3

DATA QUALITY MATURITY

CHAPTER OUTLINE

What does organizational maturity mean in the context of good data management practices? From a holistic standpoint, the differences in organizational maturity for data quality are gauged by the sophistication of the processes in place for managing the identification of flawed data as well as the levels of capability of those tasked with managing data quality. Most organizations are reactive when it comes to resolving issues, meaning that problems are addressed at the time that the impacts have manifested themselves, but long after the failure has occurred. But as the practitioners in the organization gain a more thorough understanding of the methods for identifying the sources for data flaws, they become more proactive in identifying and resolving potential issues before negative business impacts occur.

In chapter 2, we provided an overview of the processes, people, and technology that are part of a data quality program. This chapter explores the concept of a capability/maturity model for data quality management, the life cycle of the data quality program, and how the organization transitions from one that is reactive into one that is proactive in ensuring high quality data.

3.1 The Data Quality Strategy

A strategy encompasses a long-term plan of action designed to achieve a specific objective. This plan provides a way to guide the efforts to ensure that they are contributing to the achievement of stated goals. We can therefore propose that a data quality strategy directs the organization to take the steps that will reduce the business impacts of poor data quality to an acceptable level.

Chapter 1 outlined some of the challenges and benefits of data quality management, whereas chapter 2 provided a more comprehensive introduction to the data quality program. And though there are clear benefits to an organizational data quality management program, the need to coordinate the efforts of different personalities in the organization means that there are bound to be conflicts that will arise among participants as the data quality program evolves. Even though most business applications and business operations depend on data quality, it cannot necessarily be mandated across administrative boundaries. Because of this, the expected benefits of improved information value can only be achieved when all participants willingly contribute to successful data quality management.

Frequently, defined data quality activities largely focus on evaluation and procurement of data quality tools, but won't encompass the management, technical, and operational infrastructure that must be in place to support a generalized conformance to acceptability levels of properly defined and documented expectations. Yet measuring this conformance demonstrates that the effort is succeeding. This suggests that when developing the data quality strategy, consider describing an operational framework for instituting best practices in the context of a level of maturity, and lay out the roadmap to address the challenges and achieve the benefits. Apply industry best practices and combine those with quality disciplines from other industrial domains (e.g., manufacturing, software development, or service industries). Ultimately, the data quality practices and processes should be relevant within your organization, and the approach to building the program should follow the patterns for other successful organizational programs.

It is a formidable challenge to establish the appropriate level of data quality to meet the needs of the diversity of participants, regulatory bodies, policy makers, and information clients when coupled with the different technologies and practices already in place. To address these, a data quality strategy requires governance, policies, practices, technology, and operational solutions that are all-encompassing yet present themselves to all participants as pragmatic and practical. Some things to keep in mind:

- **The Information Lifecycle:** When assembling a data quality strategy, it is necessary to identify the key success objectives for the program, evaluate the variables by which success is measured, establish information quality expectations, develop the governance model for overseeing success, and develop protocols for ensuring that policies and procedures for maintaining high quality data are followed by the participants across the enterprise. Information follows a "life cycle" (e.g., create, distribute, access, update, retire), so it is necessary that the data quality framework provide protocols for measuring the quality of information at the various stages of that life cycle.

- **Performance and maturity:** A data quality framework defines management objectives that are consistent with the key success objectives and the enterprise expectations for quality information, either through integration of services across an enterprise information architecture, or through the collaborative implementation of data governance policies and procedures. Performance associated with data quality expectations can be tied to a data quality maturity model. This maturity model establishes levels of performance and specifies the fundamental best practices needed to achieve each level of performance.
- **Data governance roles and responsibilities:** Also included in your data quality framework should be a model for data governance that outlines various data quality roles for the participants in the enterprise community. This model will provide an organizational structure and the policies and procedures to be followed by the community to ensure high quality data. The governance model defines data ownership and stewardship and describes accountability for the remediation of data quality issues across the various enterprise information systems. If necessary, the model will also define procedures for the data quality certification of participants as well as ongoing auditing of data quality.
- **Meeting expectations:** To achieve assurance of high quality data, the framework should provide for the identification, documentation, and validation of data quality expectations. These expectations can be transformed into data quality rules and metrics used to assess the business impact of poor data quality, develop performance models to gauge severity of data quality issues, track data quality events and issues, and provide ongoing data quality measurement, monitoring, and reporting of conformance with customer expectations.
- **Staff training and education:** To encourage coordination with the efforts to ensure data quality, there is value in educating participants in ways to integrate data quality as an important component of the system development life

