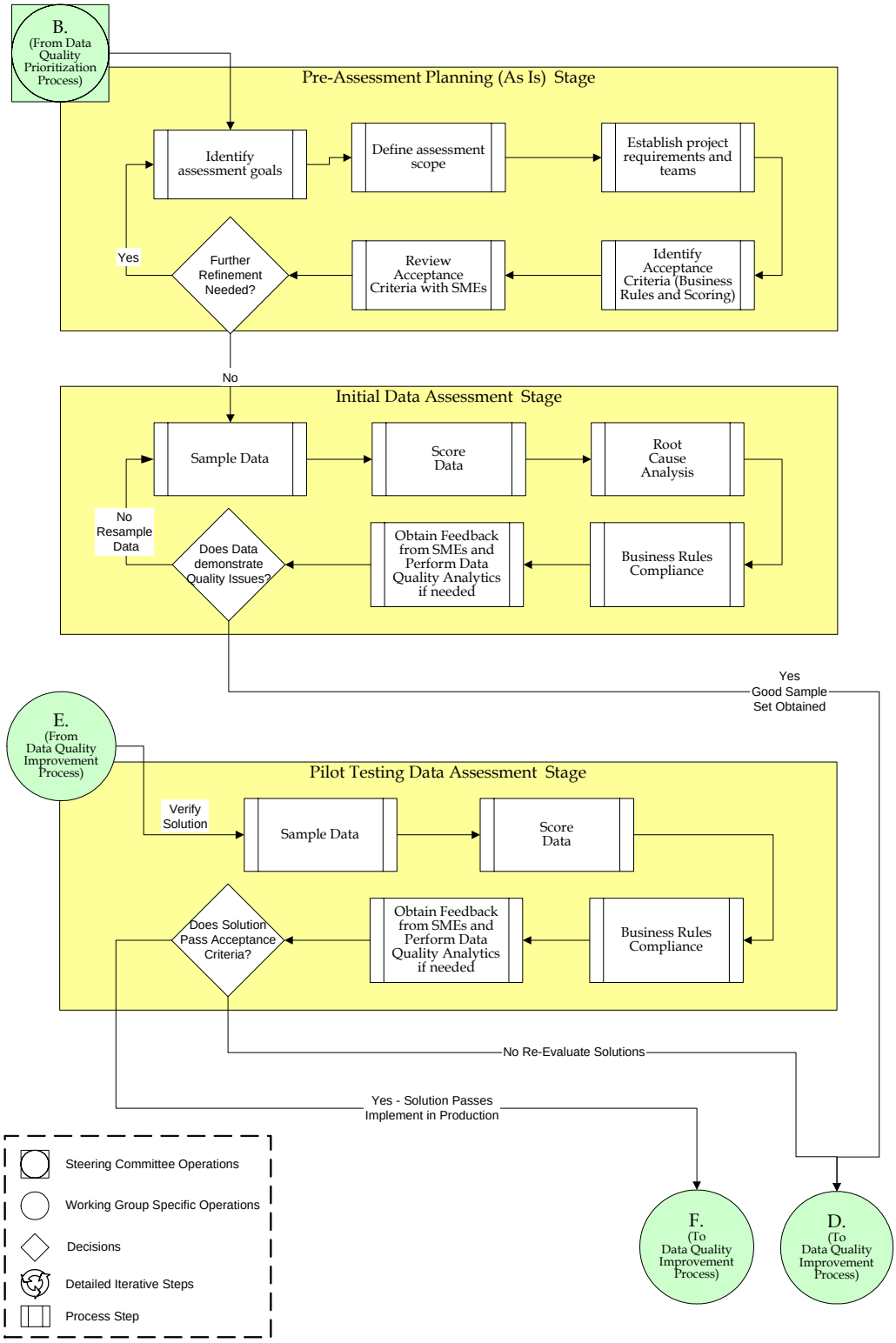


Figure 16 . Data Quality Methodology: Assessment Phase

The business tools that support this phase will include development of a Project Charter that contains high level summary information on the problem statement, project objectives, scope, key milestones, and team roles/responsibilities. The charter should also contain a business case section to determine the impact of the quality issue on other business units and individuals. After completing the charter, a proper communication plan will be instrumental in ensuring that all parties understand the quality issues and necessary changes. Feedback to this communication plan, as captured in the Voice of the Customer (Tool), will ensure system owner buy-in and additional insight into the solution for the issue. Histograms, fishbone diagrams, and other graphical tools will help visualize these processes and provide a diagrammatic view of the results.

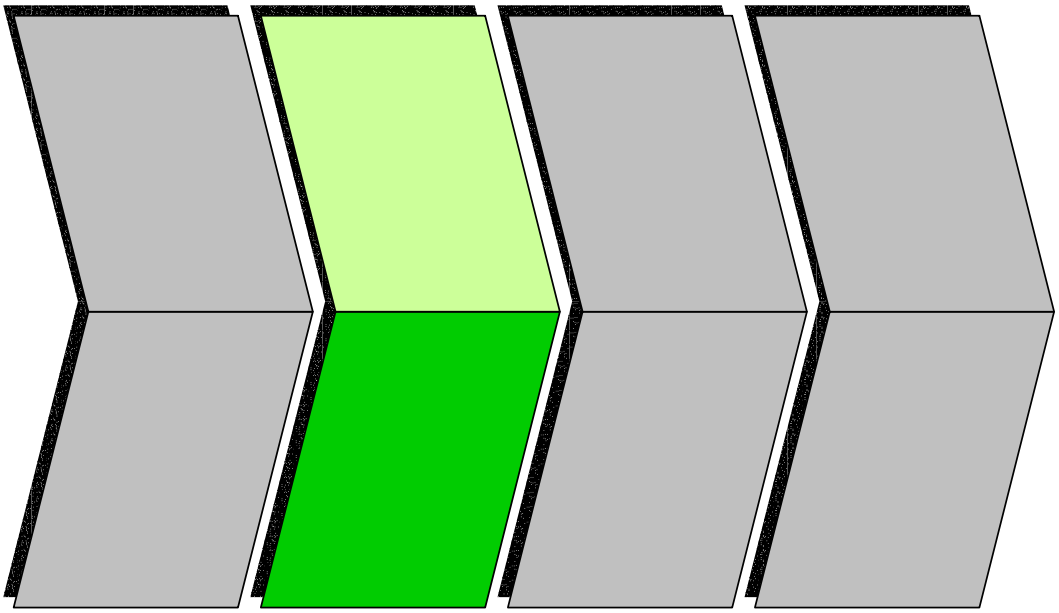


Figure 17 . Business Process Tools Mapping: Assessment Phase

It should be noted that for all Working Groups, a series of standardized templates need to be created that contain outlines of the different tools for both the Assessment and Implementation Phases with examples on how to use them. Working Groups will be required to fill out key documentation within these templates (e.g., Project Charter, Communication Plan, etc.) and choose other templates depending on the quality issue.

3.3.3.1 Pre-Assessment Planning Stage

The first stage of Data Quality Assessment will be Pre-Assessment Planning. During this stage, the Data Quality Working Group will acquire a more complete understanding of the specific issue and develops the requirements and acceptance criteria for an end state vision which rectifies the issue. This stage will involve the following processes shown and described below:

Prioritization Phase

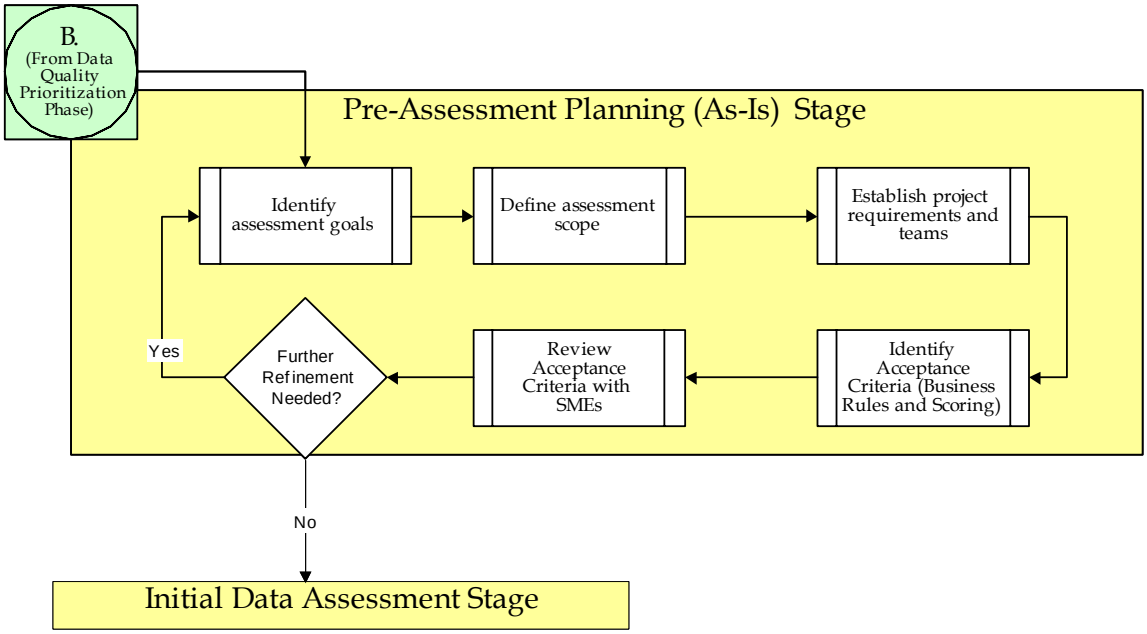


Figure 18 . Assessment Phase: Pre-Assessment Planning (As-Is) Stage

- **Identify Assessment Goals** - The goals of the assessment should be clearly defined prior to beginning the actual assessment. Some of the key questions to be considered are:
  - What is the expected outcome of an effective assessment?
  - What are the perceived data quality issues?
  - Did a critical operational failure occur?
  - What other systems, interfaces, or processes does this issue potentially affect?
- **Define Assessment Scope** - The Project Manager will work with the SME(s) to thoroughly review the data quality issue and to determine the scope of the assessment. Generally, the assessment scope outlines the system impact of the issue and any associated minor issues.
- **Establish Project Requirements and Teams –** Once the assessment scope is defined, the Project Manager will collaborate with the SME(s) to develop a Project Charter. This document will act as a high level summary of the following project concepts and documents:
  - **Business Case:** Discusses how fixing the quality issue aligns with the future state vision for FSA business, data, and technical architectures. This document should also outline how extensive the problem is and its impact on the enterprise. Questions such as whether or not to address the problem, what benefits will be derived, and the value of those benefits may be addressed in this document.
  - **Project Objectives:** The Project Charter will review the assessment goals and define the expectations for the project. It will also ensure that the goals of the project are aligned with the overall FSA business strategy. It is important to set an improvement goal for the quality issue that is aggressive but achievable.

- **Project Scope:** Information from the scope process will be incorporated in summary format into the Project Charter. This will include defining the start and end points for the project. This concept should define the boundaries of the project and identify aspects that are outside the boundaries for the project requiring future follow up.
- **High Level Suppliers, Inputs, Process, Outputs, and Customers (SIPOC) identification:** Identify the current state of the quality issue in relation to the business, data, and technical architectures. This concept will be part of the Root Cause Analysis process in the Initial Data Assessment Stage discussed below in greater detail. As a precursor, the Pre-Assessment Planning Stage should perform a high-level analysis to identify the inputs and outputs of the quality issue.
- **Project Team:** Decide upon the roles and responsibilities for the project team members who will be able to best provide a solution to the issue. Typical tasks for this process will be deciding upon a team lead, core team members, team meetings schedule, and accountability. A discussion with each team member should also be initiated to clearly communicate roles, tasks, and expectations.
- **Project Plan:** Determine key tasks for the project team in pursuit of the objectives for the quality issue. The project plan will document each tasks start date, estimated duration, major milestones, and person assigned.
- **Communication Plan:** Identifies the target audience and shares the status of the quality issue, lessons learned, solutions, and team findings. This tool will usually be created in spreadsheet format with the following column headers: Target Audience, What is Communicated, Frequency of Communication, and Delivery Method. It can be thought of as a tracking tool and feedback mechanism for all communications to other system owners, external agencies, or any others interested parties in the quality initiative.

- **Identify Acceptance Criteria (Business Rules and Scoring) –** In order to determine a proper solution, acceptance criteria and sample data that demonstrate the problem must be identified. The Working Group will develop criteria for acceptable data by examining the current business rules and input from business and/or technical SMEs. The established criteria will then be listed and weighted for an overall scoring methodology. These weighted criteria can be listed in a Data Quality scoring matrix to be utilized later during the Assessment Phase.
- **Review Acceptance Criteria with SMEs and Examine Additional Documentation if needed –** In the final process of the Pre-Assessment Planning Stage, the Working Group will review the established data quality metrics and criteria with SMEs, as well as any additional relevant documentation. If further refinement is necessary, the Working Group will loop back to the beginning of the Pre-Assessment Planning Stage to modify the assessment approach and to ensure the goals and scope have been properly adjusted. Once SMEs have verified the acceptance criteria and determined no further refinement is needed, the Working Group may continue on to the Initial Data Assessment Stage.

3.3.3.1.1 Pre-Assessment Planning Business Tools

The recommended tool for developing acceptance criteria is the Voice of Customer (VOC) process. A VOC involves interviewing system owners for all associated systems to gather data requirements and to identify any end-user considerations. A critical subcomponent tool in the

VOC process is the Critical to Quality (CTQ) analysis. The CTQ analysis translates customer needs derived from the VOC into attributes that are necessary for acceptable quality levels. These attributes are termed Critical to Quality. The steps for conducting the VOC process are outlined below:

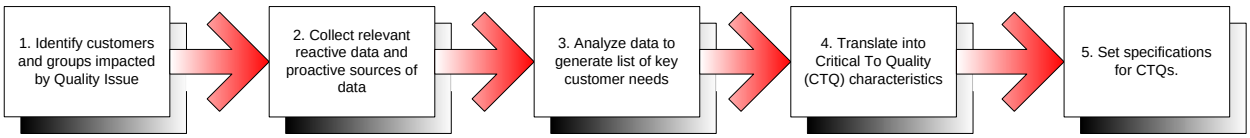


Figure 19 . VOC Process Steps

1. Create a list of customers that are impacted by the solution.
2. Create a list of reactive events (e.g., customer complaints, technical support phone calls, claims, credits, contested payments, web page activity, etc.) and a list of proactive events (e.g., customer interviews, focus groups, surveys, data gathering during product selling, market research, market monitoring, bench marketing, etc.). Reactive systems generally gather information on current and former customer issues, problems, unmet needs, and interest in particular products or services.
3. Identify the users of the product or service. Enter a general customer description and output of the process. Enter the customer requirements and rate the importance of each requirement to the customer and their current satisfaction level. Calculate the priority for each customer requirement. Determine where to focus the process improvement initiative. Customer interviews should be conducted at this stage of the process to obtain this information.
4. Based on the customer needs, define the drivers and data or process attributes that are Critical to Quality.
5. Develop specifications for CTQs so that they can be translated into metrics for data quality measurement. These metrics are used for scoring sample data during the Initial Data Assessment and Pilot Testing Data Assessment Stages.

In the VOC process outlined above, the fourth step involves the creation of a CTQ Tree that defines the business needs, the drivers for those needs, and the CTQs that determine the drivers. To demonstrate the applicability of the VOC and CTQ tools, the hypothetical CTQ tree below outlines the needs, drivers and CTQs related to issue 24 from the Data Quality Mad Dog Report (Deliverable 123.1.3).

Mad Dog Issue 24 notes that there is currently no Social Security Administration (SSA) or Immigration and Naturalization Service (INS) verification of SSN and citizenship for Parent Loan Underwriting for Student (PLUS) borrowers. This can result in the lending of Title IV aid to ineligible borrowers.[19] The VOC analysis would begin by identifying FSA (personnel) as the customer and asking them what problems are occurring regarding the Title IV lending process. In this example, the customer's (FSA) issue relevant to such lending is that an SSA match is not

[19] 123.1.3 Data Quality Mad Dog Report

conducted for all aid recipients. Following these steps, a CTQ tree, as outlined below, may be developed.

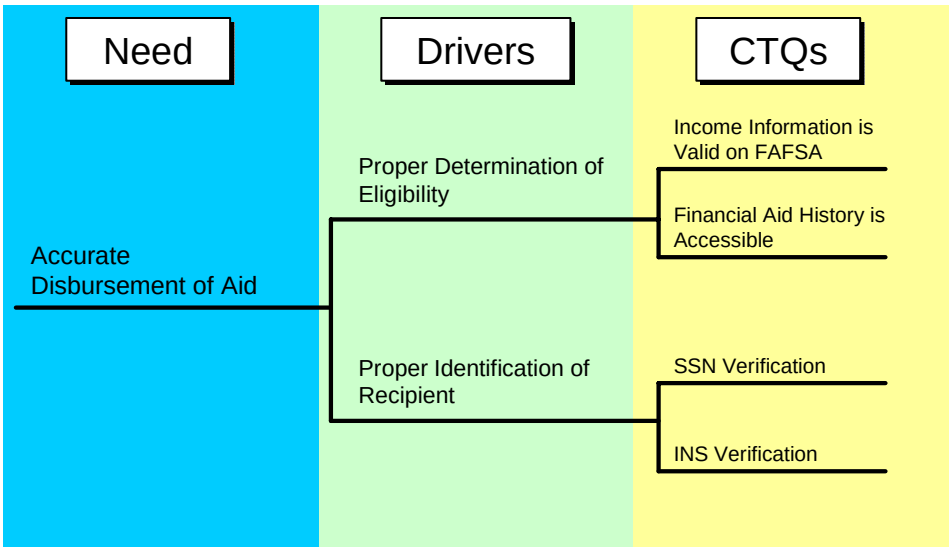


Figure 20 . CTQ Tree

As shown above, FSA has a need for accurate disbursement of aid. This accurate disbursement is driven by the proper determination of eligibility and proper identification of the recipient. Although two drivers have been depicted in the example CTQ Tree, there may be others derived by the Working Group. Each driver then branches into one or more determining attributes, such as Social Security Number (SSN) verification which is a CTQ for proper identification of the recipient. As with the drivers, multiple CTQs may exist and should be explored by the Working Group. After requirements have been gathered, the Working Group can establish specifications and thresholds for the CTQs which delineate acceptable levels of data quality. Within the hypothetical example above, acceptance criteria could include one or any combination of 99% of aid recipients passing SSN verification, 95% passing INS verification, 90% completing financial aid history, and 99% providing valid income information. In the next Initial Data Assessment Stage, the sample data would then be scored against these criteria to verify whether it is below or above acceptable quality levels.

Acceptance criteria can also be created for quality issues that deal with data in interface files. In this case, the Working Group can leverage the XML Core Component Dictionaries (Deliverable 123.1.15) and the Technical Reference and Usage Guidelines Volume 2 – Governance Model (Deliverable 123.1.14). These tools provide the framework for an enterprise dictionary of definitions of commonly exchanged data. The core components are standardized pieces of business information modeled in the eXtensible Markup Language (XML). The Working Group could compare the problematic data to the data standards in the dictionary to determine proper definition and usage within the enterprise. Acceptance criteria would be built to ensure that all data in the interface file maps to the enterprise standards. For that data which does not have an enterprise definition, the Governance Model dictates the organizational structure, guiding principals, and processes for FSA to create and manage the Core Components Dictionaries.

This model includes processes for enhancements to the core components as well as issue management.

The Cohort Default Rate (CDR) is an example of a core component that has been defined as part of the Core Component Dictionary. Using these pre-defined core components stored in the XML registry, a message specification schema file can be created for the interface between NSLDS and PEPS. The schema file can be used to ensure that information is validated between the two systems, and that data exchanged has a common definition throughout the enterprise.

