Army Data Management Program

Army Data Management Guide

Data Quality Management (DQM)

Version x.x

Prepared by:

Office of the Army Chief Information Officer

Architecture, Operations, Networks and Space Mission

Information Architecture Division

Arlington, VA 22202

**Revision History**

|  |  |  |
| --- | --- | --- |
| **Revision** | **Date** | **Description of Change** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Table of Contents

[1. Introduction 6](#_Toc42854434)

[1.1 Army Data Management Program 6](#_Toc42854435)

[1.2 Purpose 6](#_Toc42854436)

[1.3 Background 6](#_Toc42854437)

[1.4 Scope 7](#_Toc42854438)

[1.5 Applicability 7](#_Toc42854439)

[1.6 Benefits of Implementing Data Quality 7](#_Toc42854440)

[1.7 Audience 8](#_Toc42854441)

[2. Data Quality Management Mission 9](#_Toc42854442)

[2.1 Goals 9](#_Toc42854443)

[2.2 Meeting VAUTIS Goals 9](#_Toc42854444)

[3. Data Quality Definition and Fundamental Concepts 10](#_Toc42854445)

[3.1 Definition of Data 10](#_Toc42854446)

[3.2 Definition of Data Quality Management 10](#_Toc42854447)

[3.3 Data quality improvement cycle 11](#_Toc42854448)

[3.4 DOD Guidelines for Total Data Quality Management 13](#_Toc42854449)

[3.5 Data Quality Dimensions and Measure/Metrics 14](#_Toc42854450)

[3.6 Data Quality Business Rules 16](#_Toc42854451)

[4. Data Quality Management Implementation Framework 17](#_Toc42854452)

[4.1 Enterprise Data Quality Management Framework 17](#_Toc42854453)

[4.1.1 Prevention 18](#_Toc42854454)

[4.1.2 Remediation 18](#_Toc42854455)

[4.1.3 Reporting 18](#_Toc42854456)

[4.1.4 Data Quality Assessment and Inspection 19](#_Toc42854457)

[4.1.5 Data Quality dimensions and measures 20](#_Toc42854458)

[4.1.6 Data Quality Governance 20](#_Toc42854459)

[4.2 Data Quality Framework for Data Quality Creators/Owners 20](#_Toc42854460)

[4.2.1 Data Quality for Architecture and Design Discovery 21](#_Toc42854461)

[4.2.2 Data Quality for Existing Data Discovery 22](#_Toc42854462)

[4.2.3 Data Quality for Process and Process Technology Discovery 24](#_Toc42854463)

[5. DOD/Army Specific Requirements 26](#_Toc42854464)

[5.1 Constraints 26](#_Toc42854465)

[5.2 Ten Policy Guidelines 26](#_Toc42854466)

[5.3 Establishing the TQDM Environment 27](#_Toc42854467)

[5.3.1 Developing the strategic plan for data quality management 27](#_Toc42854468)

[5.3.2 Developing and Managing the Cultural Environment 27](#_Toc42854469)

[5.4 Scoping the Data Quality Project and Developing the Implementation Plan 28](#_Toc42854470)

[5.4.1 Identify Data Quality Project(s) 28](#_Toc42854471)

[5.4.2 Develop Data Quality Improvement Plan 28](#_Toc42854472)

[5.5 Evaluating the Data Quality Management Process 29](#_Toc42854473)

[6. Data Quality Management Implementation Guidance 29](#_Toc42854474)

[6.1 Army & DOD Technology/Tools 29](#_Toc42854475)

[6.2 Data Quality Used case 29](#_Toc42854476)

[6.2.1 Military Health System (MHS) Data Quality Management Control (DQMC) Program 30](#_Toc42854477)

[6.3 Challenges and Mitigation Strategies 31](#_Toc42854478)

[6.3.1 Team Coordination 31](#_Toc42854479)

[6.3.2 Avoid IT Having Exclusive Control of DQM 31](#_Toc42854480)

[6.3.3 Choosing the Right Data Management Tools 31](#_Toc42854481)

[6.3.4 Getting Stakeholders to Participate 32](#_Toc42854482)

[6.4 Best Practices 32](#_Toc42854483)

[6.4.1 Let Business Drive Data Quality 32](#_Toc42854484)

[6.4.2 Building the DQM Culture 32](#_Toc42854485)

[6.4.3 Performing Manual Data Auditing 33](#_Toc42854486)

[6.4.4 Identify Who is Responsible 34](#_Toc42854487)

[6.4.5 Applying DQM to Secondary Data Sources 34](#_Toc42854488)

[6.5 Key Implementation Considerations 34](#_Toc42854489)

[7. Summary and Further Reading 36](#_Toc42854490)

[8. Acronyms and Definitions 38](#_Toc42854491)

[8.1 Acronyms and Abbreviations 38](#_Toc42854492)

[8.2 Definitions 39](#_Toc42854493)

Figures

[Figure 1: Data Quality Management and the Data Quality Lifecycle 11](#_Toc42854660)

[Figure 2: Data Quality Management Cycle based on the Shewhart Cycle 12](#_Toc42854661)

[Figure 3: DOD TDQM Process 14](#_Toc42854662)

[Figure 4: DQM Implementation Framework 17](#_Toc42854663)

[Figure 5: DQM Dashbaord Example 19](#_Toc42854664)

[Figure 6:Discovery:Architecture &Design 22](#_Toc42854665)

[Figure 7: Discovery:Existing Data 23](#_Toc42854666)

[Figure 8:Discovery: Process & Process Technology 25](#_Toc42854667)

[Figure 9: TRICARE DQMC Web site 30](#_Toc42854668)

Tables

[Table 1: Data Quality Management Audience Overview 8](#_Toc42854639)

[Table 2: DOD DQM Quality Dimensions Definition 14](#_Toc42854640)

[Table 3: Additional DQM Quality Dimensions Definition 15](#_Toc42854641)

Appendices

[Appendix A References 41](#_Toc40862409)

[Appendix B Acronyms and Definitions 43](#_Toc40862410)

[B.1 Acronyms and Abbreviations 43](#_Toc40862411)

[B.2 Definitions 43](#_Toc40862412)

Executive Summary

The primary objectives of Data Quality Management (DQM) are as follows: 1) provide data that is fit for use and trustworthy in the eyes of the community of interest, and 2) maximizing the quality, utility , and integrity of information.

Data Quality Management (DQM) is important to meet the Army Strategic Plan goals of Visible, Accessible, Understandable, Trusted, Interoperable and Secure (VAUTIS) data. The DoD and the Army have addressed DQM without consistency in execution. The purpose of the Army Data Management Guide (ADMG) – DQM is to document the roles, responsibilities, policies, and procedures associated with DQM. It outlines how to establish a culture that promotes, defines, measures, analyzes, improves and controls data quality within an Army environment. Additionally, it will provide information on the business benefits, challenges and best practices associated with DQM.

This Data Quality Management Guide provides tangible business benefits for DOD and Army because it:

* Recommends on government- and industry-accepted practices of applying data quality criteria to ensure a consistent level of quality
* Facilitates Department-wide monitoring of correction and improvement activities that are not related to data correction request made
* Facilitates integration and coordination of data quality activities with other data management
* Provides common milestones and products for data quality management activities
* Provides the ability to share techniques, solutions, and resources throughout the Army organizations

The intended audience for this guide is largely Information Technology (IT) practitioners who are responsible for monitoring and correcting the quality of data in data sources owned and managed by DOD and Army G-6/CIO. IT practitioners coordinate with principal data stewards and business stewards to understand their expectations for the desirable level of quality and scope of data to be measured.

Send recommended Changes to the document to:

Senior Information Architect

Email:

# Introduction

## Army Data Management Program

The Army Data Management Program (ADMP) directly addresses challenges brought about by the Army’s use of computing technologies and large-scale software systems: managing the ever-expanding volume and ever-increasing diversity of data used in, and exchanged among, those systems and ensuring that end-users get the trustworthy and useful data that they need when they need it. The ADMP is comprised of a governing framework [1] of enterprise data management (EDM) capabilities and a set of Army Data Management Guides (ADMG) that provide requirements and guidance for implementing those capabilities within the Army.

This ADMG presents requirements and guidance for implementing a data quality management capability in the Army.

## Purpose

The purpose of this ADMG is to document the roles, responsibilities, policies, and procedures associated with DQM. It outlines how to establish a culture that promotes DQM and defines, measures, analyzes, improves and controls data quality within an Army environment. Additionally, it will provide information on the business benefits, challenges and best practices associated with DQM.

The intent of this ADMG is to state how DQM is to provide key concepts and information to understand and implement DQM. Additionally, it will outline how the Army can use DQM capabilities to meet the Army Data Plan goals of Visible, Accessible, Understandable, Trusted, Interoperable and Secure (VAUTIS) data.