cycle. The development of a component model for data quality services will expose the appropriate topics for training materials to facilitate data quality integration.

In addition, these concepts from chapter 2 should also be addressed in the data quality strategy:

- Provide a framework of data quality concepts
- Specify a data governance model to manage the oversight of data quality, incorporating data ownership, stewardship, and accountability for community-wide data quality
- Formalize approaches for identifying, documenting, and validating data quality expectations
- Provide practices to evaluate the business impacts of poor data quality and to develop performance models for issue management and prioritization
- Integrate methods and processes for data quality event tracking, data quality monitoring and measurement, and reporting of conformance with customer expectations
- Formulate a component service model for data quality services that is integrated with the enterprise/community interoperability model

3.2 A Data Quality Framework

Ultimately the practitioner must align the framework for data quality to meet the needs of the organization without overwhelming the individuals who will participate in the program. Casting the observance of data quality expectations within the context of key business performance metrics while minimizing intrusion and extra effort enables the program to gain traction and increase participation. The framework looks at varying degrees of maturity with respect to concepts introduced in the previous chapter, including:

- Defining data quality expectations
- Creating measurement using data quality dimensions
- Defining policies for measured observance of expectations
- Implementing the procedures supporting those policies
- Instituting data governance
- Agreeing to standards
- Acquiring the right technology
- Monitoring performance

3.2.1 Data Quality Expectations

Although the expectations associated with data quality measurements are often explicit, at times many of these expectations are implied or embedded within directives that drive other areas of importance. The data quality framework must address measuring conformance to expectations of data quality as they relate to particular participant needs. The framework must also specify:

- The relevant measures of data quality attributable to all data elements ("dimensions"),
- Metrics for evaluating conformance within each dimension, and
- Processes and services for evaluating conformance within each dimension.

3.2.2 Dimensions of Data Quality

This theme will continue to ring true throughout this book: it is said that one cannot improve something that cannot be measured. In the data quality program, the concept of "dimensions" classifies aspects of data quality expectations and provides measures to evaluate conformance to these measures. These metrics are used to quantify the levels of data quality and will be used to identify the gaps and opportunities for data quality improvement across an information flow. A thorough discussion of data quality dimensions will be presented in chapter 8.

3.2.3 Policies

The complexity of managing the different types of information policies that will be in place at your organization often leads to a limited capability for ensuring policy conformance. Whether the policies are defined internally (security, access), reflected across the customer space (e.g., privacy, sales, and support policies), or are externally imposed (e.g., legislative or regulatory industry standards), the challenge of policy management within the context of an information architecture should not be ignored. Policy management incorporates data quality dependencies among the areas of:

- Data certification (such as certification of trusted data sources or establishing trust with external data consumers),
- Privacy (including maintaining consistency with supporting the privacy framework based on limitations of use, storage, and duration stored),
- Lineage (such as tracking the origin and transference of data),
- Limitation of use (thereby overseeing the limits of the use of your organization's data outside of the enterprise), and
- Single source of truth (such as providing inquiry access through a single reference data index).

3.2.4 Procedures

The data quality procedures describe the operational aspects of a system to validate the existence and effectiveness of key data management activities. In addition, those procedures incorporate inspection to either automatically or (if necessary) manually validate data quality, and include preventative measures for proactive data quality assurance as well as control processes for identifying and providing guidance in eliminating the source of errors. These processes augment the information management activity to focus on:

- Data quality management,
- Standardized data inspection templates,
- Operational data quality,
- Issues tracking and remediation,
- Manual intervention when necessary,
- Integrity of data exchange,
- Contingency planning, and
- Validation.

