International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

DATA QUALITY AND INTEGRITY MANAGEMENT FRAMEWORK FOR ENTERPRISE ERP MIGRATION PROJECTS

Ella Foster
Independent Researcher

ABSTRACT

Enterprise Resource Planning migration projects are critical components of organizational digital transformation because they enable enterprises to replace fragmented legacy systems, modernize business processes, consolidate operational information, adopt cloud platforms, improve analytical capabilities, and establish integrated enterprise architectures. However, ERP migration initiatives frequently experience substantial difficulties due to incomplete records, duplicate entities, inconsistent master data, invalid formats, obsolete values, broken relationships, semantic mismatches, transformation errors, unauthorized modifications, missing lineage, reconciliation failures, and weak governance. Poor-quality or corrupted data can cause incorrect financial reporting, disrupted procurement, inaccurate inventory positions, payroll errors, customer-service failures, compliance violations, and reduced confidence in the migrated platform. This paper proposes a comprehensive Data Quality and Integrity Management Framework for Enterprise ERP Migration Projects. The framework integrates legacy data discovery, source-system profiling, business-critical data classification, metadata management, quality-rule definition, duplicate detection, master-data harmonization, schema mapping, semantic transformation, referential-integrity validation, migration-wave planning, secure extraction, staging-zone controls, transformation monitoring, reconciliation, lineage tracking, anomaly detection, role-based authorization, API governance, auditability, migration Digital Twins, rollback readiness, post-migration monitoring, and continuous quality improvement. The proposed methodology establishes a multi-stage migration lifecycle in which data is assessed before extraction, validated during transformation, reconciled after loading, and continuously monitored after production cutover. Quality dimensions including completeness, accuracy, consistency, validity, uniqueness, timeliness, conformity, and referential integrity are evaluated according to business-specific requirements. A migration Digital Twin maintains synchronized representations of source datasets, transformation states, target structures, quality exceptions, lineage relationships, and migration-wave progress. Machine-learning-assisted anomaly analytics identify unusual records, transformation deviations, duplicate patterns, and integrity risks that may not be captured through static rules. OpenAPI-oriented service contracts support interoperability among migration tools, ERP applications, validation services, quality dashboards, and governance platforms, while secure lifecycle management protects migration credentials and sensitive enterprise data. A representative analytical evaluation compares conventional extract-transform-load migration with the proposed framework across data completeness, duplicate reduction, transformation accuracy, reconciliation success, referential-integrity violations, migration defects, exception-resolution time, rollback readiness, cutover stability, and post-migration business incidents. The results indicate substantial improvements in migrated-data reliability, traceability, operational

Received: 16-02-2024 | Accepted: 16-03-2024 | Published: 23-03-2024

continuity, and governance. The study concludes that successful ERP modernization requires data quality and integrity to be managed as continuous lifecycle capabilities rather than as isolated cleansing activities performed immediately before cutover.

Keywords: ERP Migration, Data Quality, Data Integrity, Enterprise Modernization, Data Governance, Master Data Management, Data Reconciliation, Data Lineage, Digital Twin, ETL, OpenAPI, Migration Analytics, Legacy Systems.

I. INTRODUCTION

Enterprise Resource Planning (ERP) systems integrate finance, procurement, inventory, manufacturing, sales, supply chain, customer relationship management, human resources, and business intelligence into a unified enterprise platform. As organizations undergo digital transformation, legacy ERP systems are increasingly being replaced by cloud-enabled and service-oriented platforms that offer improved scalability, operational agility, and analytical capabilities. However, the success of ERP modernization depends not only on software deployment but also on the quality and integrity of enterprise data transferred from legacy environments. Inaccurate, incomplete, inconsistent, or duplicated data can significantly reduce the effectiveness of the new ERP system and negatively impact business operations **[1], [2]**.

Data quality has become one of the most important success factors in ERP migration projects because enterprise databases typically contain information accumulated over many years from heterogeneous applications, manual processes, mergers, acquisitions, and departmental systems. These environments frequently contain duplicate customer records, inconsistent supplier identifiers, missing financial information, obsolete product codes, invalid transaction histories, and conflicting business rules. Traditional Extract–Transform–Load (ETL) approaches mainly focus on transferring data between systems and often provide limited support for continuous quality monitoring, semantic validation, metadata governance, and integrity verification. Consequently, organizations require comprehensive frameworks that manage data quality throughout the entire migration lifecycle rather than only during pre-migration cleansing **[3], [4]**.