This ADMG provides a “blueprint”, which specifies the standards, technology and governance procedures focusing on DQM. This ADM Guide provides architectural guiding principles and assistance to the Army Information Technology (IT) community developing DQM capabilities.

## Background

Data quality management focuses on planning, collecting, organizing, storing, processing data and disposing. DQM is boundary-spanning and provides all stakeholders with high-quality Army data by balancing their different interests. Because of these particularities of DQM, large enterprises are likely to have difficulties with institutionalizing DQM, e.g., defining accountabilities, assigning people accountable for DQM within the organizational structure, and enforcing DQM mandates throughout the Army. Organizations that are trying to get value from their data need to know that their data is reliable and trustworthy. But many factors can undermine data quality:

* Lack of understanding about the effects of poor-quality data on organizational success
* Bad or insufficient planning
* Isolated design of processes and systems
* Inconsistent technical development processes
* Incomplete documentation and Metadata
* Lack of standards and governance
* Disconnect between business and IT

## Scope

This ADMG will cover the guidelines for implementing DQM practices for Army organizations managing data assets. The scope of the ADMG DQM includes:

* Definition of DQM;
* Specify business benefits that result from the adoption of DQM;
* Specify challenges associated with implementing a DQM program;
* Outline the commercial and government best practices for applying DQM principles and technology for a project;
* Provide a framework for implementing DQM capabilities for data stores along with guidance for the Army at the senior leadership and functional levels;
* Outline the standards and technology that support an implementation framework associated with DQM;
* Provide use case scenarios featuring Army projects that are currently implementing DQM programs for their data systems.

## Applicability

The content of this guide applies to Army systems and organizations as described in this section.

The constraints specified in §5.1 apply to:

* All Army organizations and systems operating, using or associated with data quality

The requirements specified in §5.2 apply to:

* All registries containing data quality policies for Army data sources, services or artifacts

The requirement specified in §5.3 apply to:

* Data Quality Project(s)
* Data Quality Improvement plan

The requirement specified in §5.4 apply to:

* Evaluating the Data Quality Management Process

## Benefits of Implementing Data Quality

Data Quality is a perception or an assessment of data’s fitness to serve its purpose in a given context. As data becomes more intricately linked with the operations or organizations, the emphasis on data quality has gained greater attention.

Data quality benefits for implementing standard data cleansing and improvement processes that is the cornerstone of Net-Centric approach for Army for Data quality in which data shall be Visible, accessible, understandable, trusted and secure which includes:

* Manage the data quality at the time when it is received or created for realized savings in both effort and cost;
* Errors traced to root causes to prevent downstream errors;
* Improved and sustainable data quality through discovery of quality rules;
* Making metrics, audits, and other information available to stakeholders to promote collaboration, and trust on how to improve the overall process after reviewing the audit reports; and
* High-quality data leads to better decisions and better decision-making.

## Audience

Refer to Table 1 below to see how different Army individuals are involved with data quality management. For each stakeholder, there is an overview of the primary responsibilities that the stakeholder has with respect to data-quality management and any relevant sections of the document that will help them achieve that goal.

Table : Data Quality Management Audience Overview

|  |  |  |
| --- | --- | --- |
| **Stakeholders** | **Core Data Quality Management Functions** | **Relevant Sections** |
| Army Chief Information Officer (CIO)/G-6  Army Chief Data Officer (CDO)  Army Data Board (ADB) | Meet Army Data Strategy VAUTIS objectives, establish the standards & framework for Data Quality Management within the Army, and promote initiatives from an executive level. | 1 |
| Mission Area Data Officer (MADO) | Convey Army Leadership’s Data Quality management goals and ensure standardized Data Quality management initiatives across the mission area. | 1, 2, 3 |
| Data Steward | Ensure adherence to data quality standards/policies set by Army Leadership | 2, 3, 5, 6.1, 6.2 |
| Functional Data Manager (FDM) | Monitor data quality management activities in their functional area and identify challenges and/or feedback to give to senior leadership. | 2, 3, 5, 6 |
| Data Resource Manager/Developer | Leverage data quality management principles to ensure that data assets are understandable, trusted and interoperable. | 4, 5, 6.3, 6.4 |
| Data Service Design/Developer | Ensure data services are registered in the appropriate locations to determine technologies and tools that need to be incorporated for data quality management initiatives. | 4, 5, 6.3, 6.4, 6.5 |
| Data Quality Manager | Oversee the processes and procedures to improve data quality. | 4, 5, 6.3, 6.4 |
| Project/Product Manager | Organize the logistics of data quality management initiatives, including staffing and task prioritization. Ensure projects and initiatives adhere to DoD/Army data quality standards. | 4, 5, 6.3, 6.4 |
| Solutions Architect | Ensure data quality is implemented and used properly throughout system architecture. Analyze and determine the best architectures to utilize for a data quality management initiative. | 4, 5, 6.3, 6.4 |
| Enterprise Architect | Verify existing architectures have and properly use data quality. Provide the enterprise-wide viewpoint when discussing what architecture should be utilized for data quality management initiatives. | 4, 5, 6.3, 6.4 |
| Security Officer/System Administrator | Ensure security validation is properly used to protect data. Verify that policies are in place throughout all stages of the lifecycle for data quality. | 4, 6.3, 6.4 |

# Data Quality Management Mission

## Goals

The following goals are intended to steer the implementation of a Data Quality Management capability. They are not requirements or directives, but statements of desired end-states. Any means may be used to achieve these goals as long as they are within the constraints, requirements, and guidance specified in this document.

The primary objective of the Data Quality Management process is to improve the quality of mission-critical data in the major information systems that the Army owns and manages. Mission-critical data is data that is essential for the Army to conduct its business, data frequently used by the Army, data that are key to the Department’s integrity and accountability, and data used to support reports.

**End State Goal:** Army Data Quality Management practices ensure that Army data is error-free and provides trustworthy information to Army decision-makers and data consumers.

**Goal 1**: Measure current data quality levels through the use of profiling, querying, reporting tools, user interviews, logs etc.

**Goal 2**: Operational procedures and automated processes to improve data quality

**Goal 3**: Users are trained on and understand data quality rules

**Goal 4**: Monitor and report data quality levels and findings.

## Meeting VAUTIS Goals

The Army Data Plan [2] articulates a set of strategic goals that are intended to:

* make data a strategic asset for the Army enterprise,
* improve the availability and usefulness of data, and
* improve the effectiveness of information sharing across the Army.

These goals are to make data:

* Visible
* Accessible
* Understandable
* Trusted
* Interoperable
* Secure

With minor variations, these goals are the same goals expressed in the DoD Data Strategy [3] and the Army Data Strategy [4].

The contents of this AMDG contribute to meeting these goals as follows:

**Trusted**: Data quality management ensures that data assets are trusted and can be used for analysis. DQM is the collection of enterprise processes and governance that ensures that enterprise data “measure up” when data quality criteria are evaluated. Therefore, the effective operation of the Army depends on high quality data.

# Data Quality Definition and Fundamental Concepts

This chapter presents a definition of data and data quality management, standards, roles and responsibilities, and an overview of the data quality improvement process method.

## Definition of Data

The definition of Data from The Intelligence Community Data Management Lexicon adapted by Army is:

“Data is a representation of facts, concepts or instructions, such as text, numbers, graphics, documents, images, sound or video, in a form suitable for communication, interpretation or processing, which individually have no meaning by and in themselves [5]”.

## Definition of Data Quality Management

All Army organizations producing or maintaining enterprise data must incorporate a comprehensive data quality management process as part of their data production and maintenance activities. Therefore, the effective operation of the Enterprise depends on high quality data. The definition of data quality from DOI adapted by the Army is:

*“a measurement or assessment of how well data meets or does not meet Army goals based on the evaluation of criteria such as relevance, accuracy, timeliness, precision, coherence, completeness, and understandability.”*[6]

The definition of DQM is:

“It refers to a business principle that requires a combination of the right people, processes and technologies all with the common goal of improving the measures of data quality that matter most to an enterprise organization”

The primary objectives of DQM are as follows:

* **Define high-quality data**, through DQ standards, rules, and requirements
* **Assess data** against those standards and communicate results to stakeholders
* **Monitor and report** on the quality of data in applications and data stores
* **Identify Issues** and advocate for opportunities for improvement

The keys to achieving these objectives are not just procedures and tools to manage data quality but includes managing data through its lifecycle by setting standards, building quality into the processes that create, transform, and store data, and measures data against standards. Managing data to this level usually requires a data quality management team. The data quality management team is responsible for engaging both business and technical data management professionals and driving the work of applying data quality techniques to data to ensure that data is fit for consumption for a variety of purposes

Throughout the data lifecycle, data quality management activities help an organization define and measure expectations related to its data. These expectations may change over time as organizational uses of data evolve (see Figure 1, [7]). Figure 1 shows the Data Quality Lifecycle where all the phases from the planning of data, to cleaning and storing are represented.