As part of a set of protocols, service-level agreements may be specified for these activities and be integrated with the set of information policies, and this will be covered in greater detail in chapter 13.

3.2.5 Governance

The definitions and the management of data quality must incorporate the participation, collaboration, and oversight management from all enterprise participants, and to this end, the data quality framework specifies a data governance structure for management and oversight, and a set of data stewardship processes for all participants. Governance is required at various touch points to ensure consistency and conformance to the framework. Chapter 7 provides a model for developing a data quality governance model, with the outline for a data quality charter along with an organization structure, roles, responsibilities, and workflow for activities by the various participants.

3.2.6 Standards

Many industries participate within wider communities of data and information sharing. A data standards program facilitates the definition of and conformance with externally and internally defined standards for information exchange. Common business terms and their definitions are reflected at the information level, and data standards management looks at the data quality issues associated with:

- Data definitions,
- Semantics, and
- Data exchange.

These standards are managed as enterprise metadata, and are discussed in chapter 10.

3.2.7 Technology

To deploy the data quality framework, the participants within the enterprise will be expected to employ tools and technology intended to support the stated data quality protocols and processes, support the level of data quality services (e.g., validation, parsing and standardization, search/locate) via a single reference data set, and validate/verify the conformance of data values and records to explicitly defined data quality expectations. This will incorporate:

- Service component design guidance,
- Data quality technology,
- Business rule–based validation,
- Verification of data accuracy,
- Searching and indexing,
- Data quality issues tracking,
- Data quality performance management, and
- Identity resolution, record matching and linkage, and record splitting and merging when necessary.

3.2.8 Performance Management

Having specified processes for governance and stewardship and identified data quality expectations and ways to determine conformance of data to those expectations, it is necessary to provide a performance management scheme for continuously monitoring enterprise-wide data quality. This includes conforming to data quality expectations, identifying where significant negative impacts are incurred due to poor data quality, tracking relevant issues and providing a means for root cause analysis, and providing a set of services for the assessment of data quality performance. Together, this scheme will support auditing and monitoring that can be reported in a management dashboard characterizing data quality performance, highlighting areas that require special attention, and providing an audit trail for root cause analysis and remediation within the governance model.

3.3 A Data Quality Capability/Maturity Model

To take an approach of performance management for data quality, it is useful to visualize how data quality management dovetails with all organizational information-dependent activities. The challenge is that in many organizations, data management evolves in lockstep with the functional application needs, often as an afterthought. The data analysts starting to look at the organizational value of data as a corporate asset expose gaps showing the disconnect between the functional requirements for a collection of siloed business application and good data management practices. One way to evaluate and then resolve this disconnect is to assess the current level of maturity associated with data quality practices and then visualize a target level of maturity that best meets the organization's needs. In turn, this approach sets this vision as a yardstick by which one organization's maturity can be compared to the way that other organizations work.

This data quality maturity model is patterned after the Capability Maturity Model (CMM) developed by the Software Engineering Institute at Carnegie Mellon University. Capability maturity models are management tools that characterize levels of organizational refinement in addressing design, implementation, manufacturing, problem resolution, and so on. These kinds of models have been applied to many application domains, including software development, programmer development, and project management domains. This data quality maturity model defines five levels of maturity, ranging from an initial level where practices and policies are ad hoc, to the highest in which processes and practices lead to continuous measurement, improvement, and optimization.

3.3.1 Initial

At the initial level, the processes used for data quality assurance are largely ad hoc, with most of the effort expended in reacting to data quality issues. Problems that arise are acute, require immediate attention, and often require significant rollback and rework. The environment is relatively unstable, and as a result it becomes challenging to trace back the sources of the introduction of flawed data and to determine the quick fix. Fixes are onetime, and most likely will not address any long-term improvement in the information or the processes. Success is often correlated to individual heroics in correcting data flaws, not on proven processes for root cause analysis and managed remediation. At this level, there is little or no sharing of information or experiences, and therefore there is limited or no ability to repeat successes.

