Work On Data Quality Practices To Cover The Data Governance Approach

Posted on March 26, 2019

Data quality management should begin with data governance because an organization cannot improve data consistently without deciding who owns decisions, which standards apply, how quality is measured, and what happens when data fail to meet requirements. The original article correctly argued that a successful initiative needs executive recognition, appropriate techniques, a business glossary, stewardship, profiling, validation, and remediation. Those practices remain central, but data governance is broader than purchasing tools or hiring consultants. It is an operating model through which people exercise decision rights over data. Good governance treats data as an organizational asset while recognizing that some data also represent customers, employees, patients, citizens, or other individuals whose rights must be protected. The goal is not perfect data in every field. It is data that are sufficiently accurate, complete, consistent, timely, valid, and traceable for the decisions and obligations they support.

Recognize Data as a Business Asset and Responsibility

The first step is executive recognition that poor data quality creates measurable risk. Incorrect customer addresses increase returned mail and failed deliveries. Duplicate supplier records distort spending analysis. Inconsistent product codes create inventory errors. Missing clinical information can threaten patient safety. Unreliable regulatory data can produce penalties and loss of trust. Executives should therefore treat data in a manner comparable to financial, human, and physical assets: assign accountability, fund maintenance, define controls, and measure performance. However, data differ from ordinary assets because copying is easy and use can affect people. Governance must balance value creation with privacy, security, ethics, records management, and legal requirements. A statement that “data are valuable” is not enough. Leaders need to identify which decisions depend on which data and sponsor improvements tied to business outcomes.

Define the Scope and Business Case

Organizations often attempt to govern all data at once and create a program too broad to deliver results. A better approach begins with a defined domain or use case, such as customer identity, financial reporting, supplier master data, product information, employee records, or regulatory submissions. The business case should explain the current problem, affected processes, expected benefits, risk reduction, and measures of success. For example, a customer-data initiative may target

duplicate records, incomplete consent status, and inconsistent identifiers across sales and support systems. The scope should include upstream sources, transformations, downstream reports, and responsible teams. Beginning with a priority domain allows the organization to demonstrate value, learn how decisions are made, and extend the model gradually.

Establish Governance Principles

Governance principles guide decisions when detailed rules do not yet exist. Common principles include data should have an accountable owner; definitions should be shared; quality requirements should reflect intended use; access should follow least privilege; personal data should be minimized; lineage should be documented; and defects should be corrected as close to the source as practical. Principles must be approved and communicated, but they should also be translated into procedures. If “data should be accurate” remains an aspiration without thresholds, monitoring, and escalation, it does not govern behavior. Principles should clarify trade-offs, such as when timeliness matters more than completeness or when a local definition may legitimately differ from an enterprise standard.

Create a Governance Structure

A governance structure assigns decision rights. An executive council may approve policy, priorities, funding, and cross-functional standards. Data owners are typically senior business leaders accountable for a domain. Data stewards translate policy into definitions, rules, issue resolution, and day-to-day coordination. Data custodians or technology teams implement controls in systems and platforms. Privacy, security, legal, compliance, analytics, and records-management specialists provide expertise. These roles should not exist only on an organizational chart. Each role needs authority, responsibilities, time, escalation paths, and performance expectations. Governance fails when stewards are held responsible for quality but cannot influence the processes creating defects.

Adopt Existing Quality and Stewardship Practices

The original article recommended using techniques already developed in data-quality and stewardship programs rather than reinventing them. This is practical because governance often grows from existing controls. Organizations may already have reconciliation procedures, exception reports, metadata repositories, master-data teams, privacy reviews, or regulatory reporting committees. Governance should connect these activities, remove duplication, and clarify enterprise decisions. Existing practices must still be assessed critically. A legacy process may detect errors without preventing them, or one department’s standard may conflict with another’s needs. The aim is to reuse proven capabilities while establishing a common operating model.

Develop a Business Glossary

A business glossary records agreed meanings for important terms. The original article correctly described it as an extension of metadata management, but a useful glossary is more than an inventory. Each term should include a clear definition, owner, steward, approved context, calculation where relevant, source, related terms, permitted values, and effective date. Consider “active customer.” Sales may define a customer as anyone with an open opportunity, finance may require a completed purchase, and support may include users with service accounts. Governance should not hide these differences. It should decide whether one enterprise definition is possible or document separate qualified terms. A glossary reduces ambiguity in reports, integration, artificial-intelligence models, and regulatory submissions.