Data integrity is equally important because enterprise records are interconnected through complex business relationships. Financial transactions depend on valid account structures, purchase orders reference approved suppliers, inventory movements depend on product master data, payroll records require valid employee identifiers, and customer transactions rely on consistent organizational hierarchies. During migration, incorrect mappings, incomplete transformations, or broken dependencies may compromise referential integrity and produce significant operational failures. Modern migration frameworks therefore integrate metadata management, automated validation, reconciliation, lineage tracking, and governance mechanisms to preserve business relationships across heterogeneous enterprise environments **[5], [6]**.

Recent advances in machine learning, Digital Twins, cloud-native architectures, Application Programming Interfaces (APIs), and intelligent enterprise integration have significantly improved ERP migration capabilities. Machine learning assists in identifying duplicate records, detecting transformation anomalies, validating data quality, and predicting migration risks, while Digital Twins provide synchronized representations of migration status and enterprise information. Standardized APIs further enable secure interoperability among migration tools,

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

ERP applications, governance platforms, and quality-monitoring services, thereby improving migration visibility, automation, and operational reliability **[7], [8]**.

Although previous studies have independently investigated data quality, enterprise integration, metadata management, Digital Twins, business process management, and ERP modernization, relatively few provide a unified framework integrating multidimensional quality assessment, referential integrity validation, metadata-driven governance, Digital Twin monitoring, secure API communication, anomaly analytics, reconciliation, and continuous post-migration quality monitoring. Therefore, this research proposes a **Data Quality and Integrity Management Framework for Enterprise ERP Migration Projects** that integrates intelligent profiling, metadata management, quality-rule enforcement, automated validation, Digital Twin monitoring, secure enterprise integration, continuous reconciliation, and lifecycle governance to improve migration accuracy, operational continuity, enterprise security, and long-term data reliability **[9], [10]**.

II. LITERATURE SURVEY

Wand and Wang (1996) established ontological foundations for data quality by demonstrating that information quality depends on how accurately information systems represent real-world entities and relationships. Their work provides a theoretical basis for evaluating semantic correctness during ERP migration, where preserving business meaning is as important as transferring data structures **[11]**.

Redman (1998) investigated the organizational impact of poor data quality and demonstrated that inaccurate information significantly affects business performance, decision-making, and operational efficiency. The study emphasized that data quality should be treated as a strategic organizational asset rather than a technical concern, supporting continuous quality management throughout ERP modernization **[12]**.

Lee, Strong, Kahn, and Wang (2002) proposed the AIMQ methodology for multidimensional information quality assessment. Their framework demonstrated that completeness, consistency, accuracy, validity, and usability should be evaluated simultaneously instead of relying on a single quality metric. These concepts provide valuable guidance for multidimensional ERP migration quality assessment **[13]**.

Batini, Cappiello, Francalanci, and Maurino (2009) surveyed methodologies for data quality assessment and improvement across heterogeneous information environments. Their research highlighted the importance of systematic quality management, continuous monitoring, and structured improvement processes. These principles directly support the proposed framework's lifecycle-oriented approach to ERP migration quality management **[14]**.

Bernstein and Melnik (2007) introduced Model Management 2.0 and demonstrated that intelligent schema mapping and metadata management simplify integration among heterogeneous information systems. Their work provides important theoretical support for automated source-to-target mapping and semantic transformation during ERP migration **[15]**.

Kumar (2014) proposed Digital Twin-based smart manufacturing systems for real-time operational optimization. The concept of maintaining synchronized virtual representations is extended in this research through a migration Digital Twin that continuously represents source systems, target structures, transformation progress, quality states, and reconciliation results during ERP migration **[16]**.

Kumar (2018) investigated machine learning techniques for process optimization and defect

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

reduction in manufacturing systems. The research demonstrated that data-driven analytics effectively identify complex operational anomalies, providing useful concepts for detecting transformation errors, duplicate patterns, and quality exceptions during enterprise data migration **[17]**.

Hitzler, Krötzsch, and Rudolph (2009) presented semantic web technologies for representing relationships among heterogeneous information sources. Their work supports semantic metadata management and knowledge representation during ERP migration, helping preserve business meaning and interoperability across enterprise systems **[18]**.

Gummadi (2020) proposed standardized API design using RAML and OpenAPI specifications to improve interoperability among enterprise applications. These concepts support secure communication between migration services, validation engines, governance platforms, Digital Twins, and target ERP environments, improving migration automation and operational reliability **[19]**.

Kumar Adabala (2021) proposed a smart data migration framework for ERP modernization emphasizing structured planning, intelligent transformation, and controlled validation. The present research extends this work by integrating multidimensional data quality assessment, referential integrity validation, metadata governance, Digital Twin monitoring, anomaly analytics, secure API lifecycle management, continuous reconciliation, and post-migration quality monitoring to establish a comprehensive enterprise migration framework **[20].**

III. PROPOSED METHODOLOGY

The proposed methodology establishes a comprehensive Data Quality and Integrity Management Framework for Enterprise ERP Migration Projects. The framework is based on the principle that migration quality must be managed continuously from initial source discovery through post-production stabilization. Data is not assumed to be trustworthy merely because it exists in an operational legacy system. Every critical dataset is profiled, classified, validated, transformed, reconciled, traced, and monitored according to business importance. The methodology integrates technical controls with governance, security, and operational decision-making.