3.3.2 Repeatable

At the repeatable level, there is some basic organizational management and information sharing, augmented by some process discipline, mostly in recognizing good practices and attempting to replicate them in similar situations, enabling some ability to repeat success. There is an introductory level of governance with limited documentation of processes, plans, standards, and practices. When representatives from the lines of business application understand these good practices and attempt to put them in place, their ability to respond to data failures is more streamlined.

The rate of adoption of data management practices varies across different lines of business. Some considerations of the impacts of poor data quality lead to introductory efforts for business evaluation and identification of gross-level measures. These measures may be for proactive data quality management. Some technology components are in place, but they may not be standardized and their behaviors are not synchronized. A focus on the need for technology eclipses the identification of business needs for tools and the definition of methods for using any acquired tools.

3.3.3 Defined

At the defined level, a structured team of data quality practitioners begin to document good practices, which are:

- An established set of data governance policies,
- Processes for defining data quality expectations,
- Technology components, and
- Processes and services for implementing data quality validation, assurance, and reporting.

Once these are documented and can be made available across the organization, there emerges a degree of consistent use. An enterprise-wide data quality team has scheduled meetings to discuss organizational issues, review methods and technology, and to exchange ideas.

A framework for determining responsibility and accountability for the quality of the organization's information is in place, and accountability is monitored by an organizational governance board with representatives from the business and IT divisions. There are tailored guidelines for establishing standards and management objectives, and there are processes in place to ensure that these objectives are met. Expectations based on defined data quality dimensions can be expressed, documented, and integrated within a (conceptual) service model. The use of technical components is standardized at both service and implementation layers.

3.3.4 Managed

At the managed level, the data quality program fully incorporates business impact analysis with the ability to express data quality expectations and measure conformance to those expectations. These measurements form the basis of clearly defined criteria for performance in relation to meeting business objectives. Metrics composed of these weighted measurements are used in evaluating statistical process control at different service levels. Measured performance characteristics can be used to assess overall system performance against success criteria. Data quality is proactive, with data flaws identified early in the information workflow. Remediation is governed by well-documented procedures, information is shared, and overall performance against quality expectations is predictable.

3.3.5 Optimized

At the optimized level, the data quality maturity governance framework is in place such that enterprise-wide performance measurements can be used for identifying opportunities for improved systemic data quality. The ability to assess success and identify causes for process variation may suggest actions to adapt standards, policies, and processes for incremental or fundamental quality improvements. Strategic improvements and continuous process monitoring of the data life cycle using dashboards are applied throughout the organization.

3.4 Mapping Framework Components to the Maturity Model

3.4.1 Data Quality Expectations

The challenge of defining data quality expectations often is tied to the convergence of understanding between the information technologists and their business clients. Although the

Table 3.1 Component Maturity Description for Data Quality Expectations

Level	Characterization
Initial	• Data quality activity is reactive • No capability for identifying data quality expectations • No data quality expectations have been documented
Repeatable	• Limited anticipation of certain data issues • Expectations associated with intrinsic dimensions of data quality (see chapter 8) associated with data values can be articulated • Simple errors are identified and reported
Defined	• Dimensions of data quality are identified and documented • Expectations associated with dimensions of data quality associated with data values, formats, and semantics can be articulated using data quality rules • Capability for validation of data using defined data quality rules • Methods for assessing business impact explored
Managed	• Data validity is inspected and monitored in process • Business impact analysis of data flaws is common • Results of impact analysis factored into prioritization of managing expectation conformance • Data quality assessments of data sets performed on cyclic schedule
Optimized	• Data quality benchmarks defined • Observance of data quality expectations tied to individual performance targets • Industry proficiency levels are used for anticipating and setting improvement goals • Controls for data validation integrated into business processes

responsibility for addressing data quality issues lies solely with IT, it is difficult to establish protocols for long-term improvement; but as the partnership between IT and the business side grows stronger, the ability to effectively define and measure against data quality expectations grows as well. The mapping of the maturity model to data quality expectations is shown in Table 3.1.