Connect the Glossary to Technical Metadata

Business terms become operational when they are linked to databases, fields, reports, APIs, and transformations. Technical metadata describes physical structures, data types, mappings, jobs, and system relationships. Linking business and technical metadata helps users discover where a concept is stored and how it is calculated. It also supports impact analysis: if a field changes, teams can identify affected reports and controls. Automated metadata scanning can accelerate collection, but human interpretation remains necessary. A tool may detect a column named “status” without knowing whether it represents account status, payment status, or employment status. Governance combines automation with accountable meaning.

Use Data Stewardship to Align Governance With Business Goals

Data stewardship provides the working connection between standards and operations. Stewards understand how data are created, why users need them, and which defects matter. They coordinate definitions, approve quality rules, investigate issues, and bring unresolved conflicts to owners or councils. The original article suggested expanding pre-existing stewardship programs into governance, which can be effective. However, stewardship should not become unpaid administrative work assigned to employees without authority. Organizations should select people with domain knowledge, allocate time, train them, and recognize stewardship in performance goals. Stewards should also represent users and affected individuals, not merely enforce technical conformity. (DAMA International, 2017)

Map Critical Data Elements

Not all data require the same level of control. Critical data elements are fields or concepts whose failure could materially affect operations, reporting, safety, customer experience, legal compliance, or strategic decisions. Examples include legal entity identifiers, account balances, patient allergies, tax status, consent indicators, product safety classifications, and employee access roles. Identifying critical elements allows the organization to focus profiling, rules, lineage, and monitoring where harm is greatest. The classification should be reviewed as business processes change. A field that appears minor in one system may become critical when reused by an automated decision process.

Profile Data Early and Repeatedly

The original article emphasized profiling data “early and often.” Data profiling examines content and structure to reveal null rates, patterns, distributions, duplicates, invalid values, outliers, and relationships. Early profiling prevents teams from designing standards based on assumptions. For example, a field described as mandatory may be empty in thirty percent of records, or a date field may contain impossible values because of migration. Profiling should occur before integration, migration, analytics, or model development and continue after deployment. A one-time assessment becomes obsolete as new sources, users, and defects appear. Profiling results should be interpreted with domain experts because an unusual value may be an error or a legitimate rare case.

Define Data Quality Dimensions

Quality is multidimensional. Accuracy asks whether data correctly represent reality. Completeness asks whether required values are present. Consistency asks whether values agree across systems and rules. Timeliness asks whether data are available and current when needed. Validity asks whether data conform to format, type, range, and business rules. Uniqueness addresses unintended duplicates. Integrity examines relationships among records and systems. A dataset can be valid but inaccurate—for example, a correctly formatted date that is the wrong date. Governance should select dimensions based on use rather than create one universal score that hides important weaknesses. (Wang & Strong, 1996)

Set Measurable Rules and Thresholds

Each critical element should have quality rules linked to business requirements. A rule might state that every active supplier must have a valid tax identifier, that a discharge record must contain a timestamp, or that a financial total must reconcile with the ledger. Thresholds should indicate acceptable performance and trigger action. Requiring one hundred percent quality may be justified for a life-critical field but wasteful for optional marketing attributes. Rules need owners, calculation

methods, frequency, and escalation paths. They should also be versioned because requirements change. Good governance makes quality expectations testable.

Document Data Lineage

Data lineage shows where data originate, how they move, which transformations are applied, and where they are consumed. It supports audit, impact analysis, root-cause investigation, and user trust. Without lineage, teams may correct a report while the source defect continues generating errors. Lineage can be technical, showing tables and jobs, and business-oriented, explaining how a reported measure is derived. Automated tools can capture many technical relationships, but manually maintained documentation may still be required for spreadsheets, external files, and informal processes. Governance should prioritize end-to-end lineage for critical data rather than attempt exhaustive mapping with no business use.

Validate Data at the Point of Entry

The original article recommended real-time verification and validation. Prevention is usually less expensive than downstream correction. Systems can enforce required fields, formats, reference values, range checks, and relationship rules when data are entered or received. User interfaces should make correct entry easy through clear labels, constrained choices, and helpful error messages. Validation must avoid rejecting legitimate exceptions. A rigid address format may exclude international users; a binary gender field may not meet the process need; an automated identity match may fail after a legal name change. Rules should be tested with real diversity and allow controlled exception handling.

Monitor Data in Motion and at Rest

Quality controls should operate during data ingestion, transformation, storage, and consumption. Batch and streaming pipelines can validate schema, volume, freshness, and business rules. Monitoring can detect sudden shifts that indicate source changes or broken processing. Dashboards should show trends and impact rather than only technical counts. A thousand invalid rows may be minor in a billion-record dataset or critical if they include every transaction from one region. Alerts should be prioritized to avoid fatigue. Monitoring also needs ownership: a warning without a responsible responder is not a control.

Remediate Non-Compliant Data

The original article correctly stated that metrics reveal non-compliant data that require remediation. Correction may involve standardizing formats, merging duplicates, completing missing attributes,

reconciling conflicting records, or restoring data from authoritative sources. Automated remediation is useful for deterministic rules, but high-impact or ambiguous cases require review. Changes should be traceable, especially when data affect rights, finance, healthcare, or regulation. The objective is not only to clean the current record. Teams should identify why the defect occurred and modify the source process, integration, training, or control. Otherwise, cleansing becomes a recurring cost.

Manage Data Quality Issues Through a Workflow

Issues should be logged with description, affected domain, severity, business impact, owner, root cause, planned action, due date, and resolution evidence. Severity should consider harm, not merely record count. A small number of incorrect access permissions may be more urgent than thousands of missing optional fields. The workflow should allow escalation when teams disagree or corrective work requires funding. Closing an issue should require verification that the fix worked. A transparent backlog helps leaders see where structural investment is needed.

Apply Root-Cause Analysis

Common root causes include unclear definitions, manual re-entry, weak validation, conflicting incentives, system limitations, poor integration, inadequate training, and changes introduced without impact assessment. Techniques such as the five whys, process mapping, and cause-and-effect analysis can reveal why errors recur. If sales employees omit fields because the interface is slow and the data do not help their work, repeating training may not solve the problem. Governance must examine process design and incentives. Sustainable quality is created by the system of work, not by blaming data-entry staff.

Measure Value and Performance

Governance metrics should include both data measures and business outcomes. Quality dashboards may track completeness, duplicates, rule failures, freshness, and issue resolution time. Business measures may track reduced returns, faster onboarding, fewer reporting adjustments, improved claims processing, lower regulatory risk, or increased confidence in analytics. Activity measures such as the number of glossary terms created can show effort but not value. Executives need evidence that governance improves decisions or reduces harm. Metrics should also reveal unintended consequences, such as validation rules that delay service or disproportionately reject certain users.

Privacy, Security, and Ethical Use

Data quality is not automatically ethical. Highly accurate personal data can still be collected without justification or used unfairly. Governance should integrate privacy, security, retention, consent where

applicable, and ethical review. Data minimization reduces the volume requiring quality management and limits exposure. Access should follow business need, and sensitive data should be protected through technical and organizational controls. When data feed automated decisions or artificial intelligence, governance should assess representativeness, bias, explainability, and the ability of affected people to correct records. A quality score cannot substitute for ethical judgment. (National Institute of Standards and Technology, 2020)

Technology's Supporting Role

Tools can support cataloging, profiling, observability, lineage, master data, rule execution, issue management, and workflow. They cannot decide contested definitions or create accountability by themselves. Organizations should select technology after clarifying the operating model and priority use cases. Otherwise, they may build an expensive catalog that users do not trust or an alert system no one owns. Integration with existing platforms and ordinary work processes is more important than feature quantity. Technology should reduce manual burden and make governance evidence visible.

A Practical Implementation Sequence

A practical program can begin by selecting one high-value domain, securing an executive sponsor, appointing an owner and stewards, documenting business outcomes, identifying critical elements, creating definitions, profiling sources, defining rules, mapping lineage, implementing preventive controls, and establishing an issue workflow. The team should measure results, publish lessons, and expand to related domains. Governance should be iterative because new regulations, systems, analytics, and business models change data requirements. A phased approach provides control without creating a bureaucracy detached from operations.

Conclusion

Data quality management begins with governance because quality depends on decisions about ownership, meaning, standards, controls, and accountability. The original sequence remains valuable: obtain executive support, reuse stewardship and quality practices, build a business glossary, align governance with business goals, profile data, validate it, and remediate non-compliance. A mature approach also identifies critical data, documents lineage, measures several quality dimensions, manages issues, performs root-cause analysis, integrates privacy and security, and demonstrates business value. Tools can automate parts of the process, but governance is ultimately a human system of decision rights and responsibilities. Reliable data result when organizations design quality into the processes that create and use it rather than repeatedly cleaning defects after harm occurs.

References

DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd ed.). Technics Publications.

ISO. (2015). *ISO 8000-61: Data quality—Part 61: Data quality management: Process reference model*.

National Institute of Standards and Technology. (2020). *NIST Privacy Framework: A tool for improving privacy through enterprise risk management*.

Redman, T. C. (2013). Data's credibility problem. *Harvard Business Review, 91*(12), 84–88.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5–33.