3.3.3.2 Initial Data Assessment Stage

After the Pre-Assessment Planning Phase, the Working Group will conduct an Initial Data Assessment. Data Assessment involves inspecting data, measuring the defects, analyzing the cause and impact of those defects, and then reporting the results of the analysis. This stage is outlined and described below:

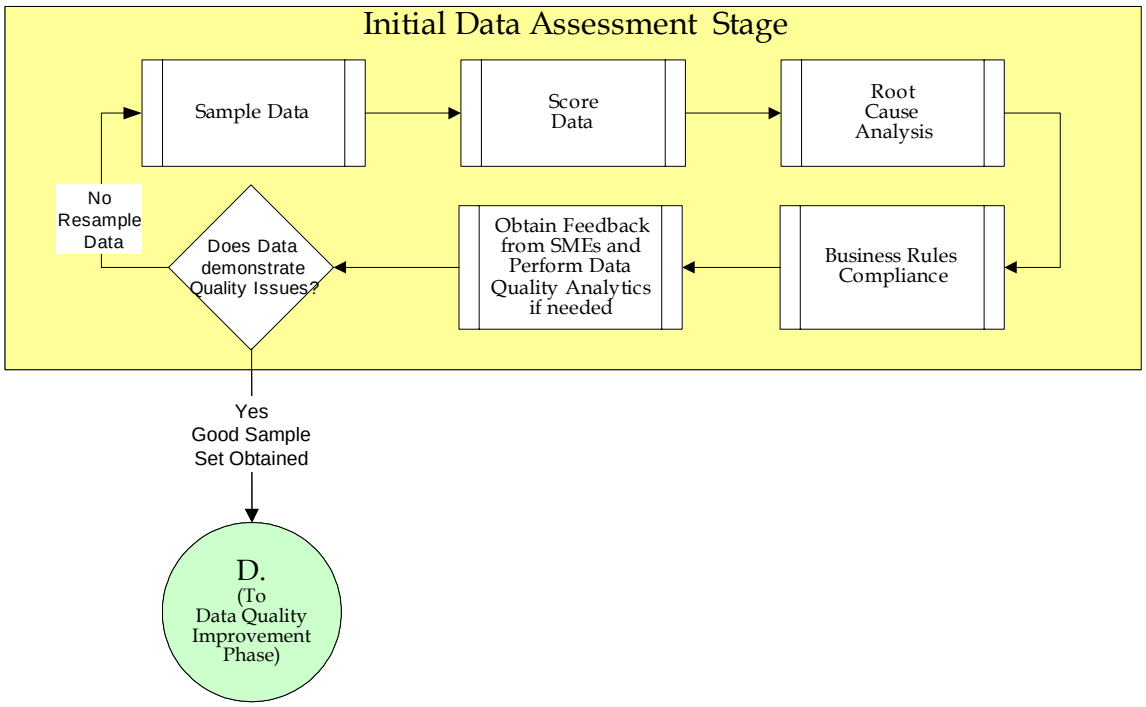


Figure 21 . Assessment Phase: Initial Data Assessment Stage

- **Sample Data –** The System Owner will coordinate with the technical support person (i.e., database administrator, system administrator, or other technician, depending on the type of issue) to create sample sets of data to be assessed for the purpose of highlighting the known data inconsistencies (Note: these sample sets of data may also identify new issues). If the sample set does not contain any of the known defects, the process for selecting sample data should be revised and the Data Assessment Stage repeated.
- **Score Data –** The sample data will be scored against the metrics established in the Pre-Assessment Phase. The Working Group will use the scoring to validate that the sample data

is within the desired quality levels. This is important since the sample data must accurately represent the issue before a root-cause analysis can be conducted. A scoring table similar to the one depicted below may be utilized to facilitate this process. Here, the sample set of data is identified in the left most column and is given a score. This scoring can be automatically generated through the use of a COTS tool or through manual research into the fields of a database. The data is rated against each metric and then weighted based on the acceptance criteria. These scores are added for an overall Data Quality Score. Thresholds are set to determine the acceptable data quality level.

In the example below, sample sets of data have been scored against hypothetical criteria, and a high water mark value of 5 is set as the threshold. Based on this data, Sample C fails as it exceeds the acceptable levels, demonstrating that it contains the quality issue. It will therefore be used for the overall root-cause analysis. In the case that this sample set does not provide a sample with an adequate score for the root-cause analysis, additional samplings of data should be taken or the criteria should be reconsidered.

	Criteria with Weights				Score	
Data Sample Set	Percent of Blank Name Fields	Percent of Duplicate Entries	Percent of Blank Address Fields	Percent of Data with Incorrect Format		**Is Data Quality Acceptable?**
	0.3	0.2	0.2	0.3		
A	0.6	1	0.4	0.6	2	Y
B	0.9	1.6	0.6	2.1	4.3	Y
C	1.8	2.4	1.2	2.7	6.3	N

(Percentage out of total data points exhibiting defect) x (Weight)=(6%) x (0.3)

Overall Data Quality Score must be below 5 to be acceptable

Table 5 Data Quality Scoring Table

- **Root Cause Analysis –** For this step, the Working Group will develop a view of the current state and attempt to determine what factors attribute to the given Data Quality Issue. When conducting a Root Cause Analysis, the possibility of both technical and functional flaws in the business process must be considered. Some examples of technical issues are as follows:
  - **Values Assigned:** Singular reference for each data element between systems. The definition of what an address is should be the same between systems.
  - **Cardinality:** Restricts the number of values allowed for any given field. For instance, though there may be more then one address for a recipient there should be a limit as to how many addresses FSA keeps track of, and there should only be one home address.
  - **Referential Integrity:** Tests for primary and foreign key integrity. Integrity of reference between data in related systems. For example, borrower address should be the same everywhere.
- **Business Rules Compliance –** The current Business Rules must to be verified against the data profile. If need be, additional rules may be added or modified to match any new

processes or data attributes. In addition, the scored data will be verified to be compliant with business rules along with the As-Is process flow.

- **Obtain Feedback from SMEs and perform Data Quality Analytics if Needed** - The data quality assessment will be presented to SMEs to verify the procedures and results. SMEs may provide feedback regarding the need for additional data measurement and analysis. At this point, if the data analysis does not effectively demonstrate the issue and its cause, the Working Group will go back to the sample data process to analyze the data again.

3.3.3.2.1 Data Assessment Business Tools

The main business tools of the Assessment Phase will help to identify functional business flaws in the root cause analysis process. These tools will include the Supplies, Input, Process, Outputs and Customers diagram (SIPOC), Pareto Chart, and Fishbone Diagram.

SIPOC Diagram

The SIPOC Diagram is a tool used for conducting an As-Is analysis of the systems being analyzed. A general SIPOC outline is shown below, followed by a business example specific to FSA. Suppliers can be either internal FSA systems or external trading partners. The Inputs are all data and information being sent by the supplier to the process. The SIPOC diagram is centered on a given process, which can be single or cross-system based. Outputs are the data files and information exiting the processes which are sent to the customers. As with suppliers, customers can range from external trading partners to internal systems.

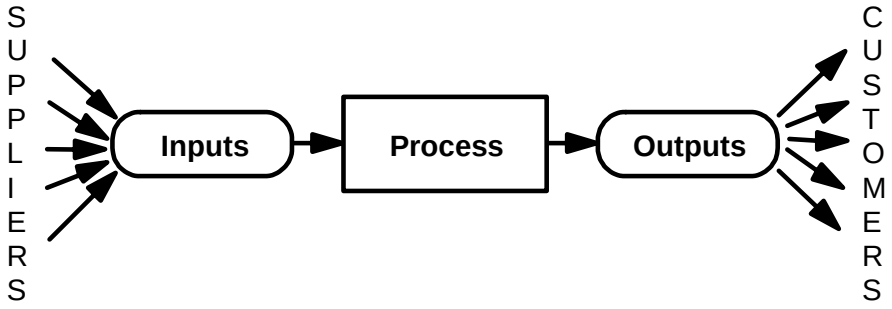


Figure 22 . Template of SIPOC Diagram

Mad Dog Issue 25 states that school data in Default Management Collection System (DMCS) is static, and can be used to illustrate the potential application of the SIPOC diagram. A Working Group, having conducted VOC and CTQ analysis surrounding this issue, would map the processes related to the issue to a SIPOC diagram. Examples A and B show the progression of the breakdown of a business issue into its inputs, processes, and output. In example A, trading partners (e.g., Schools, GAs, etc) transfer defaulted loans to FSA's DMCS for collections. All new defaulted loans in DMCS must have the school code verified or they will be rejected and sent back to the trading partner. Breaking down this process further, example B shows that this verification step encompasses an FSA DMCS representative manually updating a school file on the DMCS system. This DMCS school file is then used for verification purposes. From these SIPOC diagrams, business analysts on the Working Group may raise questions as to why DMCS has a unique school file when PEPS, the steward of school data, supplies FSA's system with a school file that contains the latest school codes. Upon further examination, it is discovered that DMCS does not currently receive the PEPS school file resulting in a small percentage of loans being rejected based on school codes that have not been updated due to the ad hoc manual updating procedures.

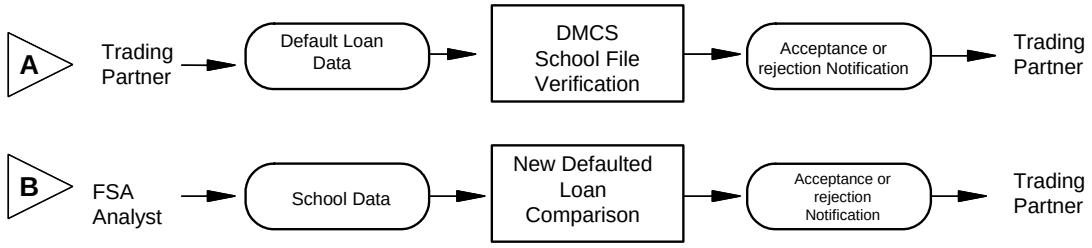


Figure 23 . Example A and B of SIPOC Diagrams

Fishbone Diagram

A Fishbone Diagram can be utilized to identify the root cause of any identified errors. This diagram can prove useful for identifying major causes and sub-issues that may lead to further system investigation and issue identification that may have otherwise been overlooked. As shown in the diagram below, the issue is listed in a box which serves as the "head" of this skeletal diagram. A line is then drawn out to represent the "spine" from which major causes of the single issue branch out as "appendages." Their proximity to the front of the diagram has no bearing on their severity, and so they are arranged in no particular order. Sub-causes, which can exist on various levels, branch out from these major causes. Depending on how detailed of an analysis is needed or the feasibility of identifying details to such an extent, the Fishbone Diagram may branch out to as many sub levels as necessary.

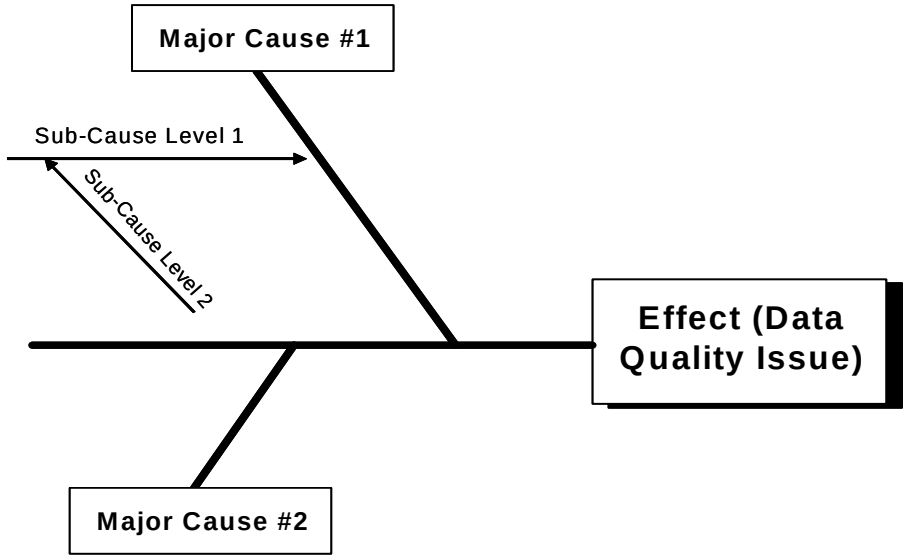


Figure 24 . Fishbone Diagram Template

As an example, one common issue that occurs in most enterprise systems is the rejection of load data into a database. This issue can be profiled using a Fishbone diagram to understand possible root causes. A hypothetical example of how this can be accomplished is shown in the diagram below. Here, the potential major causes of this issue have been mapped out as File Issues, System Issues, and Trading Partner Issues. For each cause, sub-causes branch out. For

example, a System issue may have a system crash as a sub-cause. This sub-cause can then be branched out to a lower level cause, such as hardware component failure.

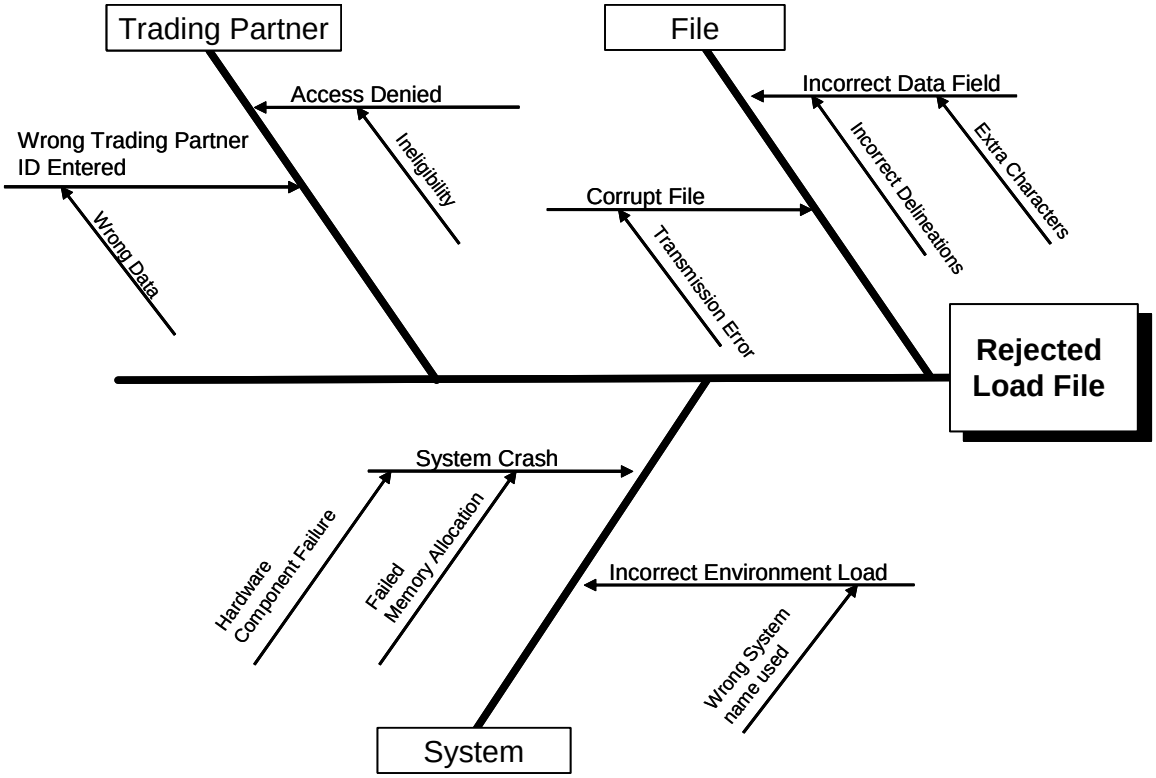


Figure 25 . Fishbone Example Diagram

Pareto Chart

In order to further understand the impact of major sources of errors, sample data can be profiled by use of a Pareto Chart analysis. Pareto Charts can be used to graphically compare the importance of potential root causes by charting their relative frequency and gauging their impact over a period of time. For example, a Working Group investigating rejected default loans within FSA may identify the major causes of rejection via a Fishbone diagram as incorrect school data, incorrect name, technical issues, and other reasons. These causes are then measured for frequency and diagrammed as shown in the hypothetical Pareto Chart below.

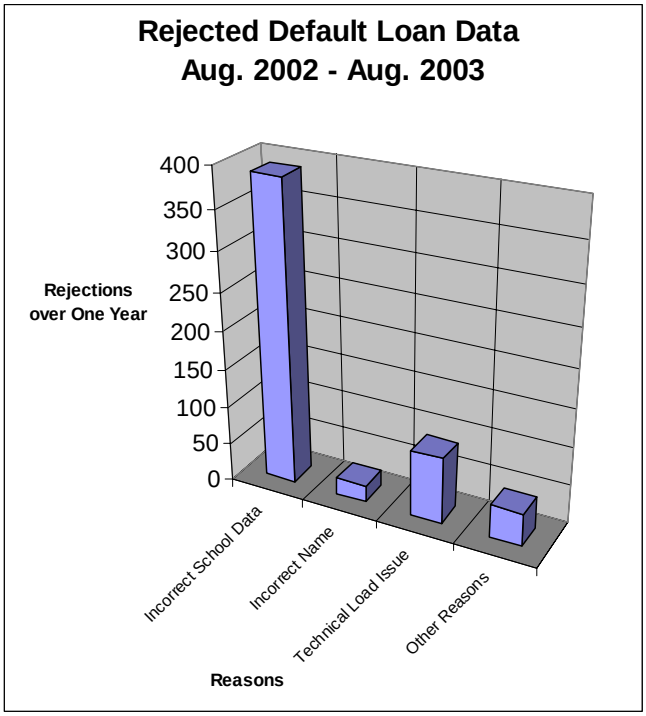


Figure 26 . Pareto Chart

This diagram highlights the fact that the majority of rejections are due to incorrect school data and not due to technical load issues. The Working Group can utilize this information when developing solutions for the issue. A solution which eliminates reasons with the most system impact will achieve a higher reduction of overall errors, or rejections in this case.

3.3.3.3 Pilot Testing Data Assessment Stage

Once the Initial Data Assessment Stage of the Assessment Phase is complete, the Working Group will enter the Data Quality Improvement Phase where solutions are developed and tested for the quality issue (see 3.3.4 - Data Quality Improvement Phase.) In order to properly verify the testing results, the Working Group will loop back to the Assessment Phase for the Pilot Testing Data Assessment.

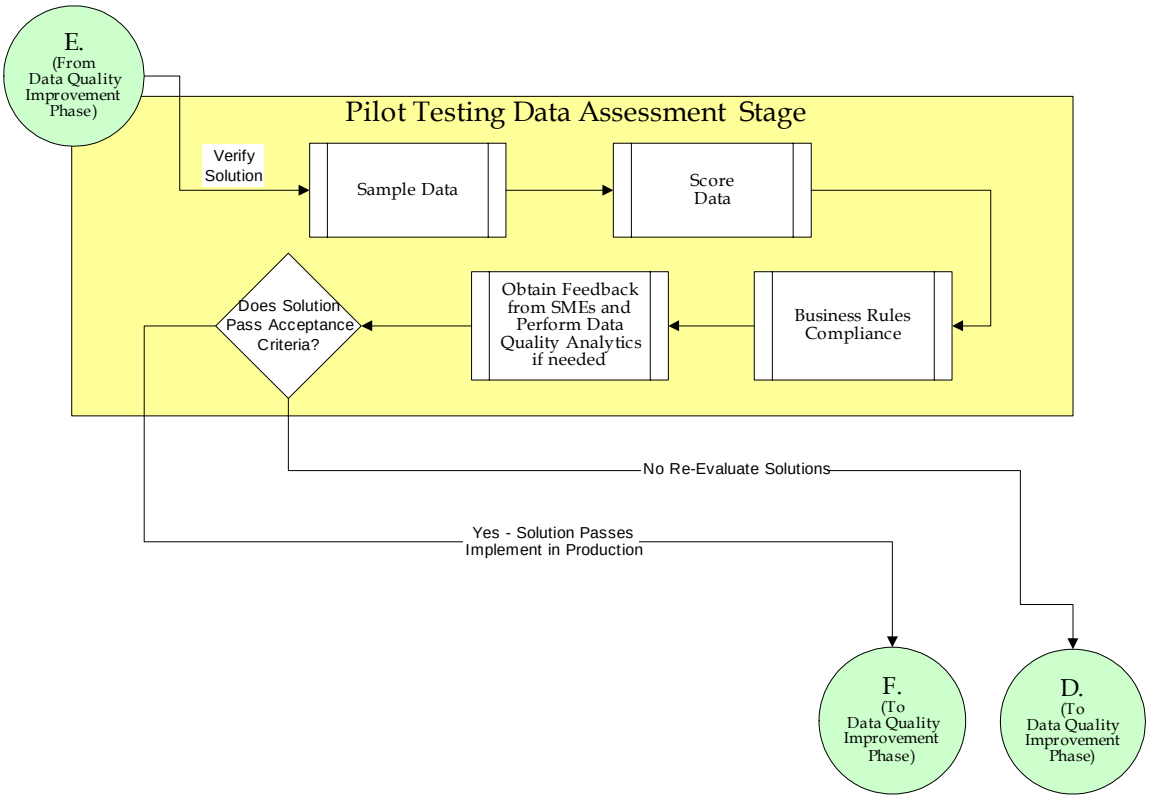


Figure 27 . Assessment Phase: Pilot Testing Data Assessment Stage

The Pilot Testing Data Assessment Stage will verify whether the developed solution improves the data quality to acceptable levels. While the steps in this stage will be similar to those in the Initial Data Assessment Stage with data sampling, data scoring, and results analysis, there are two fundamental differences:

- In the Score Data process, a table with weighted metrics can be used to assess the level of data quality with the new solution in place, and therefore, the Working Group will be interested in obtaining data that exhibits data quality *above* the acceptable threshold. This translates to testing whether the sample data is fixed by the solution and not whether the sample data contains an example of the quality issue.
- For the Pilot Testing Data Assessment there will be no Root Cause Analysis step. Instead, after data is scored, it will be reviewed again for Business Rules Compliance, followed by a SME review to determine effectiveness of the solution.