3.4.2 Dimensions of Data Quality

The ability to predict where data quality becomes critical to achieving business objectives enables data quality management to become systemic. This depends on translating the occurrence of specific business impacts into a taxonomy that systematically allows the practitioner to measure, assess, and

Table 3.2 Component Maturity Description for Defining and Using Data Quality Dimensions

Level	Characterization
Initial	• No recognition of ability to measure data quality • Data quality issues not connected in any way • Data quality issues are not characterized within any kind of management taxonomy
Repeatable	• Recognition of common dimensions for measuring quality of data values • Capability to measure conformance with data quality rules associated with data values
Defined	• Expectations associated with dimensions of data quality associated with data values, formats, and semantics can be articulated • Capability for validation of data values, models, and exchanges using defined data quality rules • Basic reporting for simple data quality measurements
Managed	• Dimensions of data quality mapped to a business impact taxonomy • Composite metric scores reported • Data stewards notified of emerging data flaws
Optimized	• Data quality service level agreements defined • Data quality service level agreements observed • Newly researched dimensions enable the integration of proactive methods for ensuring the quality of data as part of the system development life cycle

improve information value. Table 3.2 describes the mapping of the maturity model for defining and using data quality dimensions.

3.4.3 Policies

As the organization matures, the approach to managing conformance to information policies will transition from an informal approach that has limited documentation to one that completely integrates business activities, information policies, and auditable conformance, as is shown in Table 3.3.

3.4.4 Procedures

A measure of a high-performance organization lies in its well-defined processes and protocols for ensuring information quality, and this is described in Table 3.4.

Table 3.3 Component Maturity Description for Information Policies

Level	Characterization
Initial	• Policies are informal • Policies are undocumented • Repetitive actions taken by many staff members with no coordination
Repeatable	• Organization attempts to consolidate "single source of truth" data sets • Privacy and limitations of use policies are hard-coded • Initial policies defined for reacting to data issues
Defined	• Tailored guidelines for establishing management objectives are established at line of business • Certification process for qualifying data sources is in place • Best practices captured by data quality practitioners • Data quality service level agreements defined for managing observance of policies
Managed	• Policies established and coordinated across the enterprise • Provenance management details the history of data exchanges • Policy-based data quality management • Performance management driven by data quality policies • Data quality service level agreements used for managing observance of policies
Optimized	• Automated notification of noncompliance to data quality policies • Self-governing system in place

3.4.5 Governance

Table 3.5 shows how to map a data governance program to the different levels of the maturity model. The emergence of organizational data governance evolves both from the bottom up, as opportunities for information sharing are provided, and from the top down as the responsibilities are formalized.

3.4.6 Standards

Interoperability is a key to coordinated information activities, and the maturity of the organization is reflected in the way it defines and implements data standards, as is described in Table 3.6.

Table 3.4 Component Maturity Description for Data Quality Protocols

Level	Characterization
Initial	• Discovered failures are reacted to in an acute manner • Data values are corrected with no coordination with business processes • Root causes are not identified • Same errors corrected multiple times
Repeatable	• Ability to track down errors due to incompleteness • Ability to track down error due to invalid syntax/structure • Root cause analysis enabled using simple data quality rules and data validation
Defined	• Procedures defined and documented for data inspection for determination of accuracy and validity • Data quality management is deployed at line-of-business level as well as at enterprise level • Data validation is performed automatically and only flaws are manually inspected • Data contingency procedures in place
Managed	• Data quality rules are proactively monitored • Data controls are designed for incorporation into distinct business applications • Data flaws are recognized early in information flow • Remediation is governed by well-defined processes • Validation of exchanged data in place • Validity of data is auditable
Optimized	• Data controls deployed across the enterprise • Participants publish data quality measurements • Data quality management practices are transparent

3.4.7 Technology

As the maturity of the organization grows, the focus on data quality improvement transitions from acquiring tools to assembling a service-oriented approach for the entire enterprise, as is described in Table 3.7.