A screenshot of a cell phone

Description automatically generated

Figure : Data Quality Management and the Data Quality Lifecycle

## Data quality improvement cycle

Data quality improvement is a proactive step to prevent poor data from being entered into information systems based on Data quality lifecycle in Figure 1. The data correction process corrects defective data; this correction is part of the cost of non-quality data. The data quality improvement process attacks the causes of defective data. Eliminating the causes of defective data and the production of defective data will reduce the need to conduct further costly data correction activities.

A general approach to data quality improvement, as shown in Figure 2, is a version of the Shewhart / Deming cycle [7]. Based on the scientific method, the Shewhart / Deming cycle is a problem-solving model known as ‘plan-do-check-act’. Improvement comes through a defined set of steps. The condition of data must be measured against standards and, if it does not meet standards, root cause(s) of the discrepancy from standards must be identified and remediated. Root causes may be found in any of the steps of the plan-do-check-act process, technical or non-technical. Once remediated, data should be monitored to ensure that it continues to meet requirements [7].

A screenshot of a cell phone screen with text

Description automatically generated

Figure : Data Quality Management Cycle based on the Shewhart Cycle

For a given data source, a data quality improvement cycle begins by identifying that data that does not meet data consumer’s requirements and data issues that are obstacles to the achievement of business objectives. Data needs to be assessed against key dimensions of quality and known business requirements. Root causes of issues will need to be identified so that stakeholders can understand the costs of remediation and the risks of not remediating the issues. This work is often done in conjunction with Data Stewards and other stakeholders (§1.7)

**In the Plan stage**, the data quality management team assesses the scope, impact, and priority of known issues, and evaluates alternatives to address them. This plan should be based on a solid foundation of analysis of the root causes of issues.

**In the Do stage**, the data quality management team leads efforts to address the root causes of issues and plan for ongoing monitoring of data. For root causes that are based on non-technical processes, the data quality management team can work with process owners to implement changes. For root causes that require technical changes, the data quality management team should work with technical teams and ensure that requirements are implemented correctly and that no unintended errors are introduced by technical changes.

**The Check stage** involves actively monitoring the quality of data as measured against requirements. As long as data meets defined thresholds for quality, additional actions are not required, and the processes will be considered under control and meeting business requirements. However, if the data falls below acceptable quality thresholds, then additional action must be taken to bring it up to acceptable levels.

**The Act stage** is for activities to address and resolve emerging data quality issues. The cycle restarts when new issues are identified and need to be assessed. A new cycle being when:

* Existing measurements fall below thresholds
* New data sets come under investigation
* New data quality requirements emerge for existing data sets
* Business rules, standards, or expectations change

Establishing criteria for data quality at the beginning of a process or system build is one sign of a mature data management organization. Doing so takes governance and discipline, as well as cross-functional collaboration.

Building quality into the data management processes from the beginning costs less than retrofitting it. It also creates a far lower impact on the organization.

It is best to do things right the first time, though few organizations have the luxury of doing so. Even if they do, managing quality is an ongoing process. It is becoming challenging to address data quality issues with volumes of data but can be nipped in the bud if an organization is attentive to the potential risks.

## DOD Guidelines for Total Data Quality Management

DoD has developed several documents that address DQM over the years. The primary document is the DOD Guidelines on Data Quality Management [8] that introduces the concept of Total Data Quality Management (TDQM) within a DOD environment. The DoD TDQM approach borrow from other TQM methodologies in that it applies quantitative method to improve products and/or services. The TDQM approach integrates functional management techniques, existing improvement efforts, and technical tools in a disciplined and focused way to create and sustain a culture that committed to continuous improvement.

The DOD DQM Guidelines outline a TDQM process, which examines ways to improve data quality to ensure:

1. consumers of data are involved in improving data quality
2. There are clearly defined requirements for excellence
3. data conforms to these requirements.

Figure 4 outlines the DOD TDQM process, which Army organizations will be using when developing their TDQM strategy.

A close up of text on a white background

Description automatically generated

Figure : DOD TDQM Process

The DOD TDQM process consists of four (4) steps, which are:

* **Establish TDQM Environment and Culture** – Create an environment that promotes a data quality culture throughout their staff and systems;
* **Scope Data Quality Project and Develop Implementation Plan** – Outline goals and objectives for DQM and who is responsible for developing an implementation plan;
* **Implement a Data Quality Project** – Implement a project that continuously defines, measures, analyzes, improves and control data quality; and
* **Evaluate DQM Process** – Evaluate how well their data quality process is meeting their goals and make corrections as needed.

The four (4) steps of the TDQM process provide high-level guidance and incorporate best practices. One best practice involves incorporating of a continuous improvement process, which will evaluate the success of the program and then making adjustments.

“§5 will discuss how the Army should implement the Total Data Quality Management Process

## Data Quality Dimensions and Measure/Metrics

This focuses on instituting operational processes for monitoring, evaluating, and reporting results. When designing and implementing control processes, remember to include actions for addressing issues found – both to current errors and to prevent future ones. One strength for DOD TDQM as mentioned in §3.4 process is the ability to measure and evaluate the level of data quality for a data source and identify potential improvements to exhibit the characteristics of reasonable metrics defined in the context of the types of data quality dimensions .This will help organizations define and measure quality that increases the value of data assets.

Originally, the term “ACTS” **A**ccuracy, **C**ompleteness, **T**imeliness and **S**tandards was associated with quality dimensions. The DOD DQM Guidelines document expands on the concept of ACTS and defines six quality dimensions (Table 2) for Army organizations.

Table : DOD DQM Quality Dimensions Definition

| **Quality Dimension** | **Definition** | **Example Metric** |
| --- | --- | --- |
| Accuracy | Information reflects the real world entity, relationship or event it is intended to represent | * Error ratio * Deviation |
| Completeness | All necessary values for a particular data element are included so that the data set is complete | * Percent of data fields having values entered into them. |
| Consistency | Information about an entity, relationship or event is consistent across data elements, a database, data set, information system, and/or organization | * Range * Variance * Standard deviation |
| Timeliness | The age of the information is appropriate for its intended uses and available for output when required | * Time variance |
| Uniqueness | Information does not reflect unnecessary duplication | * Number or percentage of repeated values |
| Validity | Information conforms to specific formats. | * Percentage of data records where. All values are in required format |

Table 3 shows additional DQM quality dimensions that Army organizations can use in addition to the six DoD DQM quality dimensions. These dimensions are promoted by industry experts but are optional for Army organizations.

Table : Additional DQM Quality Dimensions Definition

| **Quality Dimension** | **Definition** | **Example Metric** |
| --- | --- | --- |
| Sound (Valid) | The data is error-free (at time t). The dimension encompasses most “common” data quality dimensions. It is mostly objective and does not depend in use. | * Value Range Constraint * Schema-Valid * Accurate * Complete * Consistent/Nonredundant * Reasonable |
| Comprehensive (Complete) | The data meets all business process information requirements (for purpose p of stakeholder s at time t). This dimension is subjective and is the “fit-for-purpose” goal found in data quality literature. | * Information Completeness * Relevancy * Representational Fidelity (e.g., hand-drawn tree vs photo of a tree) |
| Usable (Accessible) | An IT/data. Processing characteristic of how timely and easily the data can be accessed by a stakeholder. | * Service/API availability * Timeliness * Security * Interface standardization |
| Data Model Quality | A bad design leads to bad products. Data model\* quality is how well a data model measures up as a complete, clear and unambiguous specification. | * Completeness meeting stakeholder info requirements * Complete documentation * Clear definitions * Adherence to data model guidance * Current |
| Data Management Quality (Lifecycle/Data Processing Quality) | An enterprise function performance characteristic of how well data is managed (and quality maintained) across processing operations (such as data translation). | * Lineage * Translation accuracy * Synchronization performance |

## Data Quality Business Rules

In the DoD, TDQM (§3.4) is a process to support database migration, promote the use of data standards in conformance to business rules. The process of instituting the measurement of conformance to specific business rules requires definition. Monitoring conformance to these business rules requires:

* Segregating the data values, records, and collections of records that do not meet business needs from the ones that meet business needs.
* Generating a notification event alerting a data steward of a potential data quality issue.
* Establishing an automated or event driven process for aligning or possibly correcting flawed data within business expectations.

The first process uses assertions of expectations of the data. The data sets conform to those assertions. More complex rules can incorporate those assertions with actions or directives that support the second and third processes, generating a notification when data instances do not conform, or attempting to transform a data value identified as being in error.

Other types of rules may involve aggregate functions applied to sets of data instances. Examples include validating reasonableness of the number of records in a file, the reasonableness of the average amount in a set of transactions, or the expected variance in the count of transactions over a specified timeframe.