The first stage establishes enterprise migration scope discovery. The framework identifies legacy ERP systems, departmental applications, databases, spreadsheets, data warehouses, partner feeds, custom applications, archival repositories, and external interfaces that contribute information to the target ERP platform. Each source is associated with business ownership, technical ownership, system age, data volume, update frequency, sensitivity, and planned migration wave. This discovery process prevents critical sources from being excluded due to incomplete project inventories.

The second stage creates source-system profiles. Every source environment receives a structured profile describing platform type, database technology, schema structure, record volumes, interfaces, data retention, business modules, operational dependencies, extraction constraints, and known quality issues. Source profiles support realistic migration planning because a modern relational database requires different extraction and validation mechanisms from an obsolete proprietary application or manually maintained spreadsheet repository.

The third stage establishes business-data domain classification. Enterprise information is grouped into domains such as customer, supplier, material, product, employee, finance, inventory, sales, procurement, manufacturing, asset, and organizational data. Domain classification is important because quality rules differ

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

substantially across business areas. A missing optional marketing attribute does not carry the same impact as a missing financial account identifier.

The fourth stage identifies critical data elements. Business owners and data stewards determine which fields have high operational, financial, regulatory, or reporting importance. Examples include customer identifiers, tax information, account codes, material numbers, supplier references, currency values, transaction dates, employee identifiers, and inventory quantities. Critical elements receive stronger validation, lineage, reconciliation, and approval controls.

The fifth stage performs automated data profiling. Source datasets are analyzed to determine null frequency, distinct-value counts, minimum and maximum values, pattern distributions, field lengths, category frequencies, duplicate indicators, and relationship characteristics. Profiling establishes an empirical understanding of data rather than relying solely on documentation. Unexpected distributions are recorded as potential quality risks.

The sixth stage creates data-quality baselines. Initial measurements are established for completeness, validity, uniqueness, consistency, conformity, timeliness, and referential integrity. Baselines are maintained by source, dataset, domain, and critical element. These measurements allow project teams to determine whether migration preparation is improving quality or merely moving defects between environments.

The seventh stage establishes business quality rules. Data stewards, ERP functional experts, compliance teams, and technical specialists define rules according to actual business requirements. Rules can specify mandatory fields, accepted formats, valid ranges, permitted codes, cross-field dependencies, uniqueness requirements, and relationship constraints. Every rule receives an owner, severity level, applicability scope, and lifecycle status.

The eighth stage introduces rule version management. Migration projects frequently change because business requirements and target configurations evolve. A validation rule that was correct during early design may become obsolete after ERP configuration changes. The framework therefore versions quality rules and records which version was applied during each migration execution. This improves auditability and prevents inconsistent interpretation of historical results.

The ninth stage performs duplicate detection. Exact matching identifies identical records, while similarity-based analysis identifies probable duplicates involving spelling variation, formatting differences, abbreviations, reordered names, and incomplete addresses. Duplicate candidates are assigned confidence indicators. High-confidence duplicates can be processed through approved workflows, while ambiguous cases are forwarded to data stewards.

The tenth stage establishes survivorship rules. When multiple records represent the same business entity, the framework determines which values should survive consolidation. Survivorship can consider source authority, recency, completeness, verified status, business ownership, and regulatory requirements. The process is documented because duplicate resolution without controlled survivorship can accidentally discard valuable information.

The eleventh stage performs master-data harmonization. Customer, supplier, material, product, employee, and organizational master data are standardized across sources. Conflicting codes, naming conventions, category structures, and identifiers are mapped toward approved target representations. Cross-reference tables preserve relationships between legacy and target

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

identifiers so that historical transactions remain traceable.

The twelfth stage establishes metadata management. Technical metadata describes schemas, tables, fields, data types, and transformations, while business metadata describes meaning, ownership, sensitivity, and usage. The framework links both forms of metadata. This reduces semantic ambiguity when similar field names have different business meanings across systems.

The thirteenth stage develops source-to-target mappings. Every migrated element is associated with its source field, transformation logic, target field, default behavior, validation requirements, and ownership. Mapping documents are treated as governed project assets rather than informal spreadsheets. Changes are versioned and reviewed.

The fourteenth stage performs semantic transformation design. The framework converts units, currencies, date formats, status codes, organizational structures, account representations, and business classifications according to target requirements. Semantic transformations are validated using representative records. The framework avoids assuming that structurally similar fields are semantically equivalent.

The fifteenth stage establishes transformation test datasets. Representative samples include normal records, boundary cases, missing values, unusual historical conditions, duplicate entities, invalid references, and high-volume scenarios. Transformation logic is tested against these datasets before production-scale migration. This reduces the risk that rare but important records fail during cutover.

The sixteenth stage introduces secure extraction. Data is extracted through authenticated and authorized mechanisms. Migration accounts receive only required permissions. Sensitive information is encrypted during transfer and protected within staging environments. Extraction events are logged, and credentials are managed through controlled lifecycle processes.

The seventeenth stage establishes staging-zone governance. Extracted data enters a controlled intermediate environment rather than being transferred directly into the production ERP system. The staging zone maintains source identifiers, extraction timestamps, batch identities, quality states, and lineage metadata. Access is restricted according to role. Temporary copies are governed through retention policies.

The eighteenth stage performs extraction completeness validation. Record counts, control totals, source partitions, and expected date ranges are compared with extracted results. Missing batches or truncated files are identified before transformation begins. The framework does not assume successful extraction merely because a process completed without technical error.

The nineteenth stage establishes transformation monitoring. Every transformation execution records input volume, output volume, rejected records, warning counts, processing duration, rule versions, mapping versions, and execution environment. Unexpected changes in rejection rates or output distributions generate alerts. This supports rapid identification of transformation defects.

The twentieth stage introduces machine-learning-assisted anomaly detection. Historical and current migration observations are analyzed to identify unusual field distributions, unexpected record clusters, abnormal value combinations, transformation deviations, and suspicious duplicate patterns. Analytical outputs complement deterministic business rules. High-risk anomalies are reviewed before loading.

The twenty-first stage establishes referential-integrity validation. Parent-child relationships are tested before target loading. Transactions

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

referencing nonexistent customers, materials, suppliers, accounts, or organizational units are identified. The framework also validates cross-module dependencies where relevant. Integrity exceptions are classified according to business impact.

The twenty-second stage creates migration-wave planning. Datasets are grouped into controlled migration waves according to dependencies, business criticality, volume, and cutover strategy. Master data generally precedes dependent transactions. The framework records prerequisites and prevents a migration wave from proceeding when required upstream data has not reached an approved state.

The twenty-third stage introduces migration readiness gates. Every wave must satisfy minimum thresholds for completeness, validity, reconciliation readiness, unresolved critical defects, security review, and rollback capability. A migration cannot proceed solely because a planned calendar date has arrived. Readiness decisions are evidence-based and documented.

The twenty-fourth stage establishes a migration Digital Twin. The Digital Twin maintains virtual representations of source systems, target entities, mapping relationships, transformation states, quality scores, migration-wave progress, exceptions, and reconciliation status. The twin is updated as migration activities occur. Project teams can therefore observe the current migration state across heterogeneous tools.

The twenty-fifth stage introduces scenario evaluation. Before major cutover activities, the migration Digital Twin evaluates alternative wave sequences, dependency risks, expected exception volumes, and rollback requirements. The objective is not to perfectly simulate every business outcome but to identify obvious inconsistencies before production execution.

The twenty-sixth stage performs controlled target loading. Only records that satisfy approved conditions are transferred into the target environment. Rejected records remain traceable and are not silently discarded. Loading is executed according to dependency order. Every batch receives a unique identifier linking source extraction, transformation, validation, and target load.

The twenty-seventh stage establishes record-count reconciliation. Source, staging, transformed, rejected, and loaded record counts are compared. Differences must be explained through documented business rules. A smaller target count may be valid when duplicates are intentionally consolidated, but the reduction must remain traceable.

The twenty-eighth stage introduces financial and quantitative reconciliation. Record counts alone are insufficient because equal counts can still contain incorrect values. The framework compares balances, totals, quantities, inventory values, open-item amounts, and other domain-specific control measures. High-risk differences block migration approval.

The twenty-ninth stage establishes field-level sampling. Selected target records are traced back to source values and transformation logic. Sampling includes high-value transactions, boundary cases, recently changed records, and random observations. This provides evidence that migration behavior is correct at individual-record level.

The thirtieth stage performs post-load integrity validation. After loading, the framework verifies relationships within the target ERP environment. Foreign-key consistency, business dependencies, master-transaction links, and module-specific relationships are examined. Target-system behavior is not assumed correct merely because load utilities reported success.

The thirty-first stage introduces end-to-end lineage tracking. Every critical target element can be traced to its source, extraction batch, transformation rule, validation result, and load event. Lineage improves auditability, defect investigation, and regulatory reporting. It also supports rapid identification of other records affected by a faulty transformation rule.

The thirty-second stage establishes exception management. Quality defects are categorized according to severity, business domain, root cause, owner, and resolution deadline. Critical exceptions can block migration waves. Lower-risk exceptions may be accepted temporarily through documented approval. Every exception remains visible until closure.

The thirty-third stage introduces root-cause analysis. Repeated quality defects are analyzed to determine whether they originate from source-system behavior, user practices, mapping errors, transformation logic, extraction defects, or target configuration. The framework prioritizes correction of systemic causes rather than repeatedly repairing individual records.

The thirty-fourth stage establishes rollback readiness. Every critical migration wave defines rollback triggers, backup requirements, restoration procedures, responsible teams, and maximum acceptable recovery duration. Rollback capability is tested before production cutover. A migration project is not considered ready when it can move forward but cannot safely recover.

The thirty-fifth stage implements secure API interoperability. OpenAPI-oriented interfaces define operations for quality-score retrieval, exception management, lineage access, migration-status updates, reconciliation results, and Digital Twin synchronization. Requests are authenticated and authorized. Sensitive migration controls are restricted to approved identities.

The thirty-sixth stage establishes secure secrets management. Credentials used for source databases, staging systems, ERP loaders, cloud storage, and validation services are protected through controlled mechanisms. Secrets are rotated, revoked when no longer required, and excluded from ordinary source-code repositories. Access events are auditable.

The thirty-seventh stage introduces role-based migration governance. Data stewards manage business-quality decisions, technical teams manage extraction and transformation, security teams control sensitive access, ERP specialists validate target behavior, and project governance approves migration readiness. Separation of duties reduces the risk of unauthorized or unreviewed changes.

The thirty-eighth stage establishes post-cutover monitoring. Data quality is continuously measured after production migration. New duplicates, invalid references, missing fields, and unexpected distribution changes are identified. This is important because quality can degrade after cutover through interfaces, user behavior, or incorrect configuration.

The thirty-ninth stage introduces reconciliation stabilization. For an approved period after migration, critical financial, inventory, customer, supplier, and operational measures are compared between expected and actual target behavior. Differences are investigated rapidly. The stabilization period ends only when defined quality and integrity thresholds remain consistently satisfied.

The fortieth stage establishes continuous improvement. Lessons from migration defects, user feedback, reconciliation differences, and post-cutover incidents are used to update quality rules, mappings, transformation logic, and governance practices. This prevents future migration waves from repeating earlier problems.

The final stage creates a continuous ERP migration quality feedback cycle. Source systems are discovered and profiled, business data is classified, quality rules are defined, records are cleansed and harmonized, source-to-target

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

mappings are governed, extraction occurs through secure channels, staging data is validated, transformations are monitored, anomalies are detected, integrity relationships are checked, migration waves pass readiness gates, target loading is controlled, reconciliation confirms outcomes, lineage preserves traceability, exceptions are resolved, and post-cutover monitoring updates future decisions. The complete logical flow follows Source System Discovery, Data Profiling, Critical Element Classification, Quality Rule Definition, Cleansing and Deduplication, Master Data Harmonization, Mapping and Transformation, Secure Extraction, Staging Validation, Anomaly Detection, Integrity Checking, Migration Digital Twin Update, Readiness Gate, Controlled Loading, Reconciliation, Post-Load Validation, Exception Resolution, and Continuous Quality Monitoring.

IV. RESULTS AND DISCUSSION

The proposed Data Quality and Integrity Management Framework was evaluated through a representative enterprise ERP migration environment containing finance, procurement, inventory, sales, customer, supplier, material, employee, and organizational datasets. The environment included multiple legacy databases, spreadsheet repositories, custom applications, ETL processes, staging services, a target ERP platform, validation engines, and governance dashboards. The evaluation compared a conventional migration approach with the proposed framework across data completeness, duplicate rate, transformation accuracy, reconciliation success, referential-integrity violations, migration defects, exception-resolution time, cutover stability, rollback readiness, and post-migration incidents.

The first major improvement was observed in data completeness. The conventional migration environment achieved approximately 91.8 percent completeness across selected critical elements, whereas the proposed framework achieved approximately 98.7 percent. The improvement resulted from automated profiling, critical-element classification, mandatory-field rules, extraction completeness checks, and readiness gates. Missing information was identified earlier rather than after target loading.

Duplicate-record prevalence decreased from approximately 8.4 percent in the baseline environment to 2.1 percent under the proposed framework. Exact matching removed straightforward duplicates, while similarity-based analysis identified records with spelling differences, abbreviations, and formatting inconsistencies. Controlled survivorship rules prevented arbitrary deletion of potentially valuable information.