If the sample data does not meet the acceptable data quality criteria, then the Working Group will return to the Data Quality Improvement Phase's Solution Definition and Assessment Stage to reevaluate the solution and possibly utilize a different option. However, if the pilot test data sample meets criteria, the solution will pass and the Working Group will move on to the Production Implementation Stage.

3.3.4 Data Quality Improvement Phase

Once a sample of data has been identified that demonstrates the quality issue during the Initial Data Assessment, the Improvement Phase will be initiated for solution determination and implementation approach. There will be two process paths in this phase. For new issues or those without proposed solutions, the Working Group must formulate the criteria for solution analysis, develop a solution, and apply the solution to the criteria to determine if it fixes the quality issue. Once a solution is selected, a pilot implementation will be completed on a test environment to verify results. This stage will include looping back into the Assessment Phase for data sampling, scoring, and verification that the test solution is correct. At this point, the second process path of the Improvement Phase will be invoked as the tested solution is implemented in production.

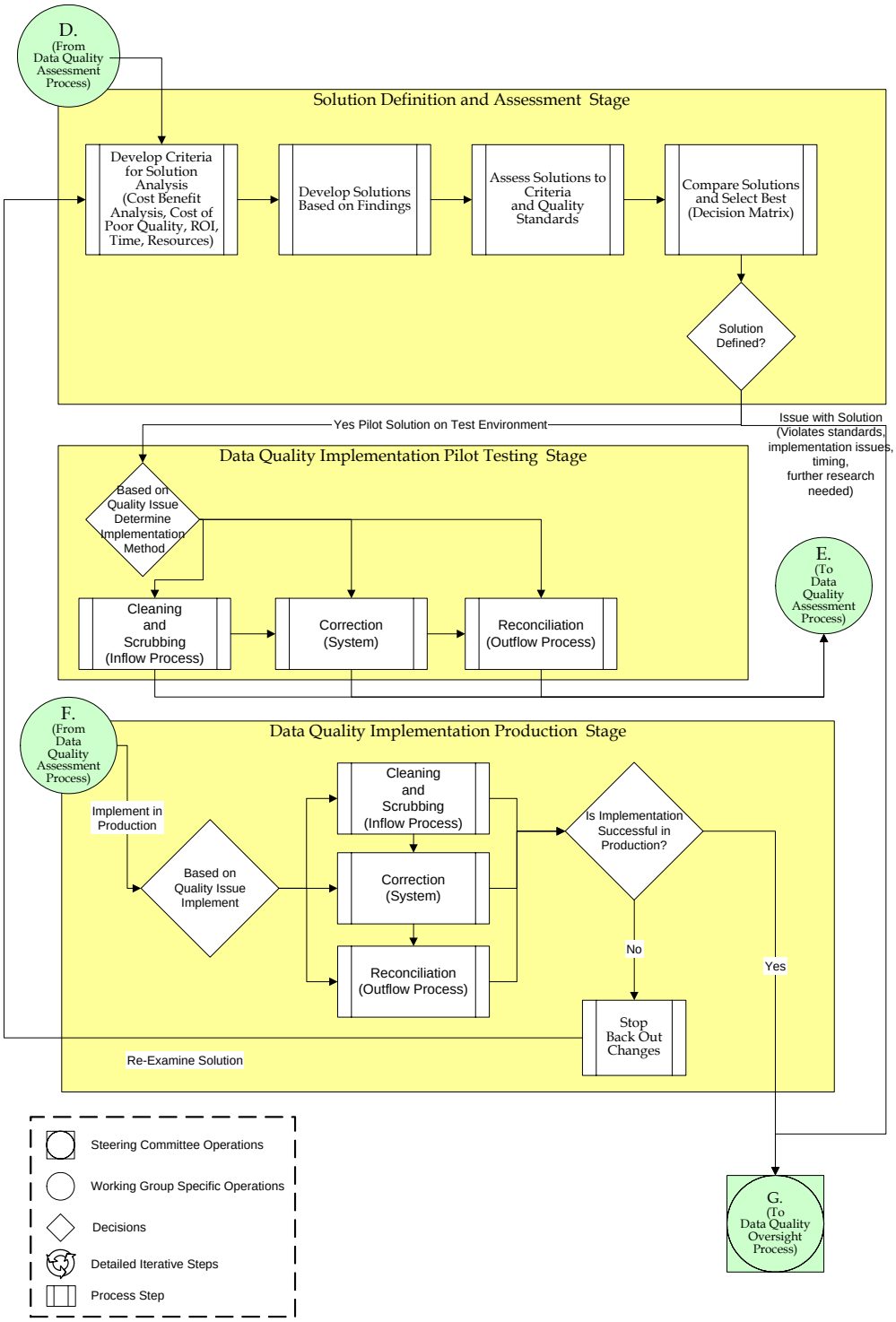


Figure 28 . Data Quality Methodology: Improvement Phase

A variety of business tools will support the Improvement Phase. The Working Groups will perform such activities as brainstorming, creating decision matrixes, conducting cost/benefit analysis, and completing risk assessments. Once a solution is determined, it is important that a change leadership analysis is completed to ensure its alignment with FSA's To-Be vision for its Data Strategy. This tool will provide a framework of the organization's vision, structure, and tasks for verification that the process improvement fits within the FSA enterprise. Once the solution is completed, an implementation plan should be created for pilot testing and ultimately production release. A control and response plan will determine if the process improvements are consistently implemented and what actions to take if they are not. Detailed examples of these tools are discussed in the sections below.

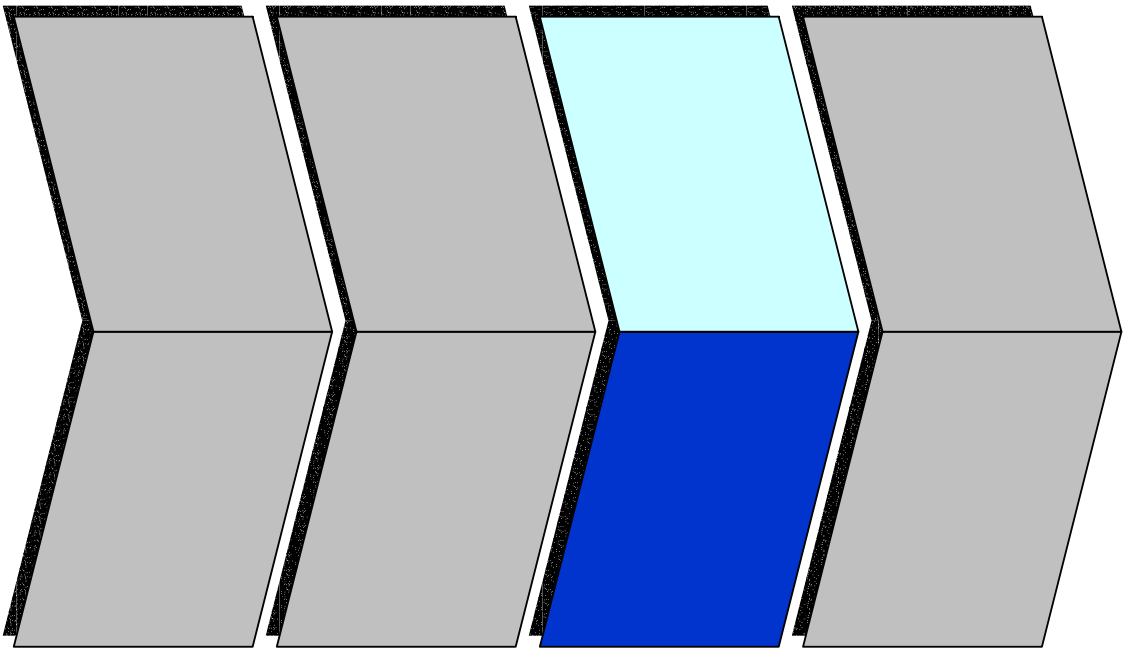


Figure 29 . Business Process Tools Mapping: Improvement Phase

3.3.4.1 Solution Definition and Assessment Stage

The Working Group will conduct the Solution Definition and Assessment Stage by following the processes below:

Prioritization

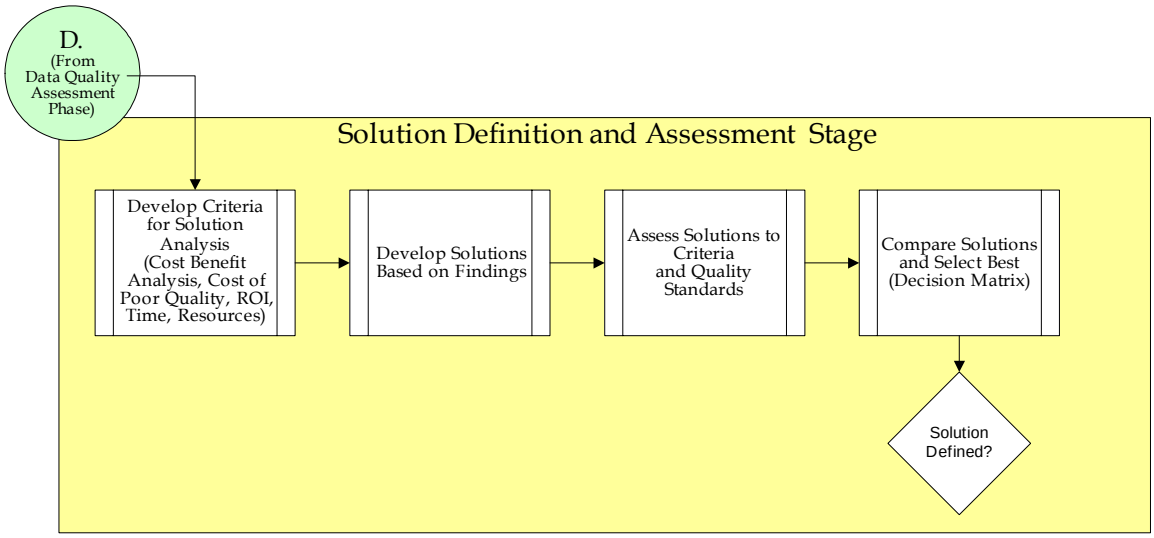


Figure 30 . Improvement Phase: Solution Definition and Assessment Stage

- **Develop Criteria for Solution Analysis –** This process will involve the Working Group developing criteria for assessing the applicability, viability, and cost-effectiveness of all formulated solutions. Guidelines will be developed around Cost-Benefit Analysis, Risk Assessment, and evaluating time/resources required. The solution's impact on the associated systems also will need to be assessed as part of this process. From a financial perspective, a key criterion is a Cost of Poor Quality analysis. As discussed previously, this tool will be used to measure the financial loss experienced by FSA if the issue were to persist unchecked. Performing an overall Cost-Benefit Analysis will include utilizing a COPQ analysis and defining a methodology for evaluating the Return on Investment for each solution.
- **Develop Solutions Based on Findings –** The Working Group can use analytics and findings from the Data Profiling step in the Assessment Phase to brainstorm possible solutions for the problem. All members of the Working Group with their unique background and expertise should participate in this activity. Additional resources may be brought in to provide further business or technical expertise if necessary. It is important for the Working Group to identify how the issue is resolved and where in the As-Is state the correction occurs, specifically, whether data entering the system group, existing within the system group, or exiting the system group needs correction. All documentation about the solution should be captured in a shared repository that contains standardized templates and procedures. This will ensure that all Working Groups create solutions based within FSA guidelines, proper documentation is generated for each solution, and knowledge capital is created.
- **Assess Solutions to Criteria –** The possible solutions will be assessed based on the developed criteria and placed in a decision matrix. Each metric measuring the solution options will be weighted based on the metric's overall importance to data quality, as it was

created in the Data Quality Assessment Phase. In the example Decision Matrix below (see Table 6 - Decision Matrix), the criteria include:

- **ROI –** This will include Cost/Benefit Analysis results taking COPQ into consideration.
- Solution deployment speed – an important consideration when examining long-term versus short-term needs.
- **Level of impact on other systems –** This may determine any need for future systems modifications.
- **Ease of implementation –** Independent of costs, the solution's ease of implementation includes necessary resource availability.
- **Customer Satisfaction –** based on previous VOC and CTQ analysis, the customer weight and satisfaction can be gauged by assigning how much of the customer need may be fulfilled by the solution.
- **Relativity to Business Objectives –** The solution must be compared to how it conforms to FSA Business Objectives. The challenge here is that the relativity to Business Objectives may not always be tangible or easily expressed in monetary terms. For example, quantification of benefits from the perspective of customer ease of use may be particularly difficult when compared to quantifying increased business decision efficiency, yet the objective remains critical to the success of the enterprise. A decision matrix will enable the objectives to be weighed in the Solution Assessment when the Cost of Poor Quality or ROI prove difficult to quantify.

  In order to quantify the assessment of each solution option, members of the Working Group will vote on a solution's ranking relative to other options based on how well they meet the established criteria. The sum of these votes will then be multiplied by the weight of the metric to determine a score for the solution option in the given criteria. Next, the scores from all of the criteria will be summed to determine the solution option's overall score.

Solution Option	Criteria with Weights					Overal Score
	ROI	Deployment Speed	System Impact	Ease of Implementation	Customer Satisfaction	
	0.3	0.2	0.1	0.3	0.1	
A	2.4	1.6	0.8	2.4	0.8	5.6
B	3	2	1	3	1	7
C	1.8	1.2	0.6	1.8	0.6	4.2

(Sum of voted ranks) x weight of metric [ex. (3+1+1+1)x0.3]

Highest Score = Best Overall Option

Table 6 Decision Matrix

- **Compare Solutions and Determine Best Option –** After assessing each solution based on the established criteria, the Working Group will either choose the solution option that best matches the set criteria (receives that highest rating or score from the assessment process) or

document that a solution is not available at this time. Reasons why a solution may not be able to be created range from violation of standards, implementation issues, and conflict with system release schedules, to further research being needed. A status will then be sent to the Oversight Phase to update the Quality Report with the issue resolution. Issues for which a solution is developed and validated, the Working Group will progress to the Pilot Testing Stage.

3.3.4.2 Pilot Testing Stage

Once a solution has been defined and selected as the best overall option, the Working Group will perform a pilot test to determine the viability of the recommendation (Note: Standards for data exchange, edit checks, and usage are created during the Assessment Phase). Depending on the solution methodology selected to resolve the data quality issue, the pilot test may involve cleaning and scrubbing data coming into the systems (i.e., Inflow Process), correcting data within the systems, or reconciling data flowing out of the systems (i.e., Outflow Process). During pilot testing, the Working Group should be careful to properly document the impact of the solution on the enterprise (e.g., if the quality of data in the inflow process is improved, do the DBAs of the systems affected understand the process change and how it could impact their databases? Do any external flows need to be validated?). Additionally, the Working Group will need to create testing standards, procedures, templates, and management guidelines used during testing (Refer to the *FSA System Integration and Testing Process Handbook*). As a first step to this process, it is important to develop test plans that document all activities required to conduct thorough and accurate tests of system parameters, customizations, interface modules, and business processes. These test plans will create the following types of test scripts to thoroughly ensure that all project requirements and objectives are met: [20]

- ***Unit Test Scripts*** – Unit test scripts are based on knowledge of how the logical unit is designed to work and can be written once the detailed design specifications are completed. Each script includes tests for field ranges, values and lengths, each function, data validation, data dependencies, and special processing contained in the module. The unit test scripts should be updated as changes occur to the requirements and/or the detailed design specifications either before coding starts or while it is in process. The developer responsible for the module will run the unit test for that module. At the end of the unit test, all errors related to the independent operation of the program should be found.
- ***Integration Test Scripts*** – While the unit test scripts focus on the smallest logical unit or module of code, the integration test scripts focus on how multiple components work together and communicate. The integration test scripts describe test cases for the interfaces and interactions between system components as they are put together to form progressively larger subsystems. Integration test scripts also test the screens together as a cohesive information function, and then are built upon to test how the information function integrates with other functions in the system. By the end of integration test, all functions will be integrated into the system and all internal interfaces will be tested.

[20] All Test Script descriptions are quoted from the "Solution Life Cycle Process Guide" - Solution Life Cycle - 5 - Construction Phase V2.0.doc

- ***System Test Scripts*** – The system test scripts expand on the high-level test scenarios and scripts developed during high-level design. The testers can reuse relevant test scenarios and conditions from the unit and integration test scripts.
- ***Performance Test Scripts*** – The performance test scripts expand on the high-level test scenarios and scripts developed during high-level design. These scripts detail the steps necessary to fully ensure that the deployed solution will be able to handle the expected workload and adequately address the business objectives stated in the Business Case and Business Performance Targets stated in the Business Performance Model.
- ***User Acceptance Test Scripts*** – User acceptance test scripts expand on the high-level test scenarios and scripts developed during high-level design. User acceptance test scripts test the requirements from a user perspective. They include enough test conditions to determine if the application meets the user acceptance criteria.

Once these tests have been successfully executed, the Test Results should be documented in the Quality Report clearly indicating the testing procedures, data used during the testing processes and results demonstrating that the solution meets the business objectives as stated in the Business Case for the quality issue. The solution methodologies for which these test scripts are developed are in the diagram below, followed by a detailed discussion of each process:

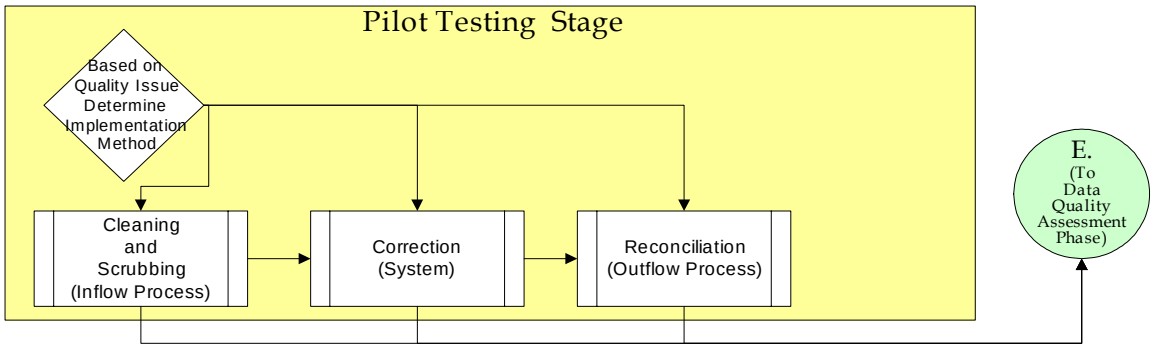


Figure 31 . Improvement Phase: Pilot Testing Stage

- **Cleaning and Scrubbing (Inflow Process)** – Data scrubbing specifically refers to either eliminating or repairing corrupt, incomplete, incorrect, or duplicate data entering the system. This process may include one or more of the following:
  - **Analyze:** Assesses the nature and cause of data defects
  - **Measure:** Quantifies the number and types of defects
  - **Parse:** Identifies and isolates data elements in data structures
  - **Standardize:** Normalize data values and formats according to customer defined rules
  - **Correct:** Verifies, scrubs, and appends database on SQL scripts or third party products
  - **Enhance:** Appends additional data increasing the value of the information
  - **Match:** Identifies duplicate records within multiple tables or databases
  - **Consolidate:** Combines unique data elements from matched records into a single source
  - **Report:** Provides reporting with the data quality process

If needed, these activities can be either manually corrected by a database administrator or through available COTS tools. It is important that data entering a system or group of systems is consistent in format and adheres to set standards. The XML Registry and Repository (Deliverable 123.1.16) proposes to supply a central access point for a consistent set of data standards. This repository could be utilized by the Oversight Phase to identify common definitions to eliminate data discrepancies between systems.

- **Correction (System) –** The data correction process will involve applying the solution internally to the system or group of systems (e.g., transfer process changes or data storing procedures). COTS tools have the ability to provide Data Correction Services which serve as mechanisms for making initial corrections and ensuring that data remains correct until a series of processes, both automated and manual, are adopted to keep data aligned across the enterprise (A list of COTS tools and their capabilities appears in Appendix F).
- **Reconciliation (Outflow Process) –** For data that is transferred to other systems and has been corrected from the Cleaning and Scrubbing Stage, Correction Stage, or as part of a solution to a quality issue, changes to the outflow transmission process may be needed. It is important to monitor the outflow process, as data exiting a system directly affects the data quality of other enterprise systems and/or external trading partners. In this step, the Working Group will contact other system owners to communicate the quality issue, solution, and potential impact on their systems. The systems owners will coordinate with the Working Group to make sure that any modifications necessary to their systems are coordinated with the implementation schedule.