Providing rule templates help bridge the gap in communicating between the business team and the technical team. Rule templates convey the essence of the business expectation. It is possible to exploit the rule templates when a need exists to transform rules into formats suitable for execution, such as embedded within a rule’s engine, or the data analyzer component of a data-profiling tool, or code in a data integration tool as mentioned in §6.1.

# Data Quality Management Implementation Framework

A data quality framework is defined as ‘ a vehicle that an organization can use to define a model of its data environment, identify relevant data quality attributes, analyze data quality attributes in their current or future context, and provide guidance for data quality improvements [9].

Data quality framework needs both objective and subjective attributes of the domain modeled by the framework to be considered to reflect the contextual nature of data quality and the many potential users of the data The purpose of this section is to provide an implementation and governance framework for developing a DQM solution that incorporates the concepts and best practices outlined in this ADMG.

## 4.1 Enterprise Data Quality Management Framework

The Total Quality Management [8] provide information for a TDQM process(§3.4) but do not address standards or technology needed to develop a complete solution. Figure 5: DQM Implementation Framework outlines the functional areas associated with implementing a solution that supports the DOD TDQM process. These functional areas provide information on the capabilities within each functional area, along with how they support the DOD TDQM process and goals. The intention is to address the issues when implementing a TDQM solution. This framework should be used to establish data quality standards for other data deemed critical by the Data Quality Management Team as mentioned in §3.2. The DQM Implementation Framework includes:

* Prevention
* Remediation
* Reporting
* Data Quality Inspection and Assessment
* Data Quality Dimensions and Measures
* Data Quality Governance

A screenshot of a cell phone

Description automatically generated

Figure : DQM Implementation Framework

### Prevention

One of the requirements for DOD TDQM process is the ability is to build the business case to allow data quality managers to see how data quality is important to the organization, and estimate how much effort it will take to address current data quality issues which includes:

* Business Process Analysis: It defines the scope and objectives and the business will gain from the project and how it will gain the value, and how much the value will exceed the investment necessary to realize that value
* Technology Analysis: It can be used in many ways to accelerate a data quality project to define data quality of the organization, facilitate data profiling and help keep data good after remediation.
* Training: It can be used to promote data quality culture where all staff is responsible for data capture, analysis and data reporting to be aware of the aware this role has on others.

### Remediation

Remediation planning is the pivotal point of the project so that all information is drawn together to create the desired future state. This takes the output from the remediation planning and starts the remediation work which is the transformation of the data, process and architecture of the organization to one where quality is at the fore and further includes:

* **Root cause analysis**: The work to determine the underlying cause of the data quality issues
* **Issue Adjudication**: This covers taking the results from the profiling and the remediation planning and remediating the process and will include process designs, process remediation, process and change management and integration into business as usual.
* **Cleansing**: This includes the results from the profiling and remediation planning and remediating the data. This will include replacement of data, rectification of data and, potentially, deletion of data that is no longer required by the organization

### Reporting

Dashboard and analytics reports provide tools that allow Army organizations to monitor the results of the data quality process. This supports the DOD TDQM requirement to measure and evaluate the results of the process and make suggestions for improvement. The dashboard allows senior leadership and data stewards to monitor the process using summary reports and graphs. They can then drill down to detail reports as needed. Additionally, dashboard and analytics reports support the DOD TDQM process by providing:

* Data score cards summarizing the metrics and issues by business rules and quality dimensions;
* Alerts on data errors that require special attention by the system administrator;
* Reports that show trends on data errors found while processing data;
* Graphic reports that show trends on data errors including Time-Series and Pareto reports;
* Reports that summarize results which Army organizations can use for their audit reports; and
* Standard and ad-hoc reports, which Army organizations can use to pinpoint the root cause of a data issues.

Figure 5: DQM Dashbaord Example provides an example of a DQM dashboard. It provides a snapshot of the data the system is processing and the results of the data quality capabilities.

A screenshot of a social media post

Description automatically generated

Figure : DQM Dashbaord Example

The DQM dashboard has three (3) sections each providing unique information. The top section contains a list of the data element profiles used to analyze data along with a pie chart summarizing the percentage of profiles by quality dimension. The middle section contains a list of business rules used to analyze and improve data along with search and save capabilities. The bottom section contains pie charts showing the percentage of data issues found by quality dimensions along with details on the business rules used to analyze the data. Each report is useful for system administrators and senior leadership to monitor the results of the DQM process.

### Data Quality Assessment and Inspection

The objective of evaluating the DQM process is to review data quality goals/benefits and to improve the processes used to manage data quality. An assessment of progress made in implementing data quality initiatives and/or projects is required. Current DOD guidance encourages the participants in the TDQM process (functional leadership and IT administrators) to review progress with respect to: 1) modifying or rejuvenating existing methods of data quality management and/or 2) determining whether data quality projects have helped to achieve demonstrable goals and benefits.

The evaluation and assessment of data quality efforts reinforces the idea that TDQM is not a program but a new way of doing business. In terms of evaluating and assessing progress made on data quality, the functional leadership is encouraged to review both the costs and benefits associated with the data quality projects.

Additionally, a periodic performance of general data management maturity assessments or focused data quality management maturity assessments should be done. Army organizations should establish long-term data quality goals to evaluate the process over time. These reports should be maintained with the other audit reports and made available to members of the organizations involved with processing and managing data.

### Data Quality dimensions and measures

One of the requirements for DOD TDQM process is the ability to measure and evaluate the level of data quality for a data source and identify potential improvements as mentioned in §3.5.

### Data Quality Governance

Army Data Management created the Chief Data Officer (CDO) position, the Army data steward role and the Army Data Board (ADB). Data Governance is a people and process function, and the most important part of implementation is to gain buy-in from key stakeholders across the organization, and to put in place a framework that will work with the individual organization and can be sustained. The Army Data Board foster a collaborative governance environment for achieving Army Data Strategy goals, with active participation from across the Army.

#### DQM Governance at the Army Senior Leadership Level

At the senior leadership level, DQM governance requires two (2) key roles, which are:

* Data Quality Committee (DQC) - Responsible for the overall data quality strategy by defining DQM initiatives, implementation plans, and reconciling DQM issues between Army organizations as mentioned in §4.2; and
* Senior Data Quality Stewards (SDQS) - Defines the key performance indicators and measurement methods for DQM and is responsible for quality standards, processes, and guidelines.

The ADB has several senior level data quality stewards as voting members on the governance board who can support the ADB in the development of key performance indicators, measurement methods and guidelines for DQM. Additionally, the data stewards can work with Army organizations to promote the DOD TDQM process and incorporate lessons learned into the process.

Further details will be mentioned in §6.3.

## Data Quality Framework for Data Quality Creators/Owners

One of the requirements for DOD TDQM process is the ability for analysis and assessment of the overall architectural organizational design, the familiar task of profiling the existing data , and also the less familiar task of business process assessment , where the project looks at the way business processes contribute to poor quality data. The reason, this stage is placed here is that assessment should be made in full knowledge of the information requirements of the organization, otherwise if not likely much time will be wasted. Discovery Stage includes 3 phases:

* Discovery: Architecture and Design
* Discovery: Existing Data
* Discovery: Process and Process Technology

The discovery stage for architecture, data and process will be explained below. The discovery stage for architecture, data and proceed can proceed simultaneously. The skill sets are different for different for each “leg” and hence different resource would be used as a matter of course. This lends itself to a very efficient project execution and is far better than analyzing each section sequentially.

### Data Quality for Architecture and Design Discovery

At the start of the discovery phase, the project needs to engage in existing architecture function and make sure for architecture and process the assessment is for risks to good quality, and for data it is the actual analysis of the data itself, directly measuring the data quality [10]. The steps include:

* **High level definition and scope**: Determine the scope in terms of domains, systems and business divisions to create high level overviews of the architecture. Architecture definition documents. Should be available to the data quality practitioner in any organization that has an enterprise or data enterprise architecture function but since the documents are under-developed the data quality practitioner may to create the documents
* **Plan:** Make a high level assessmentat an organization and business-level of how the existing architecture will affect the overall data quality of the organization
* **Detail Investigation and Description:** High level documentation for architecture may exist but the practitioner need to do a proper assessment to flesh out the understanding created in the definition phase by delving into the technology to see where data feeds come and go. The architecture practitioner will draw on all the initial investigation undertaken in the initiation stage to get more information on the problem that they identified
* **Assess Architecture**: For each part of the architecture, the data quality practitioner needs to ask the question: “How does the architecture affect the quest to improve data quality?” Here the data quality practitioner needs to identify any situation where the architecture places a hurdle over getting things right first time. The objective is to understand why the organizational design is in its current incarnation, what effect it can have, and eventually what can be done , if anything , to fix it.
* **Deliver Assessment Results:** Collate the issues under the four architecture headings, specifically business, data, application and infrastructure. At this point themes should start to emerge in the way the organization approaches information architecture and information change.

Figure :Discovery:Architecture &Design

### Data Quality for Existing Data Discovery

One of the requirements for DOD TDQM process is the ability for engaging the existing architectural function within the organization and the same applies for data. The project needs to discover data sources in the organization, to assess and profile them and to collate and present the results of work [10] . The steps include:

* **High level Definition**: Defining the scope of the problem is necessary before an organization is able to define an approach to solve it to build a picture of the data landscape and where major data store exist
* **Plan:** After identification of source systems, the next stage is to see the state of data and how it measure up against the defined data requirements
* **Profile Data:** Nature of the detail task within profiling will depend on what the organization considers to be its main concern for discovering data quality
* **Assess Profiling Results**: Assessment work will collate the issues into groups, covering headline issues and potentially those issues requiring further analysis. At this point the root causes of a number of issues may be discovered as key stakeholders recognize what caused the problem. This feedback is vital in truly understanding the root causes and planning remediation.
* **Deliver Assessment Results:** The output from profiling assessment will provide insight into the process within the organization.

Figure : Discovery:Existing Data

### Data Quality for Process and Process Technology Discovery

One of the requirements for DOD TDQM process is the ability to engage existing architecture and data functions, and the same applies to process and make sure the data stays good to examine the processes that feed data into their data stores. This step starts with identification of existing processes and discovery of how and why processes are failing to supply organizations with good quality data.

Analysis of process is undertaken in two ways. First, the processes themselves can be examined through data quality lens and issues identified where the process is failing to deliver good quality data. This may be through lack of control, overly complex process, or poor technology. Secondly, the data coming out the processes can be examined to see where process failures have occurred.

The steps include:

* **High level definition and Scope:** Processes within an organization that are relevant for data quality are those that create/capture, transport, transform(update/delete) or read/use data. Especially relevant to data quality are the data creation processes, as an error at this stage will cascade through the organization resulting in errors at all points downstream
* **Plan:** Define those in scope for the project by CRUD Matrix technique which can be looked at from the perspective of data quality , and then cut further into the processes that can cause data quality issues which will leave a small subset of the total processes within the organization. To establish the scope that is necessary for data usage , the scope is again best factored towards elements which cause the greatest impact-specifically regulatory reporting and board level decisions
* **Detail Investigation and Description:** The next step is process mapping .The objective of process mapping is to gain understanding of the process and to help identify data quality issues. It also translates into a portable format which can then be utilized together with the results from the architecture and data discovery work to plan the remediation.
* **Processed Focused Assessment**: A process focused approach looks at the process as it occurs in the organization from the perspective of data quality. It looks to where the process does not facilitate the creation of good quality first time, or does not have controls in place to pick up errors or omissions that degrade data quality. A data quality process assessment should look for:
* Lack of data quality checks
* Lack of duplication checks
* System constraints
* System failures
* Poor or no validation

The aim here is to discover how processes affect data quality, the identification of any constraints and the identification of how the quality of input is managed

* **Data Focused Assessment:** The existing data is an often-fascinating map telling you where data quality issues exist in the upstream process. Within the analysis of existing data there will be fields that are of good quality, fields that are of bad quality and some in the middle. Whilst a process-based approach is vital for completeness, the existing data should always be checked as it can be relatively quickly interrogated solutions that will improve data quality with little effort
* **Process Assessment Deliverable:** The output from data-focused assessment and process focused assessment will provide insight into the process within the organization.

Figure :Discovery: Process & Process Technology

# DOD/Army Specific Requirements

A clear statement of policy, disseminated throughout the organization, must be in place for a company to remain engaged and to succeed in maintaining a viable, continuing data quality effort, which in turn proactively supports business activities. The preceding chapters have addressed a wide range of issues that must be faced and tasks that must be undertaken by any organization with the intention of seriously and proactively pursuing the goal of attaining and maintaining high-quality information and data.

Many organizations have started the journey but have not been able to sustain the effort and keep the data quality activities going. Too often the activities are recognized by a single champion or a single department for their benefits that can be obtained from such activities. The result is that the process is not institutionalized, and inertia within the rest of the organization leads either to a slow decay or an abrupt halt to all data quality initiatives. The key to success in the data quality journey is the creation of appropriate organization policies that involve all functions and activities related to the maintenance of the data.

This section details DoD/Army requirements for establishing data quality management. Several standards and policies referenced in this section are found in the DoD IT Standards Registry (DISR).This section includes how the Army would implement TDQM.

## Constraints

Below is a list of constraints that must be followed when pursuing the end state goal:

* Do not break any laws, regulations or policies (LRP) set forth by the Federal government, the DoD, or the Army.
* Do not register unreliable and untrustworthy data assets into the enterprise-wide catalog without proper validation checks

## Ten Policy Guidelines

In developing a data quality specific to an organization, an overarching objective is to put in place a policy that will lead to the continual improvement of the overall quality of the data for use. The questions that must be addressed are many. In what areas is improvement most important?. Clearly, some of these areas will be business areas and business processes; others will be in the information technology itself. Some of the existing organizational policies may have to be changed to ensure that the firm produces and uses data of high quality.

The provenance of data quality policy should be the business side, not the information technology side, of the firm. There is no question that information technology departments are concerned about such matters as procedures for backup and recovery and establishing audit trails for changes. But a broader business perspective must be presented when establishing data quality policy for the organization.

The ten policy guidelines that are proposed as the basis of an organization’s data quality policy are distilled from the observation of data quality practices of the root conditions of data quality problems.

* The organization adopts the basic principle of treating information as product, not by-product.
* The organization establishes and keeps data quality as part of the business agenda
* The organization ensures that the data quality policy and procedures are aligned with its business strategy, business policy, and business processes.
* The organization establishes, clearly defined data quality roles and responsibilities as part of its organization structure.
* The organization ensures that the data architecture is aligned with its enterprise architecture
* The organization take a proactive approach in managing changing data needs.
* The organization has practical data standards in place.
* The organization plans for and implements pragmatic methods to identify and solve data quality problems and has in place a means to periodically review its data quality and data quality environment.
* The organization fosters an environment conducive to learning and innovating with respect to data quality activities.
* The organization establishes a mechanism to resolve disputes and conflicts among different stakeholders.

## Establishing the TQDM Environment

Based on the Ten policy guidelines mentioned in §5.2 the first step in TQDM process (§3.4), is establishing the data quality management environment. Establishing the environment includes management buy-in and the cultural conditions that encourages teamwork between functional and information system professionals. All too often, functional users of an AIS know the data quality problems that afflict an automated system but do not know how to systematically improve the data. In parallel, information system professionals know how to identify data quality problems but do not know how to change the functional requirements that drive the systemic improvement of data. This is accomplished by:

* Developing the strategic plan for data quality management; and
* Developing and managing the cultural environment

### Developing the strategic plan for data quality management

Data quality management responsibilities fall under the DoD information management and data administration initiatives (DODD 8000.1 and DODD 8320.1). The Functional Data Manager (FDM), Data Resource Developer are responsible for developing data quality goals, objectives, and action plans for their respective organizations as part of their contribution to the Army strategic plan and include:

* Overall goals and objectives for data quality management
* Strategies and projects to achieve data quality goals and objectives
* Measurable data quality objectives

### Developing and Managing the Cultural Environment

Actions plans established by the Functional Data Manager (FDM), Data Resource Developer also address the infrastructure requirements to meet data quality objectives. Infrastructure needs include developing responsibilities for improving data quality, establishing training programs and/or initiatives within functional areas, opening lines of communication between functional experts and information system professionals about problems and solutions to poor data quality, and promoting functional and information system improvements brought about by leadership to correct data quality problems by the Data quality Manager.

## Scoping the Data Quality Project and Developing the Implementation Plan

One of the major features of the TDQM approach that is promoted within DOD is the central focus on initiating and completing data quality projects. The essential requirement is to:

* Identify data quality improvement project(s) that can be successfully worked; and
* Develop an implementation for each project

The essential requirements are responsibility of the Data Quality Manager.

### Identify Data Quality Project(s)

Data quality projects are selected by users and/or System administrators. It is good business practice to listen to both the functional and AIS community. For example, users often report frustration with errors in the data recorded in system tables and/or records. Known inaccuracies in queries, reports, and data correlation problems may be good indicators of data quality issues. Second, system administrators may make recommendations based on known problems with data collection, processing errors, and internal audit and validation procedures.

Additional factors that may influence the selection of data quality projects include focusing on areas that provide the greatest opportunity for success and prototyping/demonstrating the value of data quality efforts to achieve management buy-in.

* Choose Efforts that are Opportunities for Success: The success or failure of initial TDQM efforts or projects can greatly affect how easily the organization adopts TDQM ideas. Select projects:1) that have a high chance of success, 2) that have the highest failure costs, and 3) where significant improvements can be made. Projects that address critical data quality issues can be solved with the minimum of effort will increase the attractiveness of TDQM to top management.
* Prototype Effort: If there is not top management support for data quality efforts, perform a pilot project demonstration. Choose a data quality project with low risk and low visibility that is critical to the organization’s success. Select an initial effort that is neither so large that it is doomed for failure from the start, nor so small that improvements will essentially go unnoticed.