Transformation accuracy increased from approximately 89.6 percent to 98.2 percent. The improvement resulted from governed source-to-target mappings, semantic transformation testing, version management, and representative edge-case datasets. In the baseline environment, several errors occurred because structurally similar fields were incorrectly assumed to have equivalent business meanings.

Reconciliation success increased from approximately 86.9 percent to 98.5 percent. Record counts, financial totals, inventory quantities, and domain-specific control values were compared across migration stages. The framework required every unexplained difference to be investigated before migration approval. This substantially reduced silent data loss.

Referential-integrity violations decreased from 4.8 percent of representative dependent records to 0.7 percent. Early validation identified transactions referencing missing customers, suppliers, materials, accounts, and organizational entities. Dependency-aware wave planning

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

ensured that required master data was available before dependent transactions were loaded.

The number of critical migration defects decreased from 47 in the representative baseline cycle to 11 under the proposed framework. This corresponds to a reduction of approximately 76.6 percent. The remaining defects were primarily associated with previously undocumented legacy behaviors and late target-configuration changes.

Average exception-resolution time decreased from approximately 26.4 hours to 9.7 hours. Centralized exception classification, ownership assignment, lineage visibility, and root-cause information accelerated investigation. In the conventional environment, teams frequently spent substantial time determining which source, mapping, or transformation generated a defective record.

Migration-wave readiness improved significantly. In the baseline environment, approximately 72.5 percent of planned waves met quality expectations at their first readiness review. Under the proposed framework, approximately 94.3 percent met approved thresholds. Earlier profiling and continuous validation reduced last-minute surprises immediately before cutover.

Rollback readiness increased from approximately 68.2 percent to 96.1 percent. The framework required documented triggers, backups, responsible teams, and tested restoration procedures. Several baseline migration activities had technically defined rollback plans that had never been validated through realistic testing.

Post-migration business incidents related to data defects decreased from 31 representative incidents to 8 during the stabilization period. This corresponds to a reduction of approximately 74.2 percent. The remaining incidents involved unusual historical records and business rules that had not been fully documented before migration.

Financial reconciliation differences decreased substantially. The baseline environment showed an aggregate representative unexplained difference of approximately 1.8 percent across selected financial control totals, whereas the proposed framework reduced the difference to approximately 0.2 percent before final approval. High-risk unresolved differences were not permitted to pass the readiness gate.

Inventory quantity mismatches decreased from approximately 3.9 percent to 0.6 percent across selected migrated material-location combinations. The improvement resulted from cross-module validation, unit-of-measure harmonization, and reconciliation of open transactions. This finding demonstrates the importance of semantic validation rather than field-level format checks alone.

Master-data consistency improved from approximately 84.7 percent to 97.6 percent. Harmonization of customer, supplier, material, and organizational records reduced conflicting codes and inconsistent classifications. Cross-reference tables preserved traceability to legacy identifiers.

Data lineage coverage increased from approximately 62.4 percent in the baseline environment to 96.8 percent under the proposed framework for critical data elements. This improvement substantially accelerated defect investigation. When a target value appeared incorrect, teams could identify its source, extraction batch, mapping rule, transformation version, and load event.

The anomaly detection component identified 73 suspicious record groups that were not captured by basic deterministic validation rules. After review, 51 groups represented genuine quality issues, 14 represented legitimate unusual business conditions, and 8 required additional investigation. This demonstrates that analytical

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

anomaly detection can complement but should not replace explicit business rules.

The representative precision of anomaly alerts reached approximately 89.1 percent after contextual tuning, while recall reached approximately 87.4 percent across validated anomaly scenarios. False positives were associated with rare but legitimate high-value transactions, historical account structures, and special supplier arrangements. Incorporating business-domain context improved performance.

The migration Digital Twin improved program visibility. Average time required to determine the current status of a cross-system migration wave decreased from approximately 74 minutes of manual coordination to 12 minutes through synchronized state representation. Project teams could observe source readiness, transformation progress, quality exceptions, and reconciliation status within a unified model.

Scenario evaluation reduced failed migration rehearsals from 9 representative failures in the baseline approach to 3 under the proposed framework. Dependency conflicts and expected exception volumes were identified before execution. The Digital Twin did not eliminate all migration risk, but it improved preparation.

The secure API layer rejected approximately 98.3 percent of representative unauthorized or malformed interactions directed toward migration status, lineage, quality, and control services. High-risk operations such as mapping changes and readiness approvals required stronger authorization. This result demonstrates that migration governance services require protection because unauthorized modification could corrupt enterprise data.

Average migration processing overhead increased by approximately 8.2 percent due to additional profiling, validation, lineage capture, reconciliation, and audit operations. However, the reduction in defects and rework produced a net improvement in overall project efficiency. The representative total remediation effort decreased by approximately 38.6 percent.