Once the solution has been defined and the inflow, correction, and outflow processes have been verified, the pilot test solution will be passed back up to the Data Quality Assessment Phase for final data sampling and scoring. With an acceptable score, the test solution will flow back down to the Improvement Phase for the Production Implementation Stage.

3.3.4.3 Production Implementation Stage

After a tested and viable solution is established, the Working Group will implement the solution into production with all associated parties. The actual cleaning and scrubbing, system corrections, and reconciliation processes will be duplicated from the Pilot Testing Stage into the Production Implementation Plan. If the solution is implemented but proves unsuccessful in production, the Working Group must discontinue production and back out changes to the systems. The issue will then be returned back to the Solution Definition and Assessment Phase to select and test another solution.

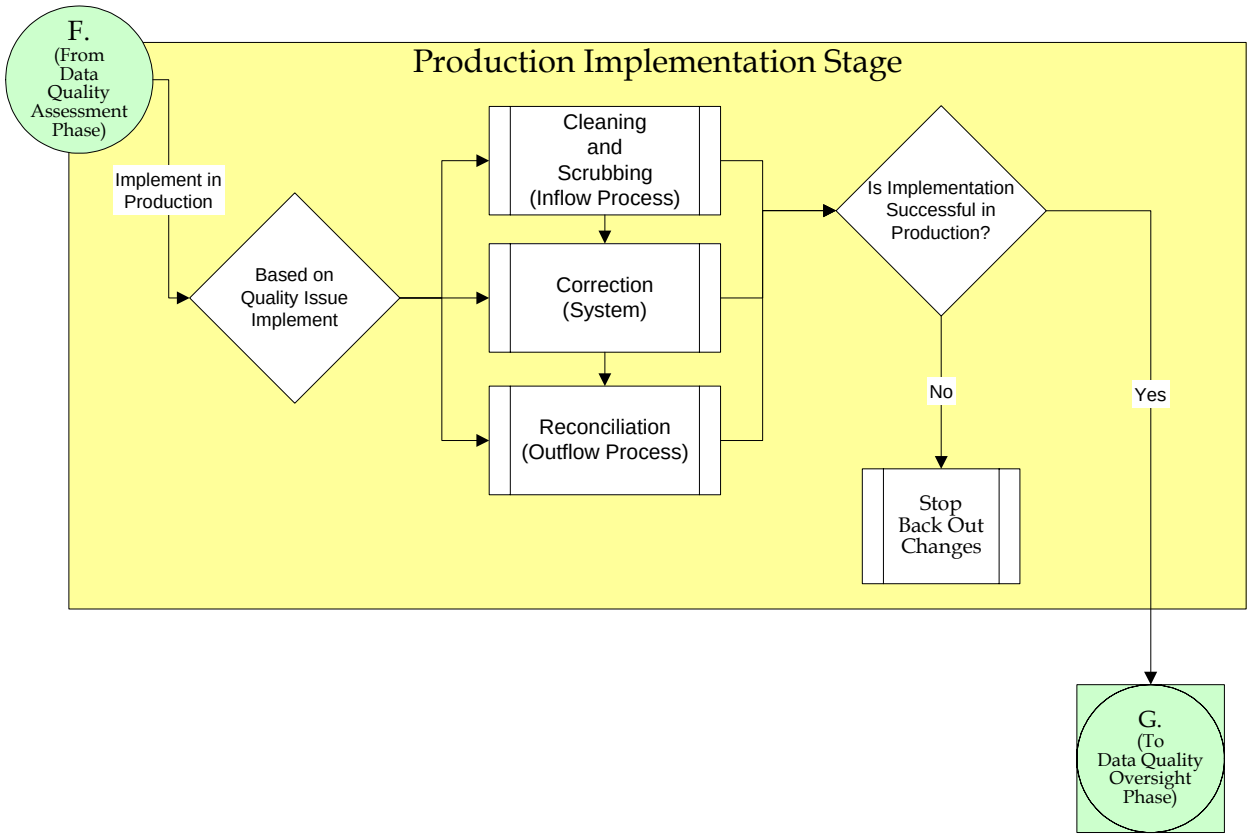


Figure 32 . Improvement Phase: Production Implementation Stage

If the implementation is successful in production, the Working Group project manager will report the issue analysis, resolution methodology, and solution results to Enterprise Steering Committee analysts in the Oversight Phase. The analysts will maintain a record of the issue within the Quality Report, which provides tracking capabilities to show the Working Group's progression and ultimate resolution of the data quality issue. As described in the Assessment and Improvement Phases above, this central database can serve other Working Groups as a reference for future issues.

3.3.4.3.1 Improvement Phase Business Tools

After a solution has been identified and tested, it will be crucial that the Change Leadership Tool is utilized to provide a framework to help align the new process within the organizations strategy, culture, structure, and management practices. If the solution cannot be aligned within the organizations culture, the solution may ultimately fail. This tool will provide the proposed solutions business case, financial impacts, and system impacts to FSA leadership in a consolidated format for final review.

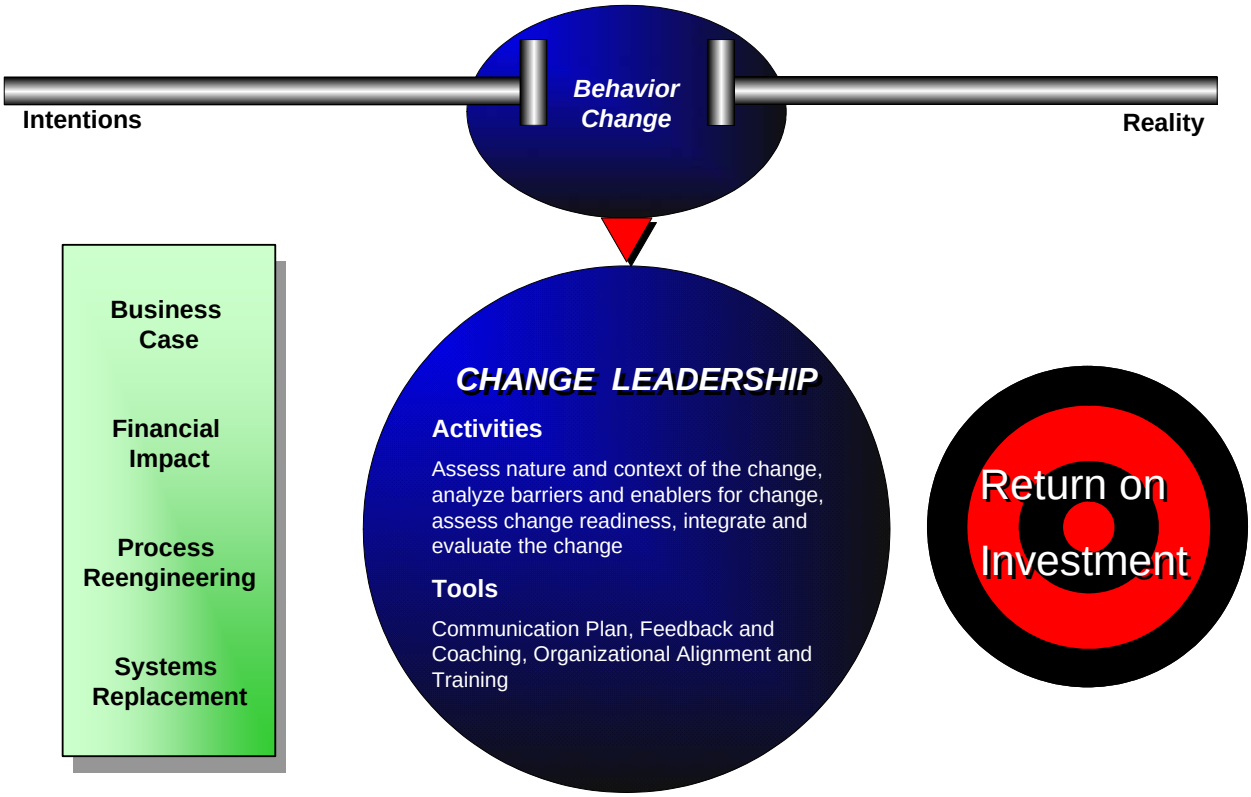


Figure 33 . Improvement Phase: Change Leadership Process

Actions within this process will include taking action on those elements out of misalignment, contacting all key stakeholders, and revising the solution as needed. By achieving alignment, FSA will achieve ownership and commitment to the new process.

Once a solution is properly aligned and tested for production release, a Control and Response Plan will identify the process steps necessary for implementation, the person responsible for each step, and inputs/outputs for each step. Unlike a typical implementation plan, this document will list the control mechanisms in place to ensure sustained process performance and will identify the actions necessary when the process improvements are not maintained. In Table 7 - Improvement Phase: Control and Response Plan, an example of a quality issue for incorrect applicant information being loaded into the CPS database is depicted

Control and Response Plan							
Quality Issue:	#15 - FAFSA mis-loads		**Core Team:**	DQ Working Group #3		**Date (Orig):**	10/31/2003
Key Contact:	Mike Brown		**Phone:**	xxx-xxx-xxxx		**Davt (Rev):**	
Process Step	**Resp.**	**Output**	**Input**	**Amount of Data**	**Frequency**	**Control Method**	**Response Plan**
Load FAFSA data	John Smith	Success	Applicant Data	50,000 rows	Daily	Field length validation, first and last name population.	Contact FAFSA technical analyst, verify system loads, etc

Table 7 Improvement Phase: Control and Response Plan

A Control and Response Plan is to be used in conjunction with an Implementation Plan to provide the details behind each activity in the implementation plan. Results from the Production Implementation Phase will flow into the Oversight Phase which tracks issue resolution and provides further analysis if needed.

3.3.5 Data Quality Oversight Phase

While the Data Quality Oversight Phase will occur at the end of the methodology, it is one of the most vital as it will continually maintain and monitor enterprise-wide data quality levels. This phase will serve a threefold purpose of scheduled monitoring for quality improvement and maintenance, ad-hoc report generation for the Steering Committee, and verification of Working Group solutions. These activities will be summarized by a Data Quality Report application which provides high-level status on quality issues, while allowing for drill down capabilities to specific issue resolution. The Oversight Phase will be conducted by the Steering Committee in conjunction with business and technical analysts and can be initiated in four distinct ways.

- The Steering Committee may desire specific inquiry into a particular issue, such as historical trends or additional metrics. An Oversight Analyst would receive this request, perform analytics or audits based on the request, potentially leveraging COTS tools, and store the results in the Data Quality Report application. The Quality Report would allow a Steering Committee member to view results in a standardized report.
- In addition to ad hoc requests from the Steering Committee or Business Owners, the Oversight Phase will have automatic initiation of scheduled reviews and maintenance of data quality throughout the enterprise. For this stage, COTS tools may be utilized to perform cross program analytics and routine financial audits. An Oversight Analyst would monitor the Quality Report to check systems' statuses and identify any issues that may arise. Based on the severity of the issue, they may proceed to rectify the problem based on the Steering Committee's discretion.
- The Oversight Phase can also be initiated after the Working Group delivers its issue results and solution methodology to the Steering Committee. The Oversight Analysts will monitor the enterprise-wide impact of the implemented solution and verify that there are no adverse effects from the system or data modifications.
- Lastly, representatives of the FSA Quality Insurance Program can update the Quality Report with results from PIR initiatives. Once a quality solution has been implemented as part of a development effort, all outcomes and information related to the original issue should be sent to the Steering Committee and updated to the Quality Report by the Oversight Analysts.

The Oversight Phase will have three stages (Enterprise Analytics, Enterprise Audits, and Enterprise Quality Report), which are independently utilized depending on the initiation request. These stages are outlined in the following process flow diagram and are explored in further detail later in this section.

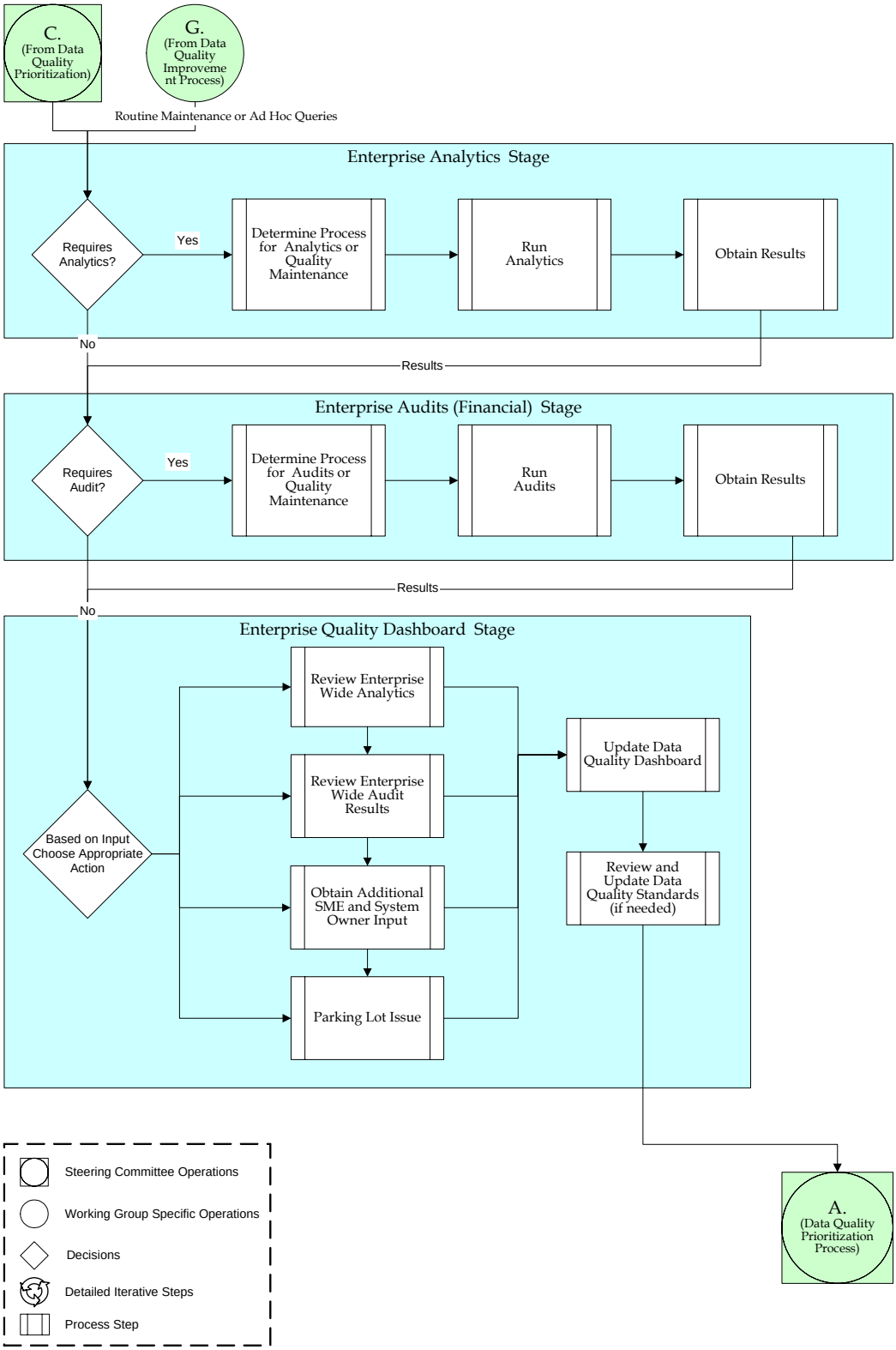


Figure 34 . Data Quality Methodology: Oversight Phase

The processes in the Data Quality Oversight Phase will be enabled by diagnostic tools and the Quality Report. Diagnostic tools may vary based on system applications and availability. A future consideration for the Quality Assurance Strategy is to perform a data quality tool selection. A good enterprise-wide analytical tool can not only fix quality issues, but assist in documentation, capture As-Is and To-Be processes, and perform data conversions. A key requirement for this tool will be its ability to provide a Quality Report. From a conceptual standpoint, a Quality Report will provide a central repository of information regarding issues and solutions and will allow the Steering Committee to effectively oversee all Data Quality initiatives. Using the Quality Report's various reports and subcomponent tools, the Steering Committee will ensure there is sustained quality improvement. The Quality Report and other Oversight tools are further detailed in the sections below.

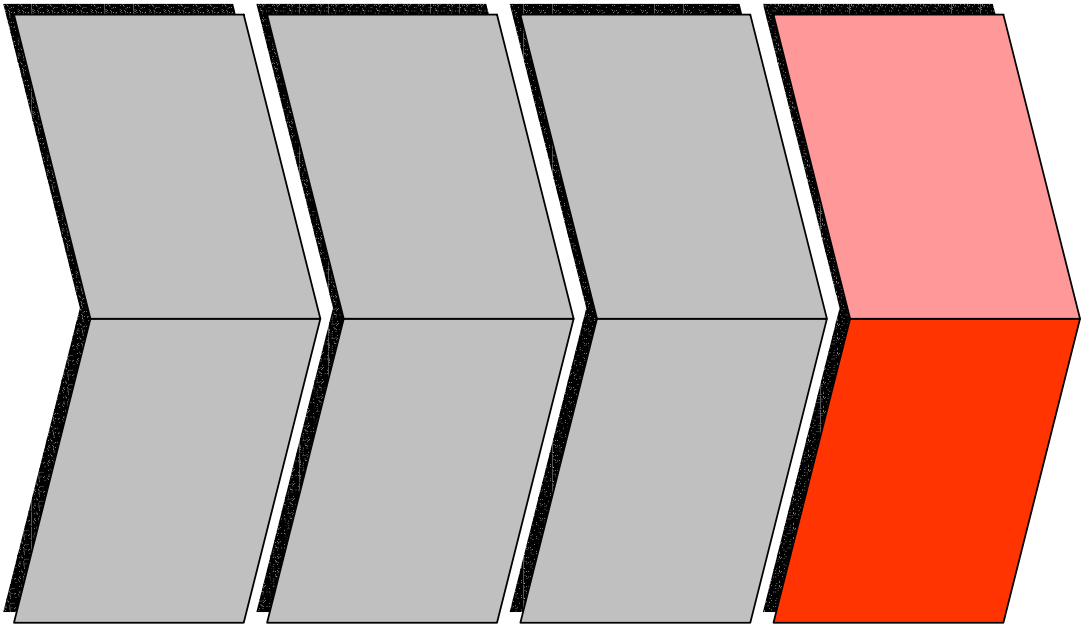


Figure 35 . Business Process Tools Mapping: Oversight Phase

3.3.5.1 Enterprise Analytics Stage

It is critical that FSA manage data at a detailed system level. However, a high-level enterprise-wide view must be considered to prevent analytics at the enterprise level (FSA Executive Quality Report) from producing flawed or inaccurate results. Depending on how the Oversight Phase is initiated, the purpose of the Enterprise Analytics Stage can be to analyze individual systems or the entire enterprise for locating and studying issues (i.e., errors in data) or to determine the systems impacted after a Working Group solution has been implemented. The Enterprise Analytics Stage will encompass three processes which are shown in the diagram and discussed below.