### Develop Data Quality Improvement Plan

Implementation plans are management documents that scope a data quality project in terms of project objectives, tasks, schedule, deliverables and resources. From a project management point of view, implementation plans developed by data steward provide information on:

* **Task Summary**: List project goals, scope and synopsis of anticipated benefits
* **Task Description**: Describe data quality tasks
* **Project Approach**: Summarizes tasks and tools to be used to baseline data quality
* **Schedule**: Identifies task start, completion dates, and project milestones.
* **Deliverables**: List reports and/or products that document the result of a data quality project. At a minimum, deliverables should include:

1. **Data Quality Baseline Assessment**: Document current data quality problems. Include exception reports on data that does not conform to established standards or business rules.
2. **After Action Report**: Technical report on the data quality improvements that were implemented. Include description of actions taken to improve data quality and rationale for taking the actions and the lessons learned and improvement metrics.

## Evaluating the Data Quality Management Process

Further , DoD TDQM process is the evaluation and assessment of progress made in implementing data quality initiatives and/or projects. Current DoD guidance encourages the participants in the TDQM process to review progress with respect to: (1) modifying or rejuvenating existing methods to data quality management and/or (2) determining whether data quality projects have helped to achieve demonstrable goals and benefits.

The evaluation and assessment of data quality efforts reinforces the idea that TDQM is not a program but, a new way of doing business. In terms of evaluating and assessing progress made on data quality, the Functional Data Manager (FDM), Data Resource Manager are encouraged to review both the costs and benefits associated with the data quality projects.

# Data Quality Management Implementation Guidance

This section of the document will describe Army & DOD technology/tools, challenges, best practices and key considerations related to the implementation of data quality management. This section is primarily concerned with data quality management from the perspective of the creators and owners. Guidance for implementing enterprise data quality management is mentioned as follows:

* **A set of goals**: “Focus resources on DOD/Army G-6 CIO’s data quality objectives.”
* **A value system**: “We value our data consumers.”
* **A mindset**: “Data are key sharable assets and are used in important information products for citizens.”
* **An environment that promotes continuous improvement**: “To eliminate the waste associated with process failures and rework caused by defective data.”

## Army & DOD Technology/Tools

The purpose of this section is to provide tools, which demonstrate how DOD and Army organizations have DQM capabilities to support validation, profiling and cleansing of data needs based on §3.3.

Data validation tools examine data as it is input into the system and rejection or correct data item errors. Extract-Transform-Load (ETL) tools can sometimes be configured to perform validation functions as the. External data is prepared and entered into an existing data set.

Data profiling or data auditing tools examine a data set to identify problems, such as missing, duplicate, inconsistent, and otherwise anomalous data, and also compute data quality metrics.

Data cleansing tools go through an existing dataset and attempt to detect, correct, or remove troublesome data items.

## Data Quality Used case

The purpose of this section is to provide used cases, which demonstrate how DOD and Army organizations have DQM capabilities to support their data collection and dissemination needs based on §3.3.

### Military Health System (MHS) Data Quality Management Control (DQMC) Program

The MHS is the program within the DOD responsible for providing health care to active duty and retired U.S. Military personnel and their dependents. Military Health System Data Quality Management Control Procedures outlines the implementation of a DQMC Program. The instructions incorporate concepts outlined in the DOD TDQM process §3.4. It required each Military Treatment Facility (MTF) to submit complete, accurate, and timely data in compliance with DOD and MHS/TRICARE Management Activity (TMA) data collection and reporting requirements. Each MTF had to establish and effectively operate the DQMC Program to provide the internal structure to improve and better assure data accuracy, completeness, and timeliness and to assure uniformity and standardization of information across the MHS. The DODI 6040.40 outlined the responsibilities, procedures, and information requirements that support the DQMC program.

TRICARE established the DQMC workgroup to provide senior leadership for their program and execute the responsibilities of the DODI 6040.40 including:

* Identification of policy and information system shortfalls impacting the implementation of DODI 6040.40 at the service and MTF level;
* Coordinate with other MHS workgroups and TMA directorates as required to communicate shortfalls and provide recommended solutions;
* Ensure the Data Quality Management Control Review List and Data Quality Statement are updated to reflect policy changes and standards as required;
* Serve as the point of contact for Government Accountability Office/Department Of Defense Inspector General (GAO/DODIG) and other external agencies on all DQMC issues; and
* Provide oversight and support in conducting the MHS Data Quality Course and provide training and education on implementation of DODI 6040.40.

TRICARE is using their public web site to support their DQMC program (see Figure 9: TRICARE DQMC Web site) and promote a TDQM environment within their organization.

A screenshot of a social media post

Description automatically generated

Figure 9: TRICARE DQMC Web site

It contains references to the DOD TDQM regulations and other materials applicable to the DQMC. Staff members can use the web site to access training materials and sign up for courses.

The DQMC workgroup has members from all DOD Services. Patient Administration Systems and Biostatistics Activity (PASMA) implemented the procedures outlined in the DQMC program for their patient information provided by several systems. They established a governance structure that includes a Data Quality Manager to oversee the entire DQMC program within PASMA. The Data Quality Assurance Team is responsible for executing the DQMC procedures and developing the audit reports. The team consists of members at command level that are responsible for DQM within their systems. PASMA has developed documentation and training materials for their staff to promote a TDQM environment.

## Challenges and Mitigation Strategies

The following are a set of challenges which could prevent an organization from adopting the DOD TDQM process or minimize the success of implementation.

### Team Coordination

The structure of traditional data management gave senior management the authority over the information used within business units, while each business unit manager had its own requirements for data quality. When senior management imposes a DQM process and set of standards, there is often push back or hesitation by business unit managers to invest time and resources in addressing DQM issues not relevant to them.

This leads to organizations customizing DQM requirements in ways that make them incomplete or inconsistent, or ignoring them altogether. Business unit managers have the practical knowledge about data to contribute to the overall DQM process. Coordination between senior and business unit managers will promote an environment that encourages teamwork. TDQM should be used as the initial baseline for data quality metrics/requirements and then have a process to enable flexibility as mentioned in section 3.6.The DOD TDQM Process addresses this challenge within several of its steps as mentioned in §3.4 and §5.3.

### Avoid IT Having Exclusive Control of DQM

The main purpose of data is to support the business/mission needs of Army organizations. Often the IT personnel entrusted with the implementation of applications assume ownership of the data used within that system. This introduces potential conflicts when functional users participate in enterprise-wide data initiatives and/or expose their information management process to data quality audits and reviews. IT personnel often assume any DQM issue is an IT issues and their responsibility to address. However, functional business units have knowledge on how the data supports their business/mission needs. The DOD TDQM process(§3.4) indicates that functional users of data should be involved in the process for improving data quality. Communities of Interests (COIs) have unique DQM needs for the data they consume. The DOD TDQM process addresses this challenge within several of its steps as mentioned in §3.4.

### Choosing the Right Data Management Tools

The case study in the DOD DQM Guidelines [4] shows how the right tools can achieve the DQM goals for an organization. However, a frequent response is to purchase an automated data cleansing or profiling tools without scoping their DQM requirements. While data quality tools do provide some benefits right out of the box, buying a tool without a well-defined understanding of specific data quality problems will not have a significant return on investment nor will it achieve long-term strategic goals.

### Getting Stakeholders to Participate

Getting stakeholders to participate can be difficult. Many times, when a data consumer discovers bad data, they hold themselves accountable to fix the error instead of reporting the issue back to the data provider. The primary concern is to fix the data necessary to execute their mission goals. This causes several problems. Others are still using the bad data and the data consumer assumes they have correctly fixed the issue. While it is acceptable for the data consumer to fix the data, should always report the issue to the data provider.

The DOD TDQM process calls for open communication between all stakeholders. The data provider should encourage all stakeholders, not just data consumers, to report data issues. The Army organization should develop an official data issue report and process to report data issues. A point of contact should be available to receive and process the reports. Army organizations could use a portal to automate the data issue reporting process. No matter the method, Army organizations should develop training materials on the data issues reporting process and share with all stakeholders.

## Best Practices

The DOD TDQM process includes several best practices such as:

* Let Business Drive Data Quality
* Keys to build the DQM Culture
* Perform Manual Data Auditing
* Identify who is Responsible
* Applying DQM to Secondary Data Sources

### Let Business Drive Data Quality

Most organizations react to data quality events instead of determining how to prevent problems from occurring in the first place (i.e.,” First Time Correct”). An example is fixing an address for a shipment after the customer complained the package went to the wrong location. A mature data quality project determines where the risks are, what the objective metrics are for determining quality levels, and the impact of data quality compliance before data quality issues appear. Army organizations should be proactive towards DQM by:

* Establishing a DQM program based on the DOD TDQM process;
* Determining how data quality issues will be handled during the development of a DQM process;
* Monitoring data quality at every stage where data is touched; and
* Creating a DQM dashboard to monitor the data quality performance measures.