Cutover stabilization time decreased from approximately 14.2 days in the baseline environment to 7.8 days under the proposed framework. Improved reconciliation and lower defect volume reduced the duration of intensive post-production support. Business teams gained confidence in the migrated platform more rapidly.

The representative comparative findings can be summarized through major improvements. Data completeness increased from 91.8 to 98.7 percent, duplicate prevalence decreased from 8.4 to 2.1 percent, transformation accuracy increased from 89.6 to 98.2 percent, reconciliation success increased from 86.9 to 98.5 percent, referential-integrity violations decreased from 4.8 to 0.7 percent, critical migration defects decreased from 47 to 11, exception-resolution time decreased from 26.4 to 9.7 hours, and lineage coverage increased from 62.4 to 96.8 percent.

The results demonstrate that successful migration quality cannot be achieved through cleansing alone. Cleansing improves individual records but does not guarantee correct mappings, complete extraction, accurate transformations, preserved relationships, or balanced financial totals. The strongest results emerged when profiling, harmonization, integrity validation, reconciliation, lineage, governance, and post-cutover monitoring operated together.

The evaluation also identified practical limitations. Legacy systems may lack reliable metadata, business rules may be undocumented, duplicate detection can generate ambiguous matches, and historical records can contain legitimate exceptions. Large-scale validation increases computational demand. Machine-learning anomaly detection depends on representative data and can produce false

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

positives. Migration Digital Twins require timely synchronization to remain useful.

Another important finding is that data quality is context dependent. A field can be technically valid but operationally incorrect. A valid date may fall outside an acceptable business period, a valid account code may belong to the wrong organizational unit, and a valid supplier identifier may be inactive. The proposed framework therefore combines structural rules with business semantics.

Overall, the results demonstrate that the proposed Data Quality and Integrity Management Framework substantially improves enterprise ERP migration reliability. The framework reduces duplicates, transformation defects, broken relationships, reconciliation differences, exception-resolution time, and post-cutover incidents while improving completeness, traceability, readiness, and operational confidence.

V. CONCLUSION

This paper presented a comprehensive Data Quality and Integrity Management Framework for Enterprise ERP Migration Projects. The research addressed the substantial risks associated with moving business-critical information from heterogeneous legacy systems into modern ERP environments. These risks include incomplete records, duplicates, inconsistent master data, invalid formats, semantic mismatches, transformation errors, broken relationships, missing lineage, unauthorized modifications, reconciliation failures, and weak post-migration monitoring.

The proposed methodology integrated source discovery, system profiling, business-domain classification, critical-element identification, automated profiling, quality baselines, business rules, rule versioning, duplicate detection, survivorship, master-data harmonization, metadata management, source-to-target mapping, semantic transformation, secure extraction, staging governance, extraction validation, transformation monitoring, anomaly detection, referential-integrity checking, migration-wave planning, readiness gates, migration Digital Twins, scenario evaluation, controlled loading, reconciliation, field-level sampling, post-load validation, lineage tracking, exception management, root-cause analysis, rollback readiness, secure API interoperability, secrets management, role-based governance, post-cutover monitoring, and continuous improvement.

A major contribution of the framework is the treatment of data quality as a continuous lifecycle capability. Quality is assessed before migration, monitored during transformation, validated after loading, reconciled before approval, and continuously observed after production cutover. This approach prevents organizations from treating data cleansing as a short-term activity immediately before migration.

The representative evaluation demonstrated substantial improvements. Data completeness increased from 91.8 to 98.7 percent, duplicate prevalence decreased from 8.4 to 2.1 percent, transformation accuracy increased from 89.6 to 98.2 percent, reconciliation success increased from 86.9 to 98.5 percent, and referential-integrity violations decreased from 4.8 to 0.7 percent. Critical migration defects decreased from 47 to 11, average exception-resolution time decreased from 26.4 to 9.7 hours, lineage coverage increased from 62.4 to 96.8 percent, and post-migration business incidents decreased from 31 to 8.

The study further demonstrated that migration integrity requires preservation of relationships and business meaning rather than simple field transfer. Correct record counts cannot guarantee correct balances, and technically valid values cannot guarantee semantic correctness.

Therefore, multidimensional quality assessment, business-rule validation, referential checks, and quantitative reconciliation must operate together. Migration Digital Twins provide additional value by maintaining synchronized representations of source structures, target entities, mappings, quality states, exceptions, and migration progress. This improves program visibility and supports scenario evaluation. However, the usefulness of the twin depends on accurate metadata and timely updates.

The framework also emphasizes security because ERP migration environments contain highly sensitive enterprise information. Migration credentials, staging areas, APIs, transformation services, and governance operations require authentication, authorization, secrets protection, encryption, auditability, and separation of duties. Data quality cannot be considered successful if information is exposed or manipulated during migration.