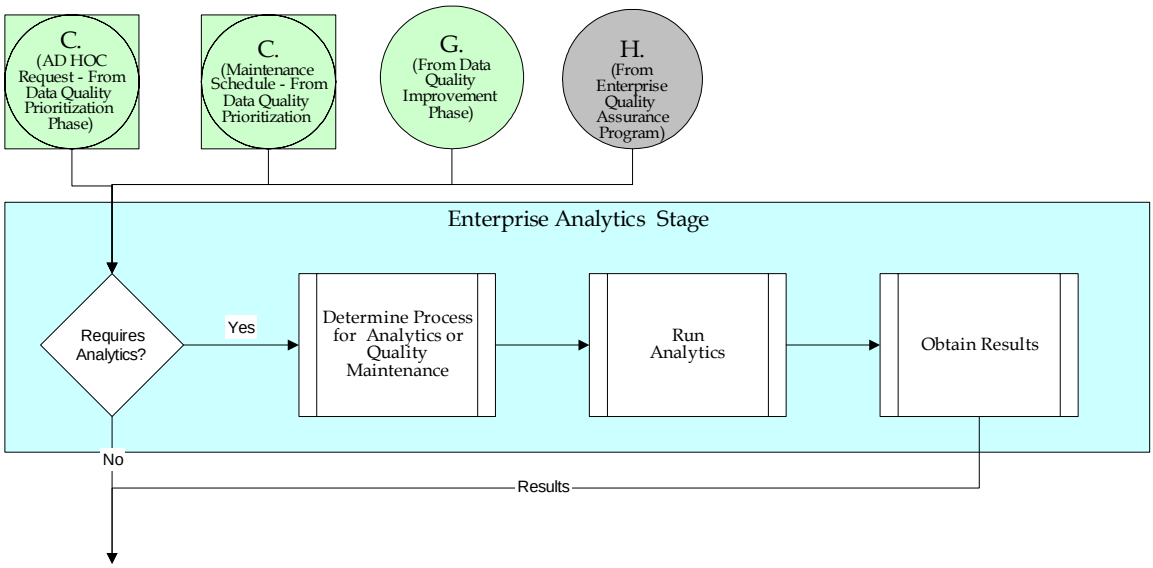


Figure 36 . Oversight Phase: Enterprise Analytics Stage

- **Determine Process for Analytics or Quality Maintenance –** Depending on the issue at hand, an appropriate COTS tool or analysis process must be selected for the systems being analyzed. Available system specific or enterprise tools should be assessed to determine desired analytic functionality for routine maintenance purposes or specific issue analysis. Business Owners, in conjunction with Oversight Analysts, will select capabilities within a single suite of tools or chose among multiple tools. Examples of the types of analytics that will be most useful for this process step include Data Mining, Trend Analysis and Forecasting.
- **Run Analytics –** Once the selection process is complete, the selected analytic tool or process will be applied to the systems involved with the issue being studied. For maintenance purposes, analytics will be run on a scheduled basis to ensure that solutions are having the intended effect and improvements have been sustained.
- **Obtain Results –** After the analytics have been run, results will be collected, reviewed for completeness, and then made available for further review in the Enterprise Quality Report Stage.

3.3.5.1.1 Analytic Tools

Enterprise Analytics will rely on software which provides statistical analysis tools and procedures. Although there are system specific tools for data quality analysis within FSA, there currently is no enterprise-wide tool that is designed for cross program analytics. Additionally, there is no centralized management of the tools that exist within FSA in order to provide tool capabilities, uses, operating procedures, and documentation. A tools analysis should be completed as a future consideration of the Quality Assurance Strategy to determine the applicability of COTS tools to FSA quality strategy and if a COTS tool can act as a centralized repository for the management of FSA system specific analytic tools. As an example, there are a

number of COTS tools in the market place that can aid in an enterprise-wide data analysis. The following are a sample of such tools with a detailed list outlining full capabilities located in Appendix F.

- **Trillium Software:** This is a suite of products which can provide capabilities such as data cleansing, standardization, and relationship matching. It has the following characteristics:
  - It can identify and match customers across multiple interaction points within an enterprise, including e-commerce applications, sales and service applications, order entry systems, call centers, help desks, and marketing campaigns to ensure a complete and accurate view of the customer.
  - This tool can verify that transactions are standardized and matched to existing data, in real time, which allows for verification of correct data and determines whether the customer needs a new account or already has an existing account.
  - This suite of products also integrates data from mainframe, relational, and other disparate sources to eliminate duplicate entities in the database.
- **Firstlogic**: This company's Information Quality Suite provides matching and consolidation technology which identifies duplicate records, consolidates data, builds relationships, and creates a single customer view across databases. It also has the following capabilities:
  - The data quality software standardizes data elements including customer information and email addresses.
  - It corrects data based on secondary data sources to improve matching and can append additional information such as phone numbers.
  - This software has the capability to identify data quality problems and business rule violations within a database.
- **Acxiom:** The Opticx application provides a statistical assessment of a customer database against 40-plus measures of database quality (e.g., Duplicate or deceased records, misspellings, undeliverable or out of date addresses, etc.). It can also display the results of any data analysis to a statistic called the Acxiom Data Quality Index (ADQI) score. The ADQI score is in relation to the average industry score, thus providing a relative quality of the customer data. ADQI scores measured over time can be used as a simple metric to track the progress of data quality improvement efforts.

3.3.5.2 Enterprise Audits Stage

Enterprise audits will enable cross-system financial checks to validate that financial data aligns across FSA systems. The Steering Committee will run the processes in this stage at their discretion based on how oversight has been initiated and the current data quality issue that is being examined. As seen in the diagram below, the processes for this stage will be exactly the same as the Analytics, with the exception that financials will be the only data element being examined.

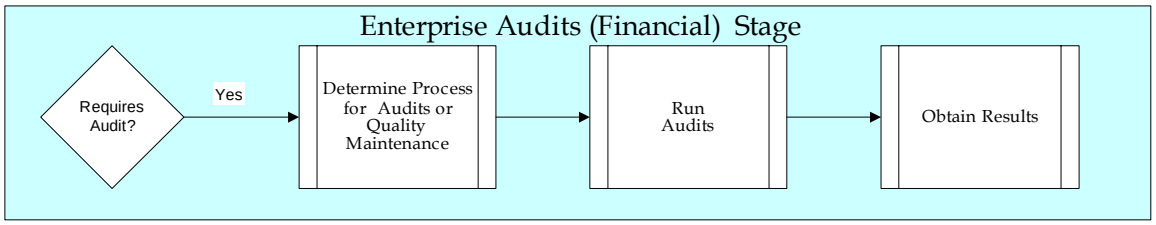


Figure 37 . Oversight Phase: Enterprise Audits Stage

- **Determine Process for Audit or Quality Maintenance –** The necessary process or tool for running the financial audit will be determined in this process. Depending upon the breadth of the audit, either a manual or COTS tool may be used.
- **Run Audit –** After a desired tool or process is selected, the audit will be conducted on the affected systems. If an issue is discovered, further manual checks can be made to ascertain specific financial discrepancies between data elements. An example of a cross-system audit check is confirming that FMS issued accurate payments based on detailed data from NSLDS or another discrete system. Here the utilization of a COTS tools could potentially detect financial discrepancies between multiple systems and files proving effective tracing capabilities. A tool could provide FSA with financial information at different levels of summarization, from an aggregated total to detailed individual transactions.
- **Obtain Results –** Results generated in the Audit Stage will be stored within the Quality Report and will be kept as part of a historical record of the issue being examined. This data should provide guidance as to the impact the issues have on the enterprise, thus allowing the Steering Committee to determine the effectiveness of a Working Group solution or for detecting any overlooked data discrepancies within the enterprise. Additionally, results will demonstrate whether improvement implementations are operating as expected.

3.3.5.3 Enterprise Quality Report Stage

The Quality Report will be the ultimate repository for all data quality related issues and analysis. Since the Quality Report will be used as an enterprise-wide reference tool, it is crucial for the Steering Committee and Oversight Analysts to maintain accurate and updated data. All results from the Analytics and Audit Phase, along with feedback from Working Groups surrounding developing issues, need to be updated into the Quality Report. This Quality Report will serve as a common communication vehicle across the enterprise and care must be taken to update information as accurately as possible and to ensure all new and current issues are clearly stated. In this stage, the maintenance of the Quality Report will be conducted based on how the Oversight Phase was initiated. Updates to the Quality Report could be made through any of the processes outlined below:

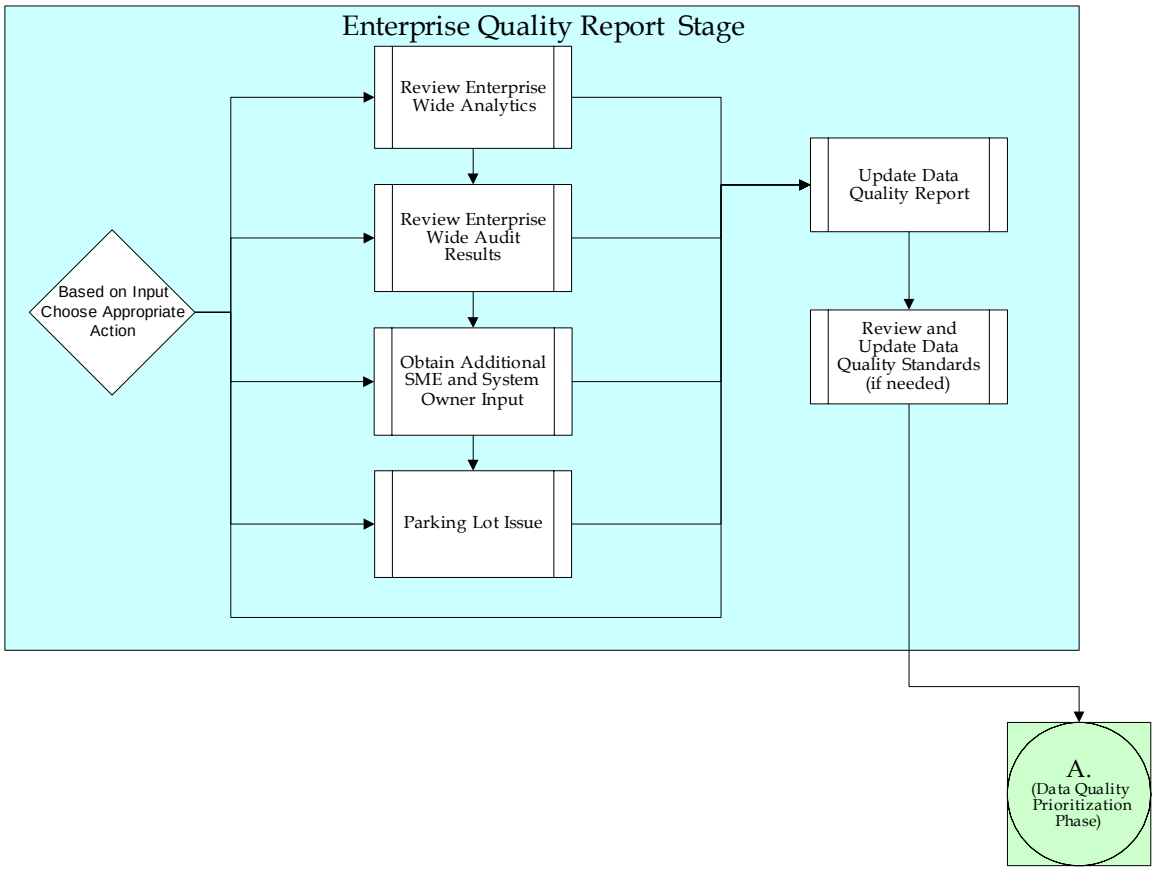


Figure 38 . Oversight Phase: Enterprise Quality Report Stage

- **Enterprise-wide Analytics Results:** After performing analytics, the Oversight Analysts will review the results to assess whether any new issues or problems have arisen as a result of a Working Group solution or during a routine check.
- **Enterprise-wide Audit Results:** For routine maintenance or for specific issue investigation, Oversight Analysts will need to examine auditing data to make certain financial discrepancies do not exist. This status will then be recorded in the Quality Report as it relates to the issue(s) or activity.
- **Additional SME and System Owner Input:** At times issues may require further research requiring SME, System Owner, or trading partner input. Additionally, trading partners may be contacted or formal surveys provided to them in order to ensure that business needs are being met and quality issues are not overlooked.
- **Parking Lot Issue:** If an issue being analyzed is determined to be currently irresolvable or to have low or minimal impact on the enterprise, it can be placed in a "parking lot" section of the Quality Report for future review by the Steering Committee.
- **Update Data Quality Standards:** Existing business rules, objectives and standards may need to be updated as a result of the Data Quality analysis (specifically from the Working Groups Data Quality Assessment Stages). For instance, if quality issue results in a new data definition zip code, a resulting update would have to be made to FSA XML Core

Component repository so that all data within FSA could align with the new standard. Additionally, business objectives should be updated as quality issues that affect or achieve those objectives are solved.

The outputs from the Quality Report Stage will be passed to the Data Quality Prioritization Phase where the Steering Committee will review all updates and changes within the Quality Report and begin a new cycle of the Data Quality Methodology.

3.3.5.3.1 Enterprise Quality Report Tool

The Quality Report will be a vital tool utilized throughout the Data Quality Assurance Strategy, especially during the Data Quality Oversight Phase. The Quality Report will assist both the Working Group and Steering Committee. During the Prioritization Phase, the Steering Committee will be able to access the Quality Report to review a list of prioritized issues, issues requiring further study, and the history of an issue with associated details including the impact on the enterprise. Working Groups will consult the Quality Report when assessing an issue to verify other instances of the issue within the same systems or other systems, and thus obtain a framework for possible analysis techniques and solution development. Within the Oversight Phase, all results from Audits and Analytics will be updated to the Quality Report and inputs from all of the other phases of the Quality Strategy will be monitored. The Quality Report will serve as a tracking, project management, and alert tool. The Quality Report also can aid the Steering Committee with setting cross line-of-business strategies and standards for operational and business intelligence applications and provide a means for continually measuring and reporting on data quality metrics.

The Quality Report will consist of a front end Graphical User Interface (GUI) used for navigating through different views and creating queries and a back-end database which stores all enterprise-wide data quality related information. In deciding upon a suitable Quality Report format, FSA must consider the specific data requirements of all intended users. A future consideration for this strategy is to perform a high level Quality Report requirements analysis and tools selection. Conceptually, some of the properties and attributes that a Quality Report should be able to exhibit are as follows:

- **Non-Prioritized Issues**: Occasionally, new issues may be uncovered by the various Working Groups which have not yet been prioritized by the Steering Committee. These new and pending issues will be maintained on the Quality Report for future prioritization and possible resolution. The issues may be viewed in a consolidated list for the entire enterprise, or broken down to the system level. The Quality Report may also alert the Oversight Committee to an issue that is automatically detected by any COTS tools monitoring data within systems.
- **Prioritized Issues List:** Issues that have been prioritized by the Steering Committee will be kept in the Quality Report along with details, such as the history and current status of the Working Group or IV&V effort involved. This information will allow the Quality Report to be used as tracking or project management tool, especially for Business Owners on the Steering Committee.

- **Issue Resolution History:** Historical data surrounding all issues identified by the Steering Committee regardless of successful resolution will be archived within the database. Details for each issue regarding all work conducted by the Working Group and all results from any necessary Enterprise Analytics and Audits performed in the Oversight Phase will be recorded. Details about issue solutions, how the solution operates, and the data collected throughout the process will be maintained for future reference.
- **Standardized Templates**: All templates utilized for issue analysis and business communication can be housed within the database for reference by Working Groups and the Steering Committee. This will provide standardized documentation and processes for the quality initiatives.
- **Business Standards**: All business standards may be placed into the Quality Report for reference. Since the Steering Committee and Working Group will utilize these, a central location such as the Quality Report ensures any changes are instantly available to personnel.
- **Best Practices:** Once the new FSA Quality Assurance Strategy has been successfully implemented for a period of time, the Quality Report could become a repository for best practices throughout the enterprise. Analysts could review the issue resolution history to identify issue/resolution trends, processes that worked best, or any non-conventional and innovative techniques which proved effective.

Once the requirements have been defined, the type of Quality Report can then be determined. Many Data Quality COTS tools contain an integrated Quality Report component. For example, Trillium and MicroStrategy each have an Executive Quality Report built into their tool. While the reporting tools contained within COTS packages are out of the box and can be used without additional modifications, the unique nature of the data within FSA may require additional customization of a ready-made solution. The extent of this customization will depend on the software being used.

If COTS tools are unable to provide the required functionality, a custom developed Quality Report is another viable solution for the enterprise. A customized solution would allow the Quality Report to meet all of FSA unique and specific requirements without extraneous functionality that a COTS tool may provide. One possible implementation of this customized solution is a web front end design that interfaces with individual COTS tools to provide analytic and auditing capabilities. This design leverages the best COTS tools for analytics and auditing while providing the flexibility of meeting FSA unique business needs.

4 Results Achieved

The Data Strategy was developed to address FSA action items aiming to achieve the following:

- A reduction in redundant data storage
- Improve customer service
- Increase accuracy of analytics
- Increase efficiency in data handling
- Reduce costs
- Maintain a clean audit
- Ultimately remove FSA from the General Accounting Office's high-risk list

In order to develop a roadmap for this plan, the Data Strategy team worked with FSA to develop high-level business objectives for each of their business, data, and technical architectures. The purpose of this section is to demonstrate results that will be achieved by the development of the Data Quality Assurance Strategy Methodology and how it satisfies FSA business objectives while supporting the resolution of data quality issues within the enterprise.

4.1 Business Objectives

As an initial phase of the Data Strategy project, meetings were held with key FSA business and system owners in order to establish a baseline of objectives for the overall Data Strategy Target Vision. Over two hundred business objectives were gathered identifying goals to achieve and representing the direction FSA wants to proceed toward an enhanced business state (see Appendix B for a complete list of business objectives). Some of these goals are met or facilitated by the implementation of this Data Quality Methodology. The tables below isolate a subset of six Data Strategy Business Objectives that will be discussed in relation to the methodology, its processes, and their steps.

Priority	Objective No.	Business Objective Type	Business Objective	Details	Raw Objectives		Data Quality Methodology Process Steps
N/A	6	High-Level	FSA needs to establish and follow common data definitions to facilitate the exchange of data internally and externally.	- Significant Digits which includes requiring that pennies be a standard on dollar fields - There is a need for external systems to be able to handle loans that are greater than $99,999.99 in value. - FSA is held back from making changes to data formats because of the trading partners' inability to correctly adapt to those changes (resistance/lack of ability). - Business processes cannot be adversely affected by standardization/common edits, but must have enough edits to enable business process. - There is a need for an enterprise wide data dictionary because more than one system edits the same data differently - Be able to merge applicant and payment data as soon as the payment data comes in (for processing and analytics)	30, 34, 35, 36, 37, 39, 156, 179, 180	➔	**Methodology Utilization:** Through the use of the XML Core Component Library common definitionos for data will be created. The Data Quality Methodology can faciliate the correction of data through the following processes: **Prioritization** Phase - Rank quality issues based on data standards **Assessment Phase -** Develop solution for data correction utilizing system specific or enterprise toolsets. **Improvement Phase** - Implement solution and ensure enterprise systems are notified accordingly. **Oversight Phase -** Verify standards are being met through Enterprise Analytics and Audits.
1	27	Data Storage	Provide data access to varied resource needs, in the formats necessary to provide meaningful business information.	- User access to data. o Query capability across the enterprise data. o Concatenate across the enterprise. o Sufficient BI tools to allow user access and flexible query capability. o Select access to people across the enterprise. - System independent access to related data. - Provide the capability to share/inform results of online Surveys. - System access needs to DM/DW for data retreival. - Enable End to End customer visibility/metrics.	79, 88, 98,99,160, 193, 194, 199	➔	**Methodology Utilization:** In order to provide meaningful data across the enterprise, it is imperative that quality controls are in place with cross enterprise verification controls: **Prioritization** Phase - Rank quality issues based on data access isssues. Look for commonality between issues. **Assessment Phase -** Develop solution for data correction utilizing trading partner input. Improvement Phase - Implement solution and ensure enterprise systems are notified accordingly. **Oversight Phase -** Verify trading partners are able to access the data that they need to perform their work. This can be accomplished through surveys or interviews.
1	34	Data Storage	Provide data mining and analytics tools that allow simple access to data and facilitate trending, forecasting and the necessary information for business decisions.	- Provide Ad Hoc Query Capabilities. - Means to Bridge the Gap between "what I want" and "knowing the data" to get the information.	105, 159, 161, 162	➔	**Methodology Utilization:** As data is consolidated into a centralized data architecture, the quality of the data must be verified so that data mining and analysis can be done. This requires that quality controls and audits are in place to ensure that standards are enforced: **Prioritization** Phase - Rank quality issues based on consolidation or standards isssues. Look for commonality between issues. **Assessment Phase -** Develop solution for data correction based on enterprise standards such as the Core Component dictionary. Improvement Phase - Implement solution and ensure enterprise systems (or trading partners) are notified accordingly. **Oversight Phase -** Through cross program analytics and audits, guarranttee that the data in the centralized data architecture is standardized and free of issues.

Table 8 High Level Business Objectives – Sample 1

These samples illustrate the priority of each business objective in the context of the type of objective being defined. Within these samples, Data Storage, Internal Data, Web Services, and high level business objective types are shown. (For a full list of objective types, please see Appendix C). The table also lists the details behind the business objective, the underlying business objectives that it summarizes, and how the phases of the Data Quality Assurance Strategy support the objective.