### Building the DQM Culture

The first step of the DOD TDQM process is to establish an environment that promotes a culture where all stakeholders are involved in data quality. It covers the need to provide strong leadership in the development of a strategic plan along with the importance of training programs and materials to achieve this goal. However, Army should address several topics when establishing the DQM culture and developing training materials.

The second topic to address is “Army-Wide Thinking”. Army organizations think only within their silo. The DOD TDQM process covers the need for opening lines of communication between all stakeholders when developing requirements. Army organizations should include data suppliers and data consumers when developing of their DQM requirements. Both will provide valuable feedback based on their experience collecting and using the data. Using existing COIs would be helpful in establishing a DQM culture and distributing training materials.

The third topic to address is “Understand how data is being used to make decisions.” Decision makers use both fact-based (data) and intuition methods to make decisions when performing their missions. The Army favors the use of fact-based decision making; however, if there is a lack of trust in the data the decision maker has a tendency to rely on intuition. The more trustworthy the data the more likely the decision maker will use it to make fact-based decisions. An example is an Army commander will use data collected by Warfighters to plan specific missions. Bad data used to plan a mission could cause an Army unit to hit the wrong target with the loss of Army personnel lives.

The Army should make the development of training programs and materials that address these topics as part of their strategic plan.

### Performing Manual Data Auditing

Tools that perform data validation (§3.5,3.6) and cleansing are very useful for any DQM process. The DQM implementation framework incorporates automated tools that perform data profiling and auditing when processing data. However, Army organizations should not rely exclusively on these tools and should perform manual audits on a regular basis. Army organizations should use the audit results to confirm whether their DQM procedures are meeting their goals and make adjustments as needed. Army organizations like the USAMC LOGSA have incorporated manual data auditing to support their TDQM process. The steps involved in the manual auditing process are: 1) planning, 2) execution and 3) reporting.

#### Planning an Audit using Data Profiling

The process for planning an audit starts with data profiling to help clarify the structure, content, relationships and derivation rules of a data source. Data profiling helps the auditors identify and understand data anomalies to focus on for the audit. Army organizations should concentrate on well-populated data elements as potential candidates for their audit. Minimally populated data elements should not be included as potential candidates.

All stakeholders, including data consumers, should be involved in the planning process. Army organizations should share the results of the data profiling analysis with all stakeholders. The stakeholders should review each data element identified and provide feedback. The auditors and stakeholders must agree on the data elements and business rules to use before the audit can proceed.

#### Execution of the Audit

Auditors need instructions and support information to execute the audit correctly. An audit package contains the overall objectives along with the scope, approach, auditors’ responsibilities and contact information. Details on the data elements selected, along with sample data are included in the audit package. The sample data should contain examples of the issues discovered during the data profiling phase. The audit package should also include the business rules to apply to each data element. Additional instructions on how auditors record the audit results should be included in the package to prepare for the next step.

#### Reporting the Audit Results

The last step is to write an official report documenting the audit results. The audit report is an official deliverable that should use the Army organization’s official format and structure. The first section should include an introduction, objectives, scope, approach, and responsibilities similar to the audit package. The second section is the detailed audit results, listing the data elements selected with a description of the business rules applied during the audit. Additionally, the report should contain audit statistics at the data record and data element level. The third section contains the audit findings, which includes summarized reports on the data errors found along with analysis on what caused the problems (process/procedural/systematic). The last section includes the recommendations to solve the issues found during the audit. The evaluation step of the DOD TDQM process should consider the recommendations found in the report.

### Identify Who is Responsible

All personnel involved in processing data should be involved in the DQM process. However, Army organizations should appoint a data steward who is responsible for executing the data quality program. This will provide an explicit focal point for data quality efforts by knowing who to hold responsible. They can also establish a data quality department to assist the data steward. The data steward will be the primary person to develop and execute the action plan, DQM training materials and manage the governance issues.

### Applying DQM to Secondary Data Sources

The Army should apply TDQM to all ADSs. Nevertheless, they should not ignore applying TDQM to secondary data sources. Secondary data sources support ADSs in several scenarios. The first example is when Army organizations, at the tactical edge, use local data sources to collect data and then feed the data to an ADS. Army organizations that support this scenario should be aware of the data quality requirements of the ADS and ensure their system supports these requirements. They should also report any issues with the data collected to the ADS so they can build business rules to analyze and correct data once it is uploaded.

Another scenario is Army organizations that cache data to a local data store from an ADS. Their users may discover issues with the data pulled down from the ADS. As stakeholders of the data, any data consumers should report all issues back to the owner of the ADS to ensure others are using corrected data.

Army organizations that manage a secondary data source may not have the budget to support all TDQM steps and procedures that an ADS owner could support. At a minimum, they should be an active participant within the COI. In addition, secondary data source owners should participate and/or use training materials provided by the owner of the ADS. This will allow them to be aware of data quality requirements from the ADS.

## Key Implementation Considerations

Data quality management is a cyclic process that involves logical step-by-step implementation. Such quantifiable steps can help in standardizing solid data management practices that can be deployed in incremental cycles to integrate higher levels of data quality techniques into the enterprise’s architecture [11]. There are number of considerations when looking at implementing improved approaches to data quality management:

* **Data. Quality Assessment** – An independent assessment of the quality of data is of prime importance to identify how poor-quality data hampers the business goals. It provides a referral point to invest and plan in data quality improvements and also measure the outcomes of successive improvements. The data assessment must be guided by an impact analysis of data on business. The business-criticality of data must be an important parameter in defining the scope and priority of the data to be assessed. This top-down approach can be contemplated by the bottom-up. Strategy of data. Profiling based assessment which can identify anomalies in data and then map these anomalies to the potential impact on business goals.

This phase must be completed with a formal report which clearly lists downthe findings. The report can be circulated amongst stakeholders, decision makers and hence drive data quality improvement actions.

* **Data. Quality Measurement** – The result of data assessment report helps to narrow down the scope to identify critical data elements. The attributes and dimensions for measuring the quality of such data, defining the units of the measurements and laying down the acceptable thresholds for these metrics is the basis of implementing improvement processes. Attributes such as completeness, consistency, timeliness are defined that act as an input to deciding the tools and techniques that should be deployed for achieving the desired levels of quality. Data validity rules are specified based on these metrics. This can help to press data. Controls into the function that acquire or modify the data within the data lifecycle.

Data quality scoreboards and dashboards can be defined for each business unit derived from these metrics and their thresholds. These. Scores can be captured, stored and periodically updated to monitor the improvement.

* **Incorporating Data Quality into the functions and processes** – Focus on building the functionality takes precedence over data quality during any application development or system upgrade. The metrics above can be used to integrate data quality targets into the system development life cycle, inbuilt as mandatory requirements for each phase of the development. Data quality analysts need to identify the data requirements for each application. A thorough traversal of the data flow within each application gives insight into probable insertion points for data inspection and control routines. These requirements must be added to the system’s functional requirements, for seamless incorporation into the development cycle, thus validating data at the point of introduction to the system.
* **Data Quality Improvement in operation systems** – Data that is shared between the data providers and consumers must be under contractual agreements that clearly define the acceptable levels of quality. Defining data standards as well as commonly agreed data. Formats helps in smooth flow of data from one business to another. The meta-data can be placed under a repository subjected to an active data center that ensure that data is represented in a fashion that is agreeable and beneficial to both the collaborating sides. The gap analysis and alignment of business needs of both the parties is done by this data control center.

Data quality inspections can be done manually or through automated routines to ascertain the working levels. Workflows can defined for periodically monitoring the data and taking remedial actions accordingly.

* **Inspect cases where Data quality standards are not met and taking remedial actions** – When data is found to be below expected levels, the remedial actions should be subjected to effective data tracking mechanisms much like the defect tracking systems in software development. Reporting data defects and tracking the actions can help to feed performance reports. A root-cause analysis done on each reported data error gives direct feedback for understanding the flaws in the business process processes

# Summary and Further Reading

The environment in which an organization operates will change. New opportunities will present themselves. New problems will arise. From these types of events, new data needs will arise. New and improved techniques and approaches will be explored and developed. A viable and vibrant data quality program should be capable of adapting to and accommodating the changing environment. A viable data quality program will require constant vigilance.

It is fitting to briefly addressing the problems of data quality that confront organization in the future and introducing research ideas and innovative applications that will be part of the continuing journey to data quality. Many of these initiatives build on the ideas and techniques discussed in the previous sections of the ADMG.