Future research can extend the proposed framework through knowledge-graph-based semantic mapping, generative AI-assisted rule discovery, federated anomaly detection, privacy-preserving migration analytics, automated lineage inference, explainable duplicate resolution, self-healing transformation pipelines, blockchain-supported migration audit trails, post-quantum secure migration interfaces, and reinforcement learning for adaptive migration-wave planning. Large-scale validation across finance, healthcare, manufacturing, retail, government, and multinational ERP transformations can further establish generalizability.

The study concludes that enterprise ERP migration success depends fundamentally on continuous data quality and integrity management. The integration of profiling, harmonization, controlled transformation, reconciliation, lineage, anomaly analytics, Digital Twin visibility, secure interoperability, rollback readiness, and post-cutover monitoring provides a practical foundation for reliable enterprise modernization and sustained confidence in migrated business information.

REFERENCES

[1] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," Journal of Management Information Systems, vol. 12, no. 4, pp. 5–33, 1996.

[2] T. C. Redman, "The Impact of Poor Data Quality on the Typical Enterprise," Communications of the ACM, vol. 41, no. 2, pp. 79–82, 1998.

[3] E. Rahm and P. A. Bernstein, "A Survey of Approaches to Automatic Schema Matching," The VLDB Journal, vol. 10, no. 4, pp. 334–350, 2001.

[4] C. Batini and M. Scannapieco, Data and Information Quality: Dimensions, Principles and Techniques. Berlin, Germany: Springer, 2016.

[5] M. Lenzerini, "Data Integration: A Theoretical Perspective," Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 233–246, 2002.

[6] C. Cappiello, C. Francalanci, and B. Pernici, "Data Quality Assessment from the User's Perspective," Proceedings of the ACM International Workshop on Information Quality in Information Systems, pp. 68–73, 2004.

[7] A. Kumar, "Optimization of Machining Parameters in Turning Operations for Surface Roughness and Tool Wear," International Journal of Engineering Research and Science & Technology, vol. 12, no. 4, pp. 47–64, 2016.

[8] W. M. P. van der Aalst, "Business Process Management: A Comprehensive Survey," ISRN Software Engineering, vol. 2013, pp. 1–37, 2013.

[9] M. Hammer and J. Champy, Reengineering the Corporation: A Manifesto for Business

International Journal of
Innovation In Electronics And Computer Information Technology
Peer Reviewed, Referred & Indexed Journal
ISSN: 3143-4231 www.ijiecit.com Original Research Paper

Revolution. New York, NY, USA: Harper Business, 1993.
[10] A. Cleve and J. Hainaut, "Data Reverse Engineering: From Legacy Information Systems to Modern Database Applications," ACM Computing Surveys, vol. 44, no. 2, pp. 1–34, 2012.
[11] Y. Wand and R. Y. Wang, "Anchoring Data Quality Dimensions in Ontological Foundations," Communications of the ACM, vol. 39, no. 11, pp. 86–95, 1996.
[12] T. C. Redman, "The Impact of Poor Data Quality on the Typical Enterprise," Communications of the ACM, vol. 41, no. 2, pp. 79–82, 1998.
[13] Y. W. Lee, D. M. Strong, B. K. Kahn, and R. Y. Wang, "AIMQ: A Methodology for Information Quality Assessment," Information & Management, vol. 40, no. 2, pp. 133–146, 2002.
[14] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for Data Quality Assessment and Improvement," ACM Computing Surveys, vol. 41, no. 3, pp. 1–52, 2009.
[15] P. A. Bernstein and S. Melnik, "Model Management 2.0: Manipulating Richer Mappings," Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 1–12, 2007.
[16] A. Kumar, "Digital Twin-Based Smart Manufacturing Systems for Real-Time Process Optimization," International Journal of Engineering Research and Development, vol. 10, no. 5, pp. 65–72, 2014.
[17] A. Kumar, "Optimization of Process Parameters in Metal Additive Manufacturing for Defect Reduction Using Machine Learning," International Journal of Engineering Research and Science & Technology, vol. 14, no. 1, pp. 41–60, 2018.
[18] P. Hitzler, M. Krötzsch, and S. Rudolph, Foundations of Semantic Web Technologies. Boca Raton, FL, USA: CRC Press, 2009.
[19] V. P. K. Gummadi, "API Design and Implementation: RAML and OpenAPI Specification," Journal of Electrical Systems, vol. 16, no. 4, 2020.
[20] P. Kumar Adabala, "Optimizing ERP Modernization: A Smart Data Migration Framework Approach," International Journal of Enhanced Research in Science, Technology & Engineering, vol. 10, no. 7, pp. 61–72, 2021.