Priority	Objective No.	Business Objective Type	Business Objective	Details	Raw Objectives		Data Quality Methodology Process Steps
2	33	Data Storage	Improve the timliness and efficiency of data feeds to accommodate varied audience data needs.	- Recognition of audience for creation and feeds to DM's. - Need for detail financial records/reporting (CMDM and FMS not sufficient). o Feeds from FMS can be out of order. o Data comes across un-reconciled.	151, 152, 157	➔	**Methodology Utilization:** To increase the timeliness and efficiency of the data, its quality must be increased through cross program examination and investigation: **Prioritization** Phase - Rank quality issues based on timeliness of data to internal systems and trading parnters. **Assessment Phase -** Develop solution to increase efficiency utilizing XML Framework and Technical Strategies. Improvement Phase - Implement solution according to standards and vision. **Oversight Phase -** Verify that FSA is implementing the Data Strategy through SME input, analytics, and audits.
2	12	Web Services	Provide access to common calculations in a standardized and central location.	- EFC Calculator. - Payment Calculator.	73, 145, 170	➔	**Methodology Utilization:** If a calculation is shared (the EFC calcuation can currently be initiated in CPS or through the Financial Aid Adminitrator of a schoo), there is a potential for data quality issues with multiple versions and non-standardized input. The methodology will: **Prioritization** Phase - Identify any quality issues with calculations and group similar issues together. **Assessment Phase -** Develop a solution for the issue by examining the Data Strategy vision for a service architecture. Improvement Phase - Implement solution in the confines the recommended strategy (could be a consolidation of the calculations into an enterprise standard). **Oversight Phase -** Verify the solution is working correctly (For the creation of an enterprise wide service, this includes ensuring access, output is correct, and invokation of the service works).
1	13	Internal Data	Centralized visibility and data flow control of the end to end interface process.	- Span of Control - systems that use interfaces of which they have no control. - Pre/Post bus reconciliation. - Increase audit/flow visibility. - Trace data touch points. - Improved work flow visibility and align processes to support it. o Processes should drive the systems and not the systems driving the processes. o Break up the silos of information.	23, 24, 27, 28,117	➔	**Methodology Utilization:** In order to impose control on the data within FSA's interfaces, the core component dictionary should be enforced for all new interfaces and quality issues that arise. This will translate into a slow migration to a standardized and centrally visiable interfave process. The Quality Methodology can facilated this by: **Prioritization** Phase - Ranking quality issues based on whether it occurs in an interface file. **Assessment Phase -** Assess solution in context of FSA's standardized core components. If a new core component needs to be developed, information Core Component Steering Committtee. Improvement Phase - Implement solution and ensure enterprise systems are notified accordingly. **Oversight Phase -** Verify complaince to core components and verification of quality issue resolution.

Table 9 High Level Business Objectives – Sample 2

The objective to "Improve the timeliness and efficiency of data feeds to accommodate varied audience data needs" shall be used to demonstrate the facilitation of business objectives through the implementation of an Enterprise Wide Quality Assurance Strategy. Table 6 indicates that this objective was derived from raw objectives 151, 152, and 157 and is grouped with a priority of 2 within the Data Storage business objective type (see Appendix C). Next, sample issues within this objective, such as a desire for better financial reporting, FMS feed timing problems, and lack of interface reconciliation, are presented. The implementation of the Data Quality Assurance Strategy Methodology, addresses each of these issues. For instance, the FMS feed timing problem will enter the Prioritization Phase, receive a rank, and be assigned to a Project Manager for Working Group processing. The Project Manager in the Assessment Phase will create a Working Group with representatives from the FMS business and technical groups as well as any system or trading partners affected by the issue. The Working Group will then formulate criteria for data sampling (potentially, FMS feed timings from multiple days) and initiate the development of solutions in the Improvement Phase. One potential solution

may be to implement controls to ensure proper sequencing of interface files or rejection of feed files initiated with an email notification to system operators. Once this solution has been tested and implemented into production, the Oversight Phase updates the Quality Report with the results.

4.2 Data Quality Examples

An effort to identify high priority data quality issues within the FSA enterprise were undertaken through the Data Quality Mad Dog effort (Deliverable 123.1.3.) The purpose of this effort was to compile data quality issues, prioritize the issues, and provide recommendations for resolving those data quality issues by working closely with FSA system owners, project teams and subject matter experts. The final product of this effort, a list of prioritized issues with viable recommendations, is included in the report as Data Quality Mad Dog Report Appendix A – Final Summary Report. It is important to understand that those issues are a recommended starting point for implementing the Data Quality Assurance Strategy Methodology.

This section demonstrates how data quality issues within FSA may be resolved through the application of the Data Quality Assurance Methodology recommended in this deliverable. <u>While the Mad Dog issues used in the following story are actual ones, the processes, tools, and other considerations are hypothetical and used only to illustrate how the Data Quality Assurance Strategy Methodology can be utilized.</u>

The following story is written in the present tense as though the Steering Committee is currently walking through each phase and stage of the Data Quality Assurance Strategy Methodology.

4.2.1 Data Quality Prioritization Phase

The Data Quality Steering Committee convenes for one of their regular meetings to prioritize the issues currently facing the enterprise.

4.2.1.1 Verification of As-Is State and Quality Issues

The Steering Committee reviews major changes to the enterprise since their last session. They discuss the process and system updates related to the following enterprise changes:

- Recipient Financial Management System (RFMS) has been successfully retired.
- The Gap Analysis for Trading Partner Management (TPM) has begun.
- Pilot Testing for the Standard Student Identification Method (SSIM) has started.
- Common Service for Borrowers (CSB) has been awarded and a detailed requirements phase has been initiated.
- FP Data Mart III has gone live.

Using the Quality Report, an analyst runs a report of all current data quality issues. The Steering Committee reviews the list. Included in the list are the following four unresolved quality issues:

- **Issue A:** Some PLUS borrowers are incorrectly being flagged for repayment status 60 days after their first disbursement date, rather then their second disbursement date (reference to Mad Dog Issue #14).
- **Issue B:** DMCS does not have current School information and is rejecting records from trading partners because school information does not match (reference to Mad Dog Issue #25).
- **Issue C:** There are no means within FSA to identify a trading partner institution (School, Lender, Servicer, GA, etc.) so that data from multiple stores within FSA can be aggregated for viewing or research (reference to Mad Dog Issue #8).
- **Issue D:** End users are using dummy identifiers to sign up for the Student Aid Internet Gateway (SAIG) (reference to Mad Dog Issue #26)

As the Steering Committee reviews the list, they update the issues when the description or complexity of the quality issue has changed since the issue was originally identified. For Issue C, they decide it is important that the capability to identify all current and historical relationships is included in the issue's wording. As such, Issue C is updated in the Quality Report and now reads, "There are no means within FSA to identify a trading partner institution (School, Lender, Servicer, GA, etc.) and all of their current and historical relationships to other entities so that data from multiple stores within FSA can be aggregated for viewing or research."

The final process in the verification stage is to compare the issues to the As-Is state of the enterprise. This step is necessary to recognize that the FSA environment is dynamic and may change from day to day. The Steering Committee discovers that the SAIG issue of end users using dummy identifiers has already been addressed by one of the operations teams. A fix has been implemented in production that performs basic data checks (e.g., SSNs of all ones are not allowed). The fix is deemed acceptable, and the Steering Committee updates the Enterprise Dashboard, indicating how and when Issue D was resolved.

All of the remaining issues are determined to still be relevant in the As-Is state of the enterprise. The next stage in the Data Quality Prioritization Phase is to determine the criteria for ranking the issues.

4.2.1.2 Determine Criteria for Ranking

The Steering Committee revisits the current list of issues and formulates a set of criteria to effectively prioritize them. They develop a three level, weighted approach:

- **Level 1:** Each issue will be ranked as a Quick Hit or a Substantial Investment. A Quick Hit is defined as an issue estimated to require less then $50,000 and less then three months for implementation. The Committee decides all Quick Hit issues will be weighted to have a higher priority then Substantial Investment issues.
- **Level 2:** The Committee decides to use the FMEA Analysis to further rank each issue within the Quick Hit and Substantial Investment groups. Using the FMEA, issues within the groups will be scored as Top 25%, Middle 50%, and Bottom 25% tiers, with Top 25% being the highest level two priority ranking.
- **Level 3:** Finally, the group decides to have Sequencing as the final criteria. Within the tiers identified in level two, the issues will be further prioritized according to any considerations for currently scheduled releases, system upgrades, and enhancements.

As previously mentioned, this is only a hypothetical example. The Steering Committee may decide to use alternate criteria for ranking the issues (the ranking order does not necessarily imply importance for top ranked issues). Some examples of criteria include:

- Immediate or long term impact to customers
- Impact to cost of contracts
- Regulatory compliance
- Correlation to overall FSA business or data strategy objectives
- Systems release schedules
- Cost of Poor Quality
- High level estimate of cost to resolve
- Peak processing times
- Life cycle phase impacts (upstream and downstream)
- Anticipated time for completion

4.2.1.3 Rank Quality Issues

Next, the Steering Committee processes the issues through the agreed upon criteria. Since a few issues in the list do not have enough information to accurately determine their prioritization, a note is entered in the Quality Report indicating that further analysis is required, and the issues are flagged for the Data Quality Oversight Phase.

Issues A, B, and C are processed through the criteria and the following results are generated:

Issue A:
Level 1/Impact Analysis: Material and Labor Costs is > $100,000
Time Required to Implement is 3 to 6 months
Issue is a Substantial Investment
Level 2/FMEA: Score is in Bottom 25% for Substantial Investment Issues

Process Step	Potential Failure Mode	Potential Failure Effects	Severity	Potential Causes	Occurrence	Current Controls	Detectability	Score
Borrowers are flagged for repayment.	Plus Borrowers are flagged for repayment too early.	Poor customer service. Either a student begins paying on a loan before they should have to or the student must manually work with FSA to correct their scheduled repayment.	4	The repayment date is caluculated as 60 days after the second disbursement date. However the field is currently not required on the XML Common Record	4	No automatic verification of whether there should have been a reported second disbursement date. All changes are manual made per request from the customer.	2	32

Table 10 Business Example Issue A: FMEA Process

Level 3/Sequencing: Updates to the Common Record should correspond with the beginning of an Award Year.

Issue B:

Level 1/Impact Analysis: Material and Labor Costs is < $50,000
Time Required to Implement is < 3 months
Issue is a Quick Hit

Level 2/FMEA: Score is in the Top 25% for Quick Hit Issues

Process Step	Potential Failure Mode	Potential Failure Effects	Severity	Potential Causes	Occurrence	Current Controls	Detectability	Score
Default or overpayment records are submitted to DMCS from external partners.	Default Records are rejected.	FSA is unable to collect on defaults or overpayments.	7	School data in DMCS is not accurately updated so when updated school data comes in on a default record from a partner the record is incorrectly rejected.	4	There is an automated check on school fields when partner records are received. However, school updates for DMCS school data are ad hoc and manual.	3	84

Table 11 Business Example Issue B: FMEA Process

Level 3/Sequencing: Sequencing has little or no impact on this issue and a fix should be implemented as soon as possible.

Issue C:

Level 1/Impact Analysis: Material and Labor Costs is > $500,000
Time Required to Implement is > 12 months
Issue is a Substantial Investment

Level 2/FMEA: Score is in the Top 25% for Substantial Investment Issues

Process Step	Potential Failure Mode	Potential Failure Effects	Severity	Potential Causes	Occurrence	Current Controls	Detectability	Score
Trading Parntner data is viewed for Enterprise analytics and research.	Unable to pull complete and accurate view of the Trading Partner.	FSA is unable to effectively oversee partners potentially leading to fraud and negligence in the programs.	10	No standardized, enterprise-wide method for indentifying Trading Partners	7	Various system specific controls. Some systems have cross-walks between the various Trading Partner IDs	10	700

Table 12 Business Example Issue C: FMEA Process

Level 3/Sequencing: The RID will be addressed as part of the TPM solution.

Once the issues are prioritized, the Steering Committee creates time frames for issue assignment and selects Project Managers who will organize Working Groups. The committee assigns Issue A to a Project Manager and indicates the issue should be implemented in time for the ’05 Award Year. Issue B is assigned to another Project Manager with the expectation that the issue will be immediately addressed.

The Steering Committee determines that a solution for Issue C should be provided as part of a new development project, TPM. As such, Issue C should be addressed by the Enterprise Quality Assurance Program and is assigned to the lead for the TPM work. All Quality Report information regarding the issue is shared to ensure the development of the TPM solution contains the necessary requirements and business rules for resolving the issue. As the TPM work progresses, all issue/solution status and results from are reported back to the Steering Committee, stored in the Quality Report, and used throughout the Data Quality Oversight Phase in the Data Quality Methodology.

4.2.2 Data Quality Assessment Phase

Per the Steering Committee's requests, the Project Managers for Issues A and B form issue specific Working Groups with cross program representatives who are affected by the issue. The first task for the Working Groups is to perform a Data Quality Assessment whereby the data is inspected, the defects are measured, and the cause and impact of the defects are analyzed.

4.2.2.1 Pre-Assessment Planning Stage

4.2.2.1.1 Issue A

The Project Manager initiates the Pre-Assessment Planning process by reviewing the data quality issue information received from the Quality Report. The Project Manager sets three high level goals for the Assessment Phase.

1. Establish a viable Working Group
2. Acquire a comprehensive understanding of Issue A
3. Obtain a sample set of data which accurately depicts the quality issue

While Issue A may impact a number of FSA systems and external partners, the Project Manager decides to set the scope of the project to focus only on schools' Common Record interface with Common Origination Disbursement (COD) and COD's interface with Direct Loan Servicing System (DLSS). Any related issues not in this scope will be reported to the Steering Committee for further review in the Data Quality Oversight Phase.

After establishing the scope, the Project Manager discusses Issue A with both COD and DLSS Subject Matter Experts to better understand the requirements for assessing the issue and implementing a solution. The Project Manager develops a project plan and determines the following resources will be required:

- Subject Matter Experts/Business Owners (2) – one from each COD / DLSS
- Database Administrator (2) – one from each COD / DLSS
- Technical Architecture Specialists (3) – one from the Common Record & one each from the COD / DLSS Interface
- Analysts (2 -3) – at least one from each COD / DLSS support staff

Once organized, the Working Group begins by identifying acceptance criteria. The group uses the Voice of Customer (Tool) to facilitate this effort. The results from this effort are as follows:

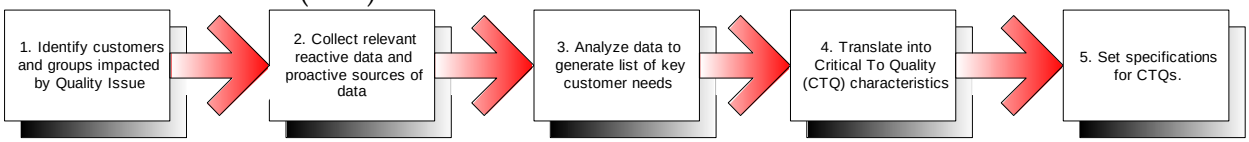


Figure 39 . Business Example Issue A: VOC Tool

1. **Impacted Customers** - Schools, COD, DLSS, Borrowers
2. **Reactive Events –** Premature repayment status for aid recipients, borrower complaints.
   Proactive Events – DLSS sends letter to Borrower welcoming them to Direct Loan Servicing
3. **Key Customer Needs –** Accurate repayment status date
4. **CTQ Tree –** (see figure below)

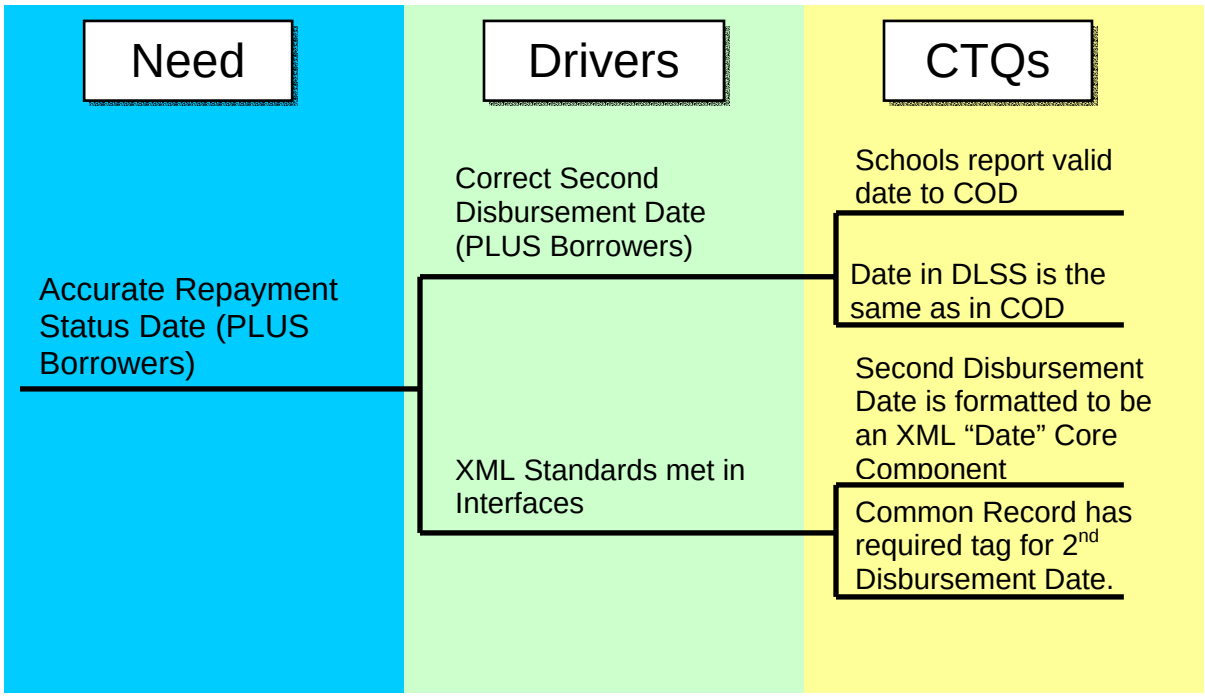


Figure 40 . Business Example Issue A: CTQ Tree 1

5. **Sample Data Criteria:** Using the CTQs, the Working Group decides on the following criteria, weighting, and acceptable score to determine a proper data sample.

Criteria	Percent of Dummy or Invalid Dates	Percent of Mismatches between COD & DLSS	Percent of Data with Incorrect Format	Percent of NULL Date Fields	**Acceptable Score**
Weight	0.3	0.2	0.1	0.4	1

Table 13 Business Example Issue A: Sample Data Criteria

4.2.2.1.2 Issue B

The Project Manager for Issue B notes that the issue is a Quick Hit and feels that the issue should not require an extensive level of Pre-Assessment Planning. The Project Manager sets a high-level goal for a thorough but expedient resolution to the issue and limits the scope of the effort to the school data in DMCS, its interfaces, and the PEPS Daily School File.

After discussing Issue B with DMCS Subject Matter Experts, the Project Manager develops a project plan and determines the following resources will be:

- Subject Matter Expert/Business Owner (2) – one from each PEPS & DMCS
- Database Administrator (1): A DBA will be required from DMCS to ensure the proper database load procedures
- Technical Architecture Specialist (1) - PEPS School File Interfaces
- Analysts (1) – Provides assistance wherever needed (documenting all processes, assist in creation of test scripts, and updates to Quality Report)

Similar to the Working Group for Issue A, this Working Group uses the Voice of Customer (Tool) to facilitate their efforts.

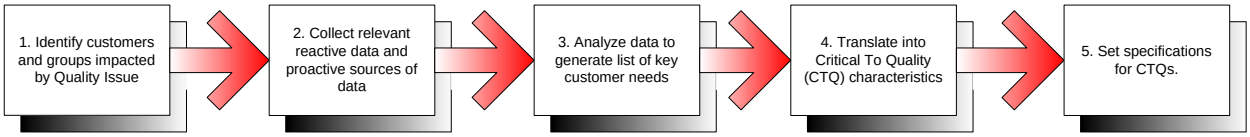


Figure 41 . Business Example Issue B: VOC Tool

1. **Impacted Customers –** FSA, DMCS, trading partners
2. **Reactive Events –** DMCS rejects default/overpayment records, trading partner submits complaints to FSA
   Proactive Events – Ad-hoc updates are made in an attempt to provide DMCS with accurate school data
3. **Key Customer Needs –** Capability to receive and process default/overpayment records
4. **CTQ Tree –** Since this issue is relatively simple and has been well defined, the Working Group creates two CTQs (see diagram below).

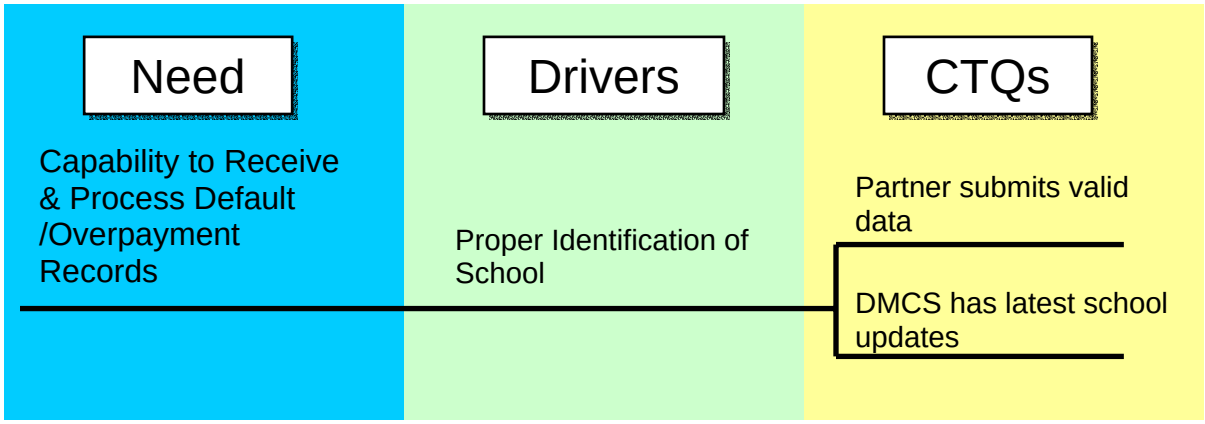


Figure 42 . Business Example Issue B: CTQ Tree

5. The Working Group decides that an acceptable level for default records being rejected due to School Data errors is .1% or less. Solutions that meet this acceptance criterion will be considered for pilot testing.

4.2.2.2 Initial Data Assessment Stage:

For both Issue A and Issue B, the Subject Matter Experts accept the established criteria, and the Working Group begins the actual data assessment.

4.2.2.2.1 Issue A

The Working Group's database administrator takes three samples of data from the DLSS system. The team's analysts then score the data using the criteria established during the Pre-Assessment Planning Stage and generate the following results:

Data Sample Set	Criteria with Weights				Score	Is Data Quality Acceptable?
	Percent of Dummy or Invalid Dates	Percent of Mismatches between COD & DLSS	Percent of Data with Incorrect Format	Percent of NULL Date Fields		
	0.3	0.2	0.1	0.4		
1	0.3	0	0.1	0.5	0.9	Y
2	0.2	0.1	0.1	0.9	1.3	N
3	0.5	0.2	0	1.4	2.1	N

Table 14 Business Example Issue A: Sample Data Criteria

Both sets #2 and #3 provide a good sample set of poor quality data. Since set #3 has the highest (or worst) score, it is selected to be used for the data sample in the overall root cause analysis.

To initiate a root cause analysis, the Working Group first creates a SIPOC diagram to better understand the As-Is process related to Issue A.

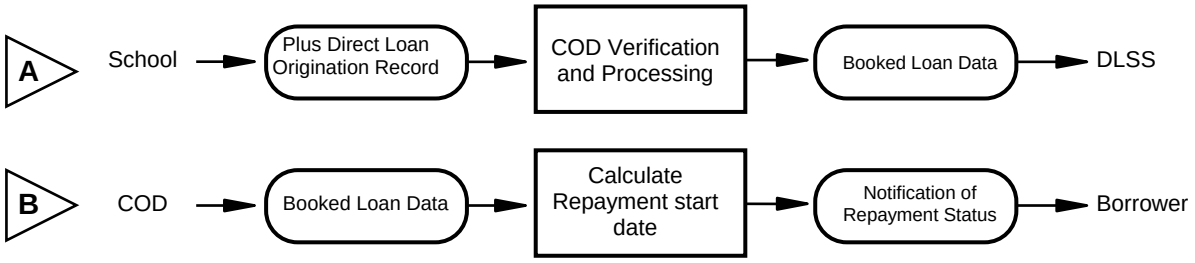


Figure 43 . Business Example Issue A: As-Is State SIPOC Diagram

The processes illustrated by the SIPOC diagram do not have any readily apparent quality concerns. However, the diagram still highlights the systems and processes that can be further researched as part of the root cause analysis. To facilitate this research, the Project Manager instructs the Analysts to work with the Subject Matter Experts and create a Fishbone Diagram and a Pareto Chart.

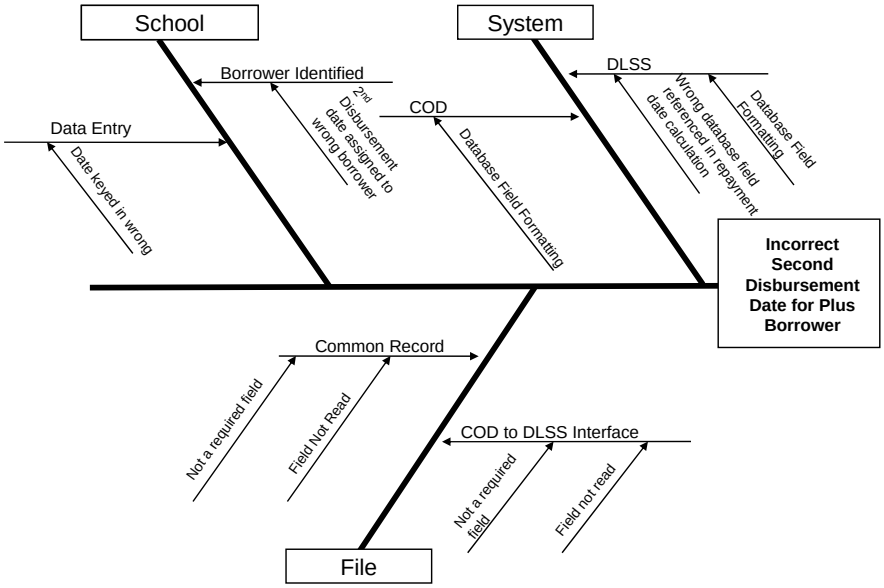


Figure 44 . Business Example Issue A: Fishbone Diagram

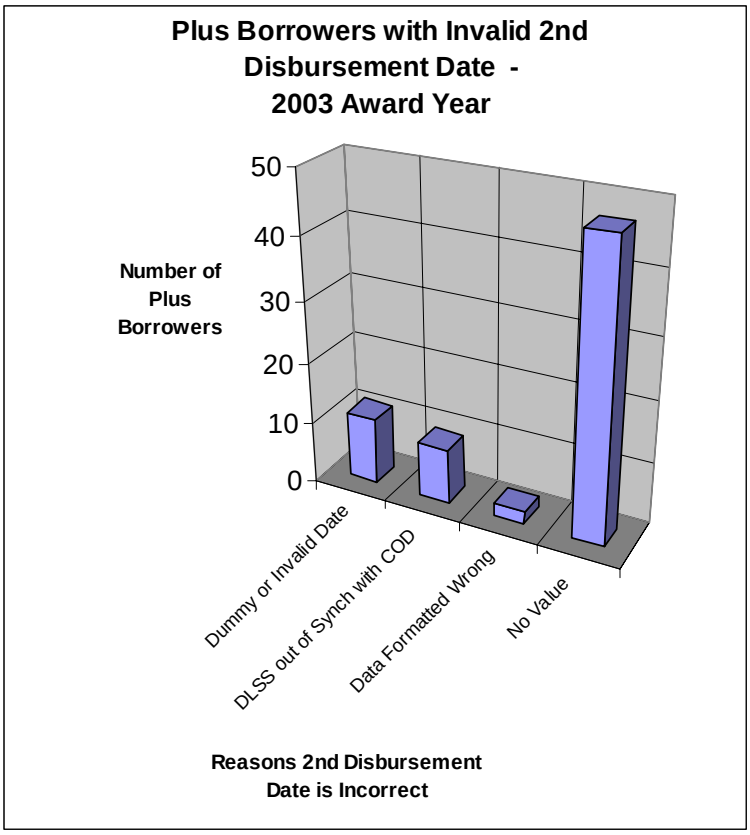


Figure 45 . Business Example Issue A: Pareto Chart

Using the Pareto Chart, the Working Group Analysts determine that the majority of PLUS Borrowers entering repayment early do not have a value reported for their Second Disbursement Date. After researching the various possible causes on the Fishbone Diagram, the group discovers that the field is not a required field on the XML Common Record.

The Analysts review their findings with the group's Subject Matter Experts and the Project Manager. The Project Manager reports back to the Steering Committee informing them that the issue may require the XML Standards and Business Rules to be updated, requiring schools to submit the Second Disbursement Date on the XML Common Record. The Project Manager then instructs the Working Group to begin the Data Quality Improvement Phase.

4.2.2.2.2 Issue B

In order to sample and score the DMCS school data, the Working Group uses three sample sets of default records and sends them through the DMCS incoming file processes. For each set, .5%, .6%, and .2% of the records, respectively, are rejected due to invalid school data for each of the respective sample sets of data. All of the samples are above the acceptable .1% level. Since the second set demonstrates the most issues, the Working Group uses it for their additional assessment efforts.

While it is known that default records sent to DMCS by trading partners are rejected due to mismatched school data, the Project Manager instructs his team to conduct a root cause analysis for verification and to determine if there are any other substantial issues with this process.

The Working Group Analysts work with the Subject Matter Experts to create a SIPOC Diagram, Fishbone Diagram, and Pareto Chart.

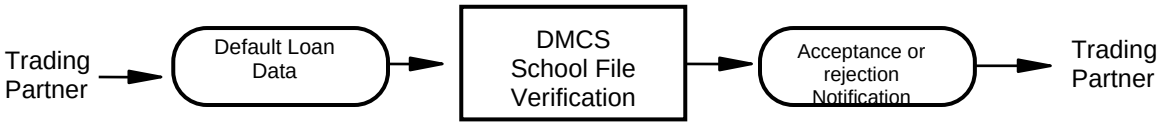


Figure 46 . Business Example Issue B: SIPOC Diagram

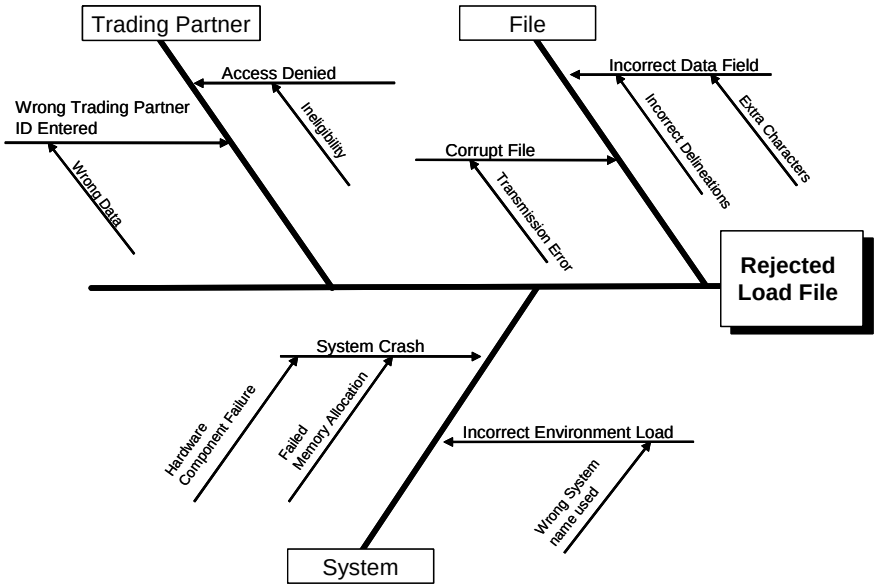


Figure 47 . Business Example Issue A: Fishbone Diagram

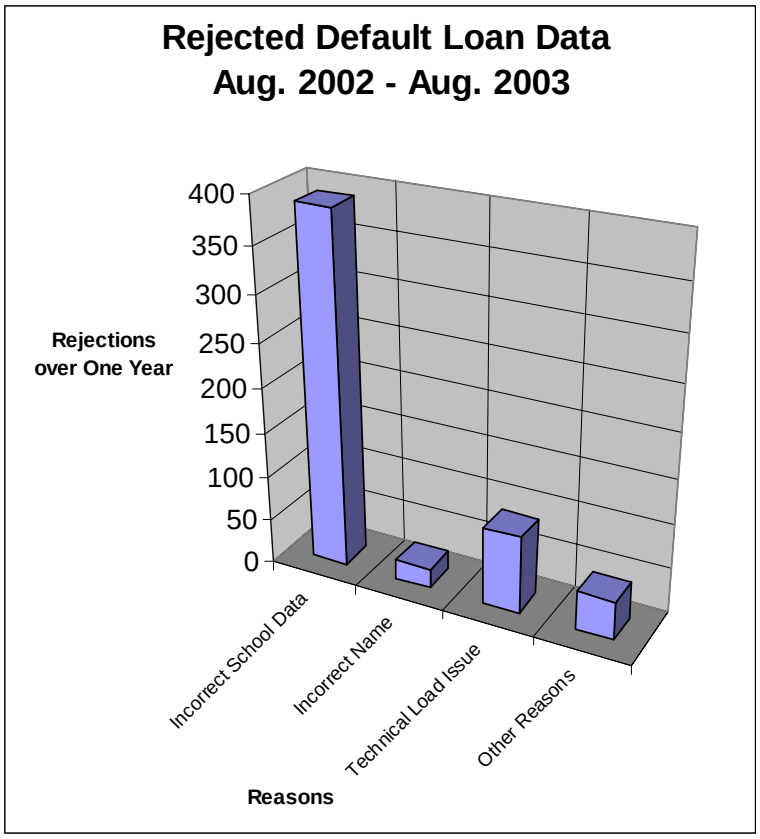


Figure 48 . Business Example Issue A: Pareto Chart

While the SIPOC Diagram and Fishbone are useful in identifying a number of potential causes for File Load Failures, the Working Group determines that the only significant quality issue is the incorrect DMCS school data. The Project Manager reports these findings back to the Steering Committee and the Quality Report is updated with a note that while there are other load failures for default records into DMCS, these errors are minor and should not be considered a significant quality concern. The Project Manager then instructs the Working Group to begin the Data Quality Improvement Phase.

4.2.3 Data Quality Improvement Phase

Based on quality findings in the Assessment Phase, the Working Groups for Issues A and B begin the Improvement Phase. During this phase, they develop, assess, and test solutions for their assigned issue.

4.2.3.1 Solution Definition and Assessment Stage

4.2.3.1.1 Issue A

The Working Group develops a set of criteria which will help determine the best solution option. Since this quality issue has led to a number of PLUS borrowers being dissatisfied with their FSA experience, the group decides to weight Customer Satisfaction as the most important

criteria. While Ease of Implementation is still a factor that should be considered, the group feels that their team is highly capable of handling complex solutions and agrees the criteria can be given a lower weighting.

Criteria	ROI	System Impact	Ease of Implementation	Customer Satisfaction
Weight	0.2	0.2	0.1	0.3

Table 15 Business Example Issue A: Solution Criteria Matrix

Next, the group begins to brainstorm for possible solution options. Each team member has a unique background and perspective and is able to provide valuable input. Three possible solutions are documented:

- **Solution Option #1 –** Require schools to submit the Second Disbursement Date on the XML Common Record. Add Cleaning and Scrubbing functionality to COD to validate the incoming Common Record data.
- **Solution Option #2 –** Use a System Correction approach by adding logic to DLSS to calculate the Second Disbursement Date when one is not given. (Note: The date would be calculated as the point half way between the 1st disbursement and the end of the loan period. However, because this work around is only an estimate, borrowers could still enter repayment too soon or too late.)
- **Solution Option #3 –** Use a Reconciliation approach. Modify the DLSS Student Status request process to include a request for PLUS Borrower Second Disbursement Date updates. After receiving a list of PLUS Borrowers, the National Clearinghouse gathers PLUS Borrower information from schools and submits it to NSLDS (similar to student status updates). DLSS would then reconcile their data with NSLDS and make any necessary updates.

These three options are then assessed against the weighted criteria. The Working Group's Analysts research the issues and provide the rest of the group with a variety of tools for comparing the issues to the criteria (e.g., a Cost of Poor Quality (COPQ) analysis is conducted to estimate the financial loss experienced by FSA if the issue were to persist unchecked). Each member in the group ranks each option according to how well they feel the solution meets the various criteria. These rankings are summed and multiplied by the assigned weight to score each option.

	Criteria with Weights				**Overall Score**
	ROI	System Impact	Ease of Implementation	Customer Satisfaction	
Solution Option	0.2	0.2	0.1	0.3	
1	2.8	1.5	1.6	3.1	6.2
2	0.7	2.4	2.6	0.2	5.2
3	1.5	1.1	0.8	1.7	3.6

Table 16 Business Example Issue A: Solution Criteria Assessment Matrix

The scoring indicates Option #1 is the best solution. The Project Manager agrees with the assessment and orchestrates the Pilot Testing Stage.

4.2.3.1.2 Issue B

The Working Group develops a set of criteria to determine the best solution option. Since this issue is a Quick Hit, the group agrees that Deployment Speed and Ease of Implementation should be weighted slightly more then some of the other criteria.

Criteria	ROI	Deployment Speed	System Impact	Ease of Implementation	Customer Satisfaction
Weight	0.1	0.3	0.1	0.3	0.2

Table 17 Business Example Issue B: Sample Data Criteria

The Working Group decides there are only two viable solution options:

- **Solution Option #1 –** Accept all default records without performing a match on the school data. Work with PEPS and the trading partner to determine whether DMCS needs to update its school data or the trading partner needs to update their data.
- **Solution Option #2 –** Have DMCS receive the PEPS Daily School File updates. DMCS would still reject all records with invalid school data.

The Working Group ranks the solution options and determines the overall scores.

	Criteria with Weights					**Overall Score**
	ROI	Deployment Speed	System Impact	Ease of Implementation	Customer Satisfaction	
Solution Option	0.1	0.3	0.1	0.3	0.2	
1	0.1	4.1	2.8	3.3	2.6	12.9
2	4.9	2.9	2.2	3.7	3.4	17.1

Table 18 Business Example Issue B: Solution Criteria Assessment Matrix

Since Option #1 would require a great deal of manual processing, it ranks poorly for the ROI criteria. Overall, Option #2 is identified as the best solution. After selecting Option #2, the Project Manager instructs the Working Group to being the Pilot Testing Stage.

4.2.3.2 Pilot Testing Stage

4.2.3.2.1 Issue A

The Working Group's Technical Architecture Specialists work with the Database Administrator to develop the Cleaning and Scrubbing fix in a test environment for COD and DLSS. Three data sets are then run through the processes in the updated test environment. The Working Group loops back to the Assessment Phase in order to verify the data sets pass the acceptance criteria.

4.2.3.2.2 Issue B

The Database Administrator creates a testing environment for the Working Group. The PEPS and DMCS Technical Architecture Specialists work together to determine all of the technical requirements for sending and receiving the Daily School File. Cleaning and Scrubbing functionality is added to DMCS to ensure the school data is correct as it enters the system. Three sets of default records are then sent to DMCS to mimic the production processing. The Working Group uses the results from the testing to assess whether the solution is successful.

4.2.4 Data Quality Assessment Phase

4.2.4.1 Pilot Testing Data Assessment Stage

4.2.4.1.1 Issue A

The Working Group scores the Pilot Testing data samples using the same criteria defined in the Initial Data Assessment Stage.

Data Sample Set	Criteria with Weights				Score	Is Data Quality Acceptable?
	Percent of Dummy or Invalid Dates	Percent of Mismatches between COD & DLSS	Percent of Data with Incorrect Format	Percent of NULL Date Fields		
	0.3	0.2	0.1	0.4		
1	0.3	0	0.1	0.1	0.5	Y
2	0.2	0.1	0.1	0	0.4	Y
3	0.4	0.2	0	0	0.6	Y

Table 19 Business Example Issue A: Pilot Testing Data Sample Results

All of the sample data sets have a score less than 1.0 and are considered acceptable. The Project Manager reviews the results and notifies the Steering Committee that the solution is ready for production implementation. The Steering Committee concurs with the Project Manager's assessment.

4.2.4.1.2 Issue B

Using the results from the Pilot Testing, the Working Group finds .05%, .01%, and .02% of the records in the respective sample sets error out due to invalid school data. Since all of the samples sets meet the acceptable .1% or lower level, the solution is deemed ready for production.

4.2.5 Data Quality Improvement Phase

4.2.5.1 Data Quality Implementation Production Stage

4.2.5.1.1 Issue A

The Working Group's Technical Architecture Specialists work with the Database Administrator to ensure the Cleaning and Scrubbing XML Common Record fix is added to the production environment. The implementation is successful and the Working Group reports the issues analysis, resolution methodology, and solution results to the Enterprise Steering Committee analysts. The analysts update the Quality Report to reflect the Working Group's progression and ultimate resolution of the data quality issue.

4.2.5.1.2 Issue B

The Working Group successfully implements the PEPS to DMCS Daily School File interface into production. All of the results and information related to the quality issue are sent to the Steering Committee to be entered in the Quality Report.

4.2.6 Data Quality Oversight Phase

As the various Working Groups finish the Data Quality Assessment and Data Quality Improvement Phases, they share all results of the implementation with the Enterprise Steering Committee. Results from issues addressed by new development efforts and the Enterprise Quality Assurance Program are also given to the Steering Committee. Analysts on the committee update the Enterprise Quality Report and continue to monitor the issues as they perform routine and ad hoc analytics and audits.

As the Steering Committee Analysts review the Quality Report, they begin to compile a new list of quality issues that will be addressed during the next Data Quality Prioritization Phase, beginning the Data Quality Assurance Strategy Methodology anew.

5 Implementation Plan

Data quality is improved through an iterative, evolutionary process rather than a single, revolutionary event. Furthermore, enhanced data quality within FSA is contingent upon a uniform adherence to the effort throughout the enterprise. Due to the interdependency of FSA systems, the failure of one system to follow enterprise quality standards jeopardizes the quality of all enterprise data.

The Data Strategy project has taken the initial steps toward the implementation of a data quality plan for the enterprise. Through the meetings and processes surrounding the creation of the Data Quality Mad Dog Report a high level methodology for obtaining uniform enterprise-wide quality issues was essentially pilot tested. The Mad Dog team worked with Business Owners from the various FSA channels to identify and evaluate key issues based on business needs, cost, value, and urgency. The Mad Dog effort served as a catalyst for defining a detailed ongoing Data Quality Methodology as presented in this document. This methodology addresses the need for continued identification, prioritization, and resolution of quality issues.

To fully implement and realize the benefits of the Data Quality Assurance Strategy, a number of additional actions are needed. These actions are outlined in the following diagram. It should be noted that this diagram and the supporting information in the following sections provide a high-level sequencing plan. Additional detail needs to be defined by the Business Owners. Timing issues and the availability of resources such as budget, skills, and management time will need to be taken into consideration.

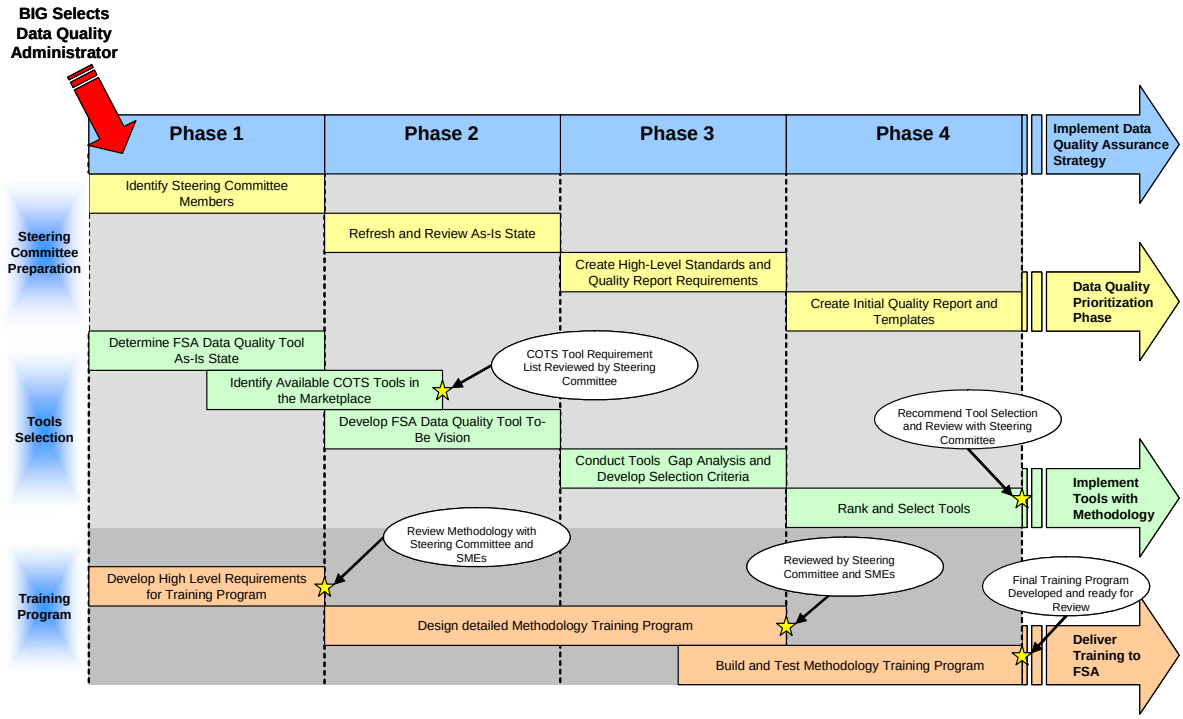


Figure 49 . Data Quality Assurance Strategy Implementation Plan

The Implementation Plan has four preparatory phases, followed by the actual implementation of the Data Quality Assurance Strategy. Throughout these phases, there will be three groups simultaneously working on different preliminary pieces of the overall Quality Strategy:

- **Steering Committee:** A Data Quality Administrator appointed and assisted by the BIG, will form a Steering Committee for the Data Quality Assurance Strategy. Preliminary activities will be establish standards, determine criteria for the Quality Report, and review Tool Selection and Training efforts.
- **Tools Selection Working Group:** The Data Quality Administrator, assisted by the BIG, will form a group of technical and business personnel to do a COTS tool analysis of data quality applications in the market place. Activities include requirements gathering, developing application criteria, assessment of COTS tools, and a final tool recommendation.
- **Training Program Working Group:** In a similar process, the Data Quality Administrator, assisted by the BIG, creates a team of cross functional representatives to create an enterprise wide training program for the Data Quality Assurance Strategy.

These groups are discussed below with action plans outlined for each phase. Note that the initiation of the Data Quality Assurance Strategy will begin with the FSA BIG selecting a Data Quality Administrator to assemble the three Working Groups. This Data Quality Administrator should have some knowledge or experience in the major Data Quality methodologies and have a strong basis of FSA business knowledge. After forming the Working Groups, the

Administrator will consult and monitor the groups as they progress through the preparatory phases.

5.1 Group 1: Steering Committee Preparation

The Steering Committee will play an important role both before and after the Implementation Plan. Before implementation, during the four preparatory phases, this group will establish standards and Quality Report requirements that will serve as the foundation for the ongoing Data Quality Methodology. Once the Quality Report and the methodology are implemented, the Steering Committee will conduct the Prioritization and Oversight Phases as part of the on-going quality methodology (for more detail regarding the Steering Committee's continuing post-implementation responsibilities see Section 3.1 - Data Quality Steering Committee.)

5.1.1 Phase 1

In this first phase, the Data Quality Administrator will create the Steering Committee based on the roles and responsibilities outlined in Section 3.1. After this team is created, they will receive an overview of the Data Quality Assurance Strategy Methodology by the Training Working Group (currently in the process of developing high-level training requirements). This overview will be a condensed version of the Quality Assurance Strategy to provide the Steering Committee a list of their roles and responsibilities, tasks, and objectives.

5.1.2 Phase 2

After the Steering Committee has been established, they will review the FSA enterprise As-Is state and discuss any recent changes to FSA systems, processes, data flows, and business rules. This current state view of FSA will allow the Steering Committee to determine if any data quality issues have been resolved by recent system modifications. The major release schedules of all projects within the enterprise must also be taken into consideration.

5.1.3 Phase 3

The Steering Committee members will modify and, if necessary, add to current business objectives and standards. As cross-lifecycle representatives, they are able to create new standards that may be necessary due to updates to the current As-Is process flows. Based on these standards, the Steering Committee will outline Quality Report requirements and develop rudimentary tolerances and metrics utilized to measure data quality throughout FSA. Specific metrics should be set to measure quality levels. These could be in the form of defect percentages within given databases or percentage of files that are meeting the set standards. Formatting and procedural standards can be considered here, such as XML Core Components, necessary data elements for stored files, and procedures for proper database correction. These requirements will aid the Tools Selection Working Group as they conduct their tools selection (see Section 5.2.4 - Phase 4).

5.1.4 Phase 4

In the final preparatory phase before implementing the Data Quality Assurance Strategy, the Steering Committee will assemble an initial Quality Report and business templates to be utilized by the analysts and Working Groups throughout the new methodology. It is recommended that FSA initially use a customized Quality Report which provides a high level view of quality levels across the enterprise. Using the set tolerances, the Steering Committee will formulate a Quality Report to outline current issues across FSA systems. This will provide an overall view of the current "health" of the enterprise and will help measure the breadth and severity of FSA's current issues. Initially, a customized Quality Report may simply list issues for all FSA systems with percentage levels of defective data. This initial Quality Report could be as simple as an Excel spreadsheet that effectively tabulates the existing issues and developed metrics.

Once the Tool Selection Working Group finalizes its recommendations for a Quality Report, the Steering Committee will take the existing issues and metrics held within the initial Quality Report (e.g., Excel spreadsheet) and uploaded it to the new recommended tool.

5.1.5 Implementation

Since the Data Quality Mad Dog Report (Deliverable 123.1.3) determined and evaluated key data quality issues based on business needs, cost, value, and urgency, it will provide the initial set of data quality issues to be reviewed by the Steering Committee. The Steering Committee will then identify those issues that will be addressed by future system modifications and those that will require Working Groups. For issues requiring Working Groups, the Steering Committee may initially want to focus on the Mad Dog Quick Hits, and then, as familiarity with the Quality Report and Data Quality Methodology increases, more complex issues can be addressed.

5.2 Group 2: Tools Selection

Tools are a fundamental component of the analytic and audit processes in the Data Quality Strategy Methodology. They will deliver analytical capabilities and will provide measurable data which will facilitate the assessment and audit of FSA data quality. The Data Quality Administrator will develop a Tool Selection Working Group in charge of determining the need for and selection of data quality tools for analytics and audits. This team will consist of technical and business analysts who will conduct market research and develop criteria for COTS tools. There will also be cross program Business Owners who can provide guidance and input into the selection process to ensure FSA business objectives are taken into consideration.

5.2.1 Phase 1

The Tool Selection Working Group will first determine the As-Is data state of quality tools within FSA. All current tools and improvement techniques will be listed and analyzed to determine applicability and their capabilities. Halfway through this phase, the Working Group will also begin to identify COTS tools that are currently available in the marketplace.

5.2.2 Phase 2

During this phase, the Working Group will complete their marketplace COTS tools analysis and initiate a high level requirements gathering for an FSA specific Enterprise-Wide Data Quality tool. Requirements will be gathered from the Working Group's knowledge of current FSA data quality tools, business needs, and current technological offers in the market place. The result from this process will be a list of what requirements are necessary to implement the new Data Quality Methodology with proper analytics and audits available to FSA. At this point, the Tools Selection Working Group will review a high-level methodology presentation developed by the Training Working Group (the same that the Steering Committee reviews), to ensure that their requirements are created in the context of the Quality Assurance Strategy Methodology.

5.2.3 Phase 3

In the third phase, the Tool Selection Working Group will conduct a tools gap analysis, comparing the FSA As-Is and To-Be tool availabilities and capabilities to the requirements gathered in phase two. This will identify functionality that the current tools within FSA lack and potential gaps that market place COTS tools can provide. The information gathered in the gaps analysis coupled with FSA requirements for an enterprise wide tool will help formulate selection criteria for ranking in the next phase. Examples of selection criteria could be features such as a tool having a Quality Report, hardware requirements for a tool, costs, being able to hook into a variety of database architectures, or the ability to facilitate the design build and test of quality solutions.

5.2.4 Phase 4

After establishing the quality tool selection criteria in Phase 3, the Working Group will rank the tools by assessing each tool against the criteria. The Working Group will select the top ranked tools and present their findings to the Steering Committee. Once approved, the data quality tool(s) will be purchased and implemented within the enterprise.

5.2.5 Implementation

Once the selected data quality tool(s) are incorporated as part of the Data Quality Assurance Methodology, the Tool Selection Working Group must consider any necessary training around the tools' use. Both customized, in-house training and standard vendor training should be considered and a project plan for tool training development should be developed. They will work with the Training Program Development Working Group to incorporate any selected tool training into the overall Data Quality Assurance Methodology training effort. While the Tool Selection Working Group is not a permanent team, a few of it members could be members of the Steering Committee, providing technical and function assistance regarding the data quality tools.

5.3 Group 3: Training Program Development

All levels of FSA personnel dealing with data quality will need to be trained in the new Data Quality Assurance Strategy. The Steering Committee and the Tools Selection Working Group will receive an overview of the methodology in Phase 1 to help them in their efforts throughout

their other preparatory efforts. However, by the end of Phase 4, a more detail training program will be available for these groups and for all of FSA. Training will be developed and coordinated by the Training Working Group, formed by the Data Quality Administrator with the assistance of the BIG. This group will identify a training team consisting of data quality experts and business analysts to develop a training program for the enterprise. The group will be responsible for developing high level requirements incorporating input from FSA business as well as technical representatives, designing a training program that is inclusive of FSA quality business objectives, and use of appropriate data quality tools. As a final step, the Working Group will decide upon a delivery method for this application (whether web based training or a standalone application) and schedule the training for deployment within FSA, potentially through the FSA University organization.

5.3.1 Phase 1

The Training Working Group will have two deliverables in this phase. The first will be to develop a high-level overview of the Data Quality Assurance Strategy Methodology. This overview can be in the form of a presentation that will outline for the Steering Committee members and the Tools Selection Working Group how the various phases of the methodology interact and the type of analysis that needs to be conducted.

The second objective of this phase is for the Training Working Group to develop high level requirements for a new training program. As part of this, data quality experts will be leveraged and Business Owners will be polled to ensure proper context for the training. The Working Group will also evaluate delivery methods (web based, paper, or a stand alone electronic application) and pick one to present to the Steering Committee. Technical specifications will need to be created and a cost analysis needs to be conducted to provide the Steering Committee with enough information regarding time and resource requirements to pick a proper solution. After the Working Group develops the detailed requirements of the training program, the program and delivery design will feed into a detailed design phase.

5.3.2 Phase 2

In this phase and through the next, the Working Group will prepare detailed designs for a training program that will educate FSA personnel. This program will consist of a new training application that will explain in detail how each stage and process is conducted within the phases and the specific tools and their applications as outlined in Section 3.

In this phase, the Working Group will also decide how the training program should be delivered to the other Working Groups, analysts and Business Owners within FSA. It is recommended that Computer Based Training (CBT) be utilized. This could be a single or series of online training modules detailing the methodology and necessary steps. Thus, personnel can train at their convenience and have a reference tool available to them for any retraining or refreshing.

5.3.3 Phase 3

The detailed training program design will continue through this phase which will cumulate in a presentation to the Steering Committee for design approval. The Working Group will also begin half way through this phase to start preliminary development activities of the training program (project plans, issues database, documentation repository, etc.).

5.3.4 Phase 4

In the final preparatory phase, the Working Group will continue to build out the training program application with its associated documentation (issue lists, coding standards, test scripts, test scenarios, test plan, etc.). This group will also decide how they will present the data and what type of scheduling will need to occur in order for FSA employees to receive this training. All of these efforts will also be coordinated by working with the FSA University organization.

Once the training application has been built, the training application will initiate the normal testing life cycles as defined by the FSA Solution Life Cycle in the *FSA System integration and Testing Process Handbook* (Unit Testing, Integration Testing, System Testing, Performance Testing, and User Acceptance Testing). It is important to point out that the last testing process, the User Acceptance Test, identifies a small group of Business Owners who will assess the training and provide feedback as part of a survey. Through this feedback, the Training Team will be able to determine how effective the training application is and where improvements may be needed before deploying the training throughout FSA.

5.3.5 Implementation

As the Data Quality Assurance Strategy is being implemented, the Training Working Group will deploy the methodology training across FSA to all Working Groups and analysts involved. This training may involve scheduling classroom time with employees or centralizing access to any web based training that has been developed, again leveraging the FSA University organization where appropriate.

After all preparatory phases have been completed, the first iteration of the Data Quality Methodology cycle will enable all those involved with the Data Quality process to become familiar with the new methodology and will enable the methodology to mature within FSA.

6 Future Considerations

The FSA Data Strategy effort has culminated in a conceptual model for enabling an enhanced target state vision. Each Data Strategy deliverable provides a unique piece in this enterprise wide model. This document defines a quality process to maintain the integrity of the data before, during, and after the business is transformed to the To-Be vision.

Although this document defines a Data Quality Methodology, it does not actually implement the strategy and there are future considerations that are required before the Data Quality Assurance Methodology can be fully realized as a piece in the overall Data Strategy To-Be vision. Throughout this document a number of these future considerations were identified. The following list highlights these previously identified considerations and gives other considerations not yet discussed.

- The Implementation Plan in Section 5 has four preliminary phases. To successfully complete these phases, FSA will need to create a detailed project plan with specific personnel assignments and actual dates for completion.
- The Enterprise Wide Data Quality Assurance Strategy requires the formation of a Steering Committee. Without a Steering Committee overseeing the quality methodology's implementation and utilization, the Data Quality Assurance Strategy will lack the resources to be successful. As FSA management selects the personnel for the Committee, they should refer to the general role requirements outlined in Section 3.1 to ensure the team members have the appropriate skills.
- The quality methodology's success is also contingent on the Steering Committee's ability to create Working Groups and assign them specific tasks and issues. Section 3.2 defines the basic roles in a Working Group. To facilitate the formation of these future Working Groups, the process for selecting Working Group team members needs to be further detailed and a set of protocols need to be defined. All FSA personnel should be informed of the selection process and protocols.
- A Tool Analysis needs to be initiated to determine enterprise audit and analytic data quality tools that will benefit the data quality effort within FSA. If COTS products are selected, marketplace tool selections need to be completed.
- The type of Quality Report must be considered. One option is for the Quality Report to be a front end reporting tool which handles queries and displays results using back-end COTS tools for data extraction and manipulation. Another option is for the Quality Report to operate independently but have COTS tools which collect information and send updates to the Quality Report's database.
- The success of the Data Quality Methodology's implementation it contingent on all pertinent FSA personnel receiving training. Section 5 outlines the high level training requirements. These requirements need to be further detailed and the possible use of training tools needs to be further researched.

- A number of the issues identified in the Data Quality Mad Dog Report will not require Working Groups since the issues will be resolved when the To-Be vision is implemented. The issues should still be tracked in the Quality Report to ensure they are successfully addressed. If parts of the To-Be vision are not implemented or delayed in their implementation, the Steering Committee may need to assign Working Groups to some of the issues.
- While the Data Quality Mad Dog Report provides an initial prioritization of a number of FSA's data quality issues, this prioritization should be re-examined by the Steering Committee. As the Committee begins their initial iteration of the Data Quality Methodology Prioritization Phase, they should develop the measures they want to use for prioritization, research the Mad Dog issue documentation, and if necessary, re-prioritize the issues.
- The functional business tools defined within the Quality Methodology need to be created into a series of standardized templates for Working Groups to utilize. This will ensure a consistent level of documentation for all Working Groups. Additionally, key templates should be identified for required completion to ensure a consistent, minimal level of documentation for each quality issue.
- The Cleaning and Scrubbing (Inflow Process) process refers to either eliminating or repairing corrupt, incomplete, incorrect, or duplicate data entering the system. If needed, these activities can be either manual corrections by a database administrator or automated updates through available commercial off-the-shelf (COTS) tools. Further research should be conducted to ensure FSA uses the most suitable option.
- There are a number of quality efforts that are underway within FSA; some systems have their own CMM program and there is discussion of an Enterprise CMM initiative, there is also an enterprise wide Solution Life Cycle (SLC) and a variety of auditing functions that the Case Management Office performs. An analysis must be done to determine the commonalities between these programs and lay out a road map for FSA in the use of each and how they are tied to the Quality Assurance Strategy.

It is essential that these future considerations be addressed; they represent some of the next steps towards realizing FSA's target state vision. Once these steps are taken and the Data Quality Methodology is fully functional in the target state, FSA will benefit from increased data quality. Increased data quality will improve the program's integrity, enable better management of the enterprise's loan and grant portfolio, and facilitate higher levels of customer service.

Appendix A: Acronyms

Due to the size of the Appendices, separate files have been created for ease of distribution. Please see file

Appendix B: Business Objectives – Raw

Appendix C: Business Objectives – Data Quality Subset

Appendix D: Data Quality Tools – Diagrams

Appendix E: Data Quality Tools – Tables

Appendix F: Data Quality Tools – COTS

Appendix G: Data Quality Methodology Overview