This ADMG discusses DQM governance that outlines the key roles and responsibilities at the senior and functional levels within the Army. DQM is a continuous process; therefore, the Army needs to stay focused in the upcoming years. Cost savings is one of the factors for implementing a DQM solution. However, data quality is critical to ensure the Army organizations can trust the data they need to perform their mission goals. For the way ahead, the Army should execute several tasks including:

* The ADB should create training materials, DQM portal on AKO/DKO or other site to promote DQM principals along with providing dashboard and set goals to promote DQM across the Army;
* The ADB should continue to support the identification and registration of Army ADSs;
* The Army should work with the DoD to update the DQM Guidelines and release it as an official DOD Directive;
* The Army should work with organizations with existing DQM programs to collect lessons learned and share that information;
* The Army should establish a DQM maturity model to measure how well organizations are meeting DoD DQM goals over time.

1. References

|  |  |
| --- | --- |
| [1] | Chief Data Office, Department of the Army, "Army Data Management Program (ADMP)," HQDA CIO/G-6, Department of the Army, Washington, D.C., To Be Published. |
| [2] | Office of the Army Chief Information Officer/G-6, "Army Data Plan," Army CIO/G-6, Ft. Belvoir, 2019. |
| [3] | Office of the Army Chief Information Officer, Department of Defense, "DoD Data Strategy," Department of Defense, Washington, D.C., 2020. |
| [4] | Office of the Army Chief Information Officer/G-6, "Army Data Strategy," HQDA CIO/G-6, Army Architecture Integration Center, Information Architecture Division, Ft. Belvoir, 2016. |
| [5] | Office of the Director of National Intelligence, "The Intelligence Community Data Management Lexicon," Office of the Director of National Intelligence, January 2020. |
| [6] | Office of the Chief Information Officer, United States Department of the Interior, "Data Quality Management Guide," Department of the Interior, Washington, D.C., 2008-08. |
| [7] | L. Sebastian-Coleman, Navigating the Labyrinth: An Executive Guide to Data Management, Basking Ridge, New Jersey: Technics Publications, 2018. |
| [8] | D. Kiran, Total Quality Management: Key Concepts and Case Studies, Butterworth-Heinemann, 2016. |
| [9] | K. Kerr, "The Development of a Data Quality Framework and Strategy for the New Zealand Ministry of Health," 2000-01. |
| [10] | J. Parkinson, The Data Quality Blueprint: A Comprehensive Step by Step Guide to an Effective & Long Lasting Enterprise-Wide Data Quality Solution, Holifast Limited, 2016, p. 660. |
| [11] | Headquarters, Department of the Army, "AR 25-1 Information Management - Army Information Technology," Department of the Army, Washington, D.C., 2019-07-15. |
| [12] | Office of the Army Chief Information Officer, "Army Information Architecture, Version 4.1," US Army CIO/G-6, Ft. Belvoir, 2013-06-05. |
| [13] | DAMA International, DAMA-DMBOK : Data Management Body of Knowledge, 2nd Edition ed., D. Henderson, S. Earley and L. Sebastian-Coleman, Eds., Basking Ridge, NJ: Technics Publications, 2017, p. 626. |
| [14] | Headquarters, Department of the Army, "AR 25-2 Information Assurance," Department of the Army, 2012-02. |
| [15] | "DoD IT Standards Registry (DISR)," [Online]. Available: https://gtg.csd.disa.mil/disr/standards/search/simple.html. [Accessed 09 04 2020]. |
| [16] | "ISIXSIGMA," [Online]. Available: https://www.isixsigma.com/. [Accessed 12 05 2020]. |
| [17] | Chief Information Officer, U.S. Department of Defense, "Net-Centric Data Strategy," Department of Defense, Washington, D.C., 2003-05-09. |
| [18] | Informatica, "PowerCenter," [Online]. Available: https://www.informatica.com/products/data-integration/powercenter.html. [Accessed 14 05 2020]. |
| [19] | The Data Leaders, "The Leader's Data Manifesto," [Online]. Available: https://dataleaders.org/. [Accessed 09 04 2020]. |
| [20] | IBM, "Unstructured Information Management Architecture SDK," [Online]. Available: https://www.ibm.com/developerworks/data/downloads/uima/. [Accessed 15 05 2020]. |
| [21] | S. Sarsfield, The Data Governance Imperative, IT Governance Publishing, 2009, p. 164. |
| [22] | Chief Information Officer, Department of Defense, "Department of Defense Information Enterprise Architecture, Version 1.2," Department of Defense, Washington, D.C., 2010-05-07. |
| [23] | Chief Information Officer, Department of Defense, "DoD Information Enterprise Architecture," [Online]. Available: https://dodcio.defense.gov/IntheNews/DoDInformationEnterpriseArchitecture.aspx. [Accessed 12 05 2020]. |
| [24] | Chief Information Officer, U.S. Department of Defense, "DoDAF Architecture Framework, Version 2.02," [Online]. Available: https://dodcio.defense.gov/Library/DoD-Architecture-Framework/. [Accessed 09 04 2020]. |
| [25] | A. D. Giordano, Performing Information Governance: A Step-by-Step Guide to Making Information Governance Work, 1st ed., Upper Saddle River, NJ: IBM Press, 2015, p. 648. |
| [26] | G. Gulipalli, "5 Best Practices of Effective Data Quality Management," 25 09 2015. [Online]. Available: https://www.invensis.net/blog/data-processing/5-best-practices-effective-data-quality-management/. [Accessed 12 06 2020]. |

# Acronyms and Definitions

## Acronyms and Abbreviations

|  |  |
| --- | --- |
| ADB | Army Data Board |
| ADC | Army Data Council |
| ADS | Authoritative Data Source |
| AIA | Army Information Architecture |
| API | Application Program Interface |
| AR | Army Regulation |
| CDO | Chief Data Officer |
| DAMA | Data Management Association |
| DMBOK | Data Management Body of Knowledge |
| DoD | Department of Defense |
| DoDAF | DoD Architecture Framework |
| DQM | Data Quality Management |
| FDM | Functional Data Manager |
| HQDA | Headquarters, Department of the Army |
| IA | Information Assurance |
| IAW | In Accordance With |
| IC | Intelligence Community |
| IES | Information Exchange Specification |
| IESS | Information Exchange Specification Standards |
| ISM | Information Security Markings |
| ADQM | Army Data Quality Management |
| ADS | Authoritative Data Source |
| DODAF | Department of Defense Architecture Framework |
| DoDI | Department of Defense Instruction |
| ISO | International Standards Organization |
| IT | Information Technology |
| OBT | Office of Business Transformation |
| XML | Extensible Markup Language |
| XSD | XML Schema Description language |

## Definitions

**Business Rule**

A recommendation, requirement, directive, stipulation, or imperative that asserts what should/shall be done to meet or implement a principle. A business rule may be an end-state objective.

**Chief Data Officer (CDO)**

A designated Senior Official within each IC Element responsible for the management of data as an asset and the establishment and enforcement of data-related strategies, policies, standards, processes and governance

**Data Asset**

“Any entity that is comprised of data. For example, a database is a data asset that is comprised of data records. A data asset may be a system or application output file, database, document, or web page. A data asset also includes a service that may be provided to access data from an application. For example, a service that returns individual records from a database would be a data asset. Similarly, a web site that returns data in response to specific queries (e.g., www.weather.com) would be a data asset. A human, system, or application may create a data asset.” (From DoDD 8320.02.)

**Data Exchange**

The physical and mechanical transference of data (via copying or transmission) from one location to another without consideration of meaning or intent of the data.

**Data Integration**

The process of combining data from two or more sources and producing a single unified, consistent, and cohesive view of the combined data. Generally, the objective is to produce a set of data that represents the same information that is represented by the input data sets, though this need not always be the case.

**Data Governance**

Discipline comprised of responsibilities, roles, functions, and practices, supported by authorities, policies, and decisional processes (planning, setting policies, monitoring, conformance, and enforcement), which together administer data and information assets across an IC Element to ensure that data is managed as a critical asset consistent with the organization’s mission and business performance objectives

**Data Lifecycle**

A conceptualization of a birth-to-death value chain for data, which often includes phases such as plan and task, acquire and assess, process and transform, discover and access, analyze and exploit, and preserve or dispose.

**Information Exchange Specification (IES)**

A set of specifications that govern the physical data format for data exchanged among members of a community. The specifications must include schemas, definitions of schema elements & relationships, and specifications of constraints. The specifications may include other relevant or supporting content that augments the primary specification content.

**Principle**

A generalized statement of position that is accepted as true or valid, and often reflects values, beliefs, or convictions on the “right” or “best” way to do or achieve a result or fulfill a mission. Principles guide decision-making and actions; a principle is not an end-state objective.

**Provenance**

Description of the origin or source of data, its history of stewardship or custodianship and location(s), which can be used to form assessments about its quality, reliability, or trustworthiness. Within a specific mission context only selective provenance attributes may be considered as relevant.