3.4.8 Performance Management

Creating a performance-based organization requires a focus on defining performance objectives and using the tools, methods, and protocols to measure conformance to those objectives

Table 3.5 Component Maturity Description for Data Governance

Level	Characterization
Initial	• Little or no communication regarding data quality management • Information technology is default for all enterprise data quality issues • No data stewardship • Responsibility for data corrections assigned in an ad hoc manner
Repeatable	• Best practices are collected and shared among participants. • Key individuals from community form workgroup to devise and recommend data governance program and policies • Guiding principles and data quality charter are in development
Defined	• Organizational structure for data governance oversight defined • Guiding principles, charter, and data governance management policies are documented • Standardized view of data stewardship across the enterprise and stewardship program is in place • Operational data governance procedures defined
Managed	• Data governance board consisting of representatives from across the enterprise is in place • Collaborative data quality governance board meets regularly • Operational data governance driven by data quality service level agreements • Teams within each division or group employ similar governance framework internally • Reporting and remediation frameworks collaborate in applying statistical process control to maintain control within defined bounds
Optimized	• Data quality performance metrics for processes are reviewed for opportunities for improvement • Staff members rewarded for meeting data governance performance goals

and ultimately meet those performance goals. The mapping of the maturity model to performance management is shown in Table 3.8.

3.5 Summary

One of the more important objectives of the process of defining a data quality strategy and framework is to better understand how to integrate performance-based data quality activities into the entire system. Measurements and metrics are designed

Table 3.6 Component Maturity Description for Data Standards

Level	Characterization
Initial	• No data standards defined • Similar data values represented in variant structures • No data definitions
Repeatable	• Data element definitions for commonly used business terms • Reference data sets identified • Data elements used as identifying information specified • Certification process for trusted data sources being defined • Data standards metadata managed within participant enterprises • Definition of guidelines for standardized exchange formats (e.g., XML)
Defined	• Enterprise data standards and metadata management • Structure and format standards defined for all data elements • Exchange schemas are defined
Managed	• Certification of trusted data sources in place • Master reference data sets identified • Exchange standards managed through data standards oversight process • Data standards oversight board oversees ongoing maintenance of internal standards and conformance to externally defined standards
Optimized	• Master data concepts managed within a master data environment • Taxonomies for data standards are defined and endorsed • Conformance with defined standards is integrated via a policy-oriented technical structure • Straight-through processing is enabled for standard data

around the understanding of how poor data quality impacts the business. An organization's level of maturity of data quality management can be assessed, and performance objectives can be defined. Architecting a framework that is essentially driven by governance and performance, intended to achieve a targeted level of maturity, will enable the description of a program with well-defined milestones and deliverables.

Table 3.7 Component Maturity Description for Data Quality Technology

Level	Characterization
Initial	• Internally developed ad hoc routines employed • "Not invented here" mentality
Repeatable	• Tools for assessing objective data quality are available • Data parsing, standardization, and cleansing are available • Data quality technology used for locate, match, and linkage
Defined	• Standardized procedures for using data quality tools for data quality assessment and improvement in place • Business rule–based techniques are employed for validation • Technology components for implementing data validation, certification, assurance, and reporting are in place • Technology components are standardized across the federated community at the service and at the implementation layers
Managed	• Automatic data correction guided by governance policies and defined business rules • Impact analysis and what-if scenarios supported by dashboard and reporting tools
Optimized	• Nontechnical users can define and modify data quality rules and dimensions dynamically

Table 3.8 Component Maturity Description for Performance Management

Level	Characterization
Initial	• Impacts are manifested and recognized long after failure events take place
Repeatable	• Characterization of areas of impact of poor data quality • Data profiling used to identify data failures in process
Defined	• Impact analysis framework in place • Data quality service components identify flaws early in process • Data quality service components defined • Issues tracking system in place to capture issues and their resolutions
Managed	• Data quality metrics fed into performance management reporting • Auditing based on conformance to rules associated with data quality dimensions • Consistent reporting of data quality management for necessary participants • Performance dashboards are in place • Role-based access to performance information • Well-defined visualization of data quality component contribution to business impacts
Optimized	• Enterprise-wide performance can be improved through policy modification via rules environment