

DATA QUALITY IN THE AGE OF ARTIFICIAL INTELLIGENCE: A COMPREHENSIVE REVIEW OF FRAMEWORKS, CHALLENGES, AND GOVERNANCE STRATEGIES

Mohammad Rahim Albous¹

¹Abdullah Al Salem University, KUWAIT

ABSTRACT

Data quality has emerged as a fundamental pillar of scientific integrity, organizational performance, and evidence-based decision-making in the contemporary digital landscape. As artificial intelligence systems become increasingly integrated into data-driven processes across healthcare, business, and public administration, the importance of robust data quality frameworks has intensified significantly. This comprehensive narrative review synthesizes scientific and technical literature published between 1996 and 2022, complemented by international standards including ISO/IEC 25012 and ISO 8000, to provide an integrated overview of data quality frameworks, governance structures, and ethical considerations in the era of AI. Across diverse sectors, accuracy, completeness, consistency, timeliness, and accessibility consistently emerge as universal quality dimensions that underpin reliable data utilization. Evidence from healthcare, business, and public administration demonstrates that poor data quality leads to substantial financial losses, operational inefficiencies, and erosion of stakeholder trust. Emerging frameworks increasingly integrate FAIR principles Findability, Accessibility, Interoperability, and Reusability—while incorporating ethical safeguards, including bias mitigation in AI systems. This review reveals that data quality is not solely a technical concern but fundamentally a socio-organizational challenge that requires robust governance structures and continuous assurance throughout the data lifecycle. Embedding quality and ethical governance into data management practices is crucial for producing trustworthy, reusable, and reproducible data that supports sound science and informed decision-making across all sectors.

KEYWORDS: Data Quality; Data Governance; FAIR Principles; Data Lifecycle; Artificial Intelligence Ethics; Healthcare Data; Data Management; ISO Standards

INTRODUCTION

In today's data-driven world, the quality of information has become a decisive factor in

advancing scientific knowledge, ensuring reliable decision-making, and achieving organizational success. Across healthcare, business, and public administration alike, low-quality data can lead to flawed conclusions, wasted resources, and significant social and economic harm. Data quality is commonly understood as the extent to which data are fit for use in supporting operations, planning, and decision-making. This definition highlights that quality is not an absolute property but rather a context-dependent attribute that must be assessed relative to specific goals and user requirements. Frameworks such as ISO/IEC 25012 and ISO 8000 define data quality in terms of measurable dimensions, including accuracy, completeness, consistency, and other attributes that collectively determine fitness for purpose.

Most studies treat technical, ethical, and sectoral aspects of data quality separately, with integrated approaches remaining scarce. This fragmentation hinders practitioners who must navigate disparate studies to implement comprehensive data management practices. The separation of these dimensions creates challenges for organizations seeking to develop holistic approaches to data quality that address both technical requirements and broader organizational and ethical considerations. This review addresses this gap by synthesizing literature on data quality across three interconnected domains: standards and frameworks, sectoral applications, and emerging links with governance, ethics, and artificial intelligence.

The importance of data quality for research cannot be overstated, as ensuring data quality is a prerequisite for transparency, reproducibility, and accountability in the scientific process. Robust data quality practices enhance the traceability of datasets and the credibility of scientific conclusions, aligning with broader initiatives such as the FAIR principles, which stipulate that data should be Findable, Accessible, Interoperable, and Reusable. These principles have become foundational for modern data management, providing a framework for ensuring that data remain valuable and usable over time.

This narrative review offers a unique contribution by synthesizing three areas that are seldom connected: established theoretical frameworks, practical applications in the business world, and emerging trends such as ethical governance and the use of artificial intelligence in data management. Unlike works that focus on a single dimension, this review offers a holistic perspective that connects technical challenges with their organizational and ethical implications, underscoring their direct impact on scientific reproducibility and strategic decision-making. Data quality is typically assessed through a set of measurable dimensions—accuracy, completeness, consistency, timeliness, relevance, and validity—which collectively provide a framework for identifying and addressing potential issues. These dimensions are widely applied across domains, including customer relationship management, where failures in

implementation have often been linked to poor data quality, underscoring the importance of robust evaluation tools and processes.

MATERIAL AND METHODS

This comprehensive narrative literature review was conducted to synthesize the evolution of data quality concepts, frameworks, and practices, explicitly bridging academic theory and industry- and standards-based practices. The reporting adheres to good-practice elements for transparent search descriptions, with the time window of 1996 to 2022 chosen to capture foundational work, including the consumer-centric fitness for use perspective emerging in the mid-1990s, and contemporary developments in AI, ethics, and standards. Searches were last updated on October 8, 2022.

The adoption of a narrative rather than systematic review design was justified for several reasons. First, the field of data quality spans diverse domains including healthcare, business, public administration, and software engineering, with heterogeneous methodologies that preclude meta-analytic synthesis. Second, the objective was conceptual integration—connecting theoretical frameworks with practical applications and emerging trends—rather than estimating quantitative effects. Third, the deliberate inclusion of grey literature alongside peer-reviewed articles was essential to capture industry practices and normative frameworks that systematic reviews typically exclude.

The 1996 to 2022 timeframe was strategically chosen to encompass three critical periods: the foundational consumer-centric perspective of data quality, the formalization of international standards including the ISO 8000 series and ISO/IEC 25012, and contemporary developments in AI ethics, governance frameworks, and FAIR principles for research data. This 29-year window captures both the field's historical evolution and its current state.

Database selection balanced comprehensiveness with relevance. PubMed provides coverage of biomedical and health sciences, Scopus offers multidisciplinary breadth with strong indexing of computer science and business literature, Web of Science ensures access to high-impact journals across disciplines, and Google Scholar facilitates the discovery of grey literature, preprints, and non-indexed sources. Together, these sources cover the theoretical, applied, and standards-based literature essential to the synthesis.

Search strings combined controlled vocabulary and free-text terms for the core concepts of data quality, dimensions, governance, ethics, and FAIR principles. For PubMed, searches combined MeSH terms and free-text keywords including data quality, information quality, artificial intelligence, governance, ethics, and FAIR, with filters applied for publication date and article type. For Scopus and Web of Science, searches

were adapted using platform-specific field tags with document type filters applied. Google Scholar searches employed simplified query strings due to platform limitations, with the first 200 results screened by relevance.

All records were exported and deduplicated using EndNote and then exported to Rayyan for screening. Screening proceeded in two stages: title and abstract screening for relevance, followed by full-text assessment to confirm eligibility. Screening was performed in duplicate, with disagreements resolved by discussion and senior author adjudication as needed. A piloted data extraction form captured bibliographic data, sector and study type, definitions of data quality, dimensions and measurement approaches, governance and ethical components, alignment with FAIR and standards, tools and metrics, and reported outcomes and use cases. Extraction was conducted independently by two reviewers with consensus adjudication. The findings were synthesized thematically, highlighting dimensions, methodologies, and applications of data quality across the reviewed literature.

DEFINITIONS AND FRAMEWORKS OF DATA QUALITY

The conceptualization of data quality has evolved significantly over the past three decades, with various frameworks and standards emerging to provide structure and guidance for organizations seeking to manage data as a strategic asset. Understanding these foundational frameworks is essential for appreciating how data quality is assessed and assured in practice.

ISO 8000, the international data quality standard, originated in the NATO Codification System, a framework for cataloging material items with uniform descriptions and codes that improves inventory management and supply chain efficiency. In 2004, NATO Allied Committee 135 and the Electronic Commerce Code Management Association formalized its promotion as the basis for an international standard. This collaboration led to ISO 8000, which built on the NATO Codification System's emphasis on precise, standardized data descriptions and management practices refined over decades. ISO 8000 defines quality data as portable data that meets stated requirements. Data must be separable from a software application and comply with explicit, unambiguous requirements set by the user or consumer. The standard provides comprehensive frameworks for both conceptualizing and measuring data quality, thereby aligning with broader initiatives in data portability and preservation and enabling organizations to define, request, and verify data quality using standard conventions.

Complementing ISO 8000, ISO 25012 provides a foundational model with fifteen dimensions of data quality, including accuracy, completeness, consistency, and traceability. Initially designed for software, this model is now widely applied across domains, providing a comprehensive framework for assessing the multiple dimensions of data quality that are relevant to diverse organizational contexts.

Wang and Strong developed a widely used framework with four dimensions that has been highly influential in shaping subsequent thinking about data quality. The intrinsic dimension addresses whether data should be accurate, reliable, and error-free. The contextual dimension recognizes that relevance depends on the task or use case. The representational dimension emphasizes that clarity and proper formatting enable correct interpretation. The accessibility dimension ensures that data must be available to users when and where needed. This framework has been particularly influential because it recognizes that data quality is multidimensional and that different dimensions may be more or less important depending on the specific context of use.

Building on this foundational work, subsequent researchers have extended and refined data quality frameworks. Strong, Lee, and Wang emphasized data quality as fitness for use in operations, decision-making, and planning, extending the earlier framework with an organizational perspective. Pipino, Lee, and Wang introduced practical tools for data quality assessment, defining data quality through measurable attributes that reflect accuracy, completeness, consistency, and timeliness. Ehrlinger and Wöß expanded the framework beyond classical dimensions, highlighting timeliness, completeness, plausibility, integrity, and multifacetedness while recognizing that data quality is a multidimensional construct influenced by context and use.

Haug and Zachariassen offered a cost-oriented perspective, suggesting that perfect data quality is neither achievable nor optimal. Instead, the right level of quality balances costs of maintenance against costs of poor data. This pragmatic perspective recognizes that organizations must make strategic decisions about the level of data quality they pursue, balancing the benefits of higher quality against the costs of achieving and maintaining it.

In data analytics, data quality is paramount because analytical models and algorithms depend on accurate and complete inputs to generate meaningful insights and support sound decision-making. Models trained on flawed data can yield unreliable predictions, leading to misguided strategies. Conversely, high-quality data enhances the accuracy, fairness, robustness, and scalability of machine learning models, underscoring the importance of specialized tools in the era of data-centric AI. Within customer relationship management, data quality is especially critical. Reliable and consistent datasets enable accurate predictive modeling, such as customer lifetime value estimations and RFM analysis, which supports effective segmentation, investment optimization, and customer equity assessment systems.

DIMENSIONS OF DATA QUALITY

A comprehensive understanding of data quality requires examination of its core dimensions. These dimensions provide operational criteria for assessing and assuring quality across diverse contexts and applications.

Accuracy refers to the extent to which data represent the real-world construct they describe. It can be quantified through error rates, field- and record-level measures, and newer metrics that distinguish between random and systematic errors. Accuracy is especially critical for ensuring reliable insights in analytics and marketing, where decisions must reflect actual market conditions. Accuracy is particularly challenging in CRM contexts, where customer attributes change frequently, requiring ongoing validation procedures to maintain reliable profiles. Ensuring accuracy requires comparison against authoritative sources or ground truth, with measurement approaches including field-level validation, cross-referencing with external data sources, and statistical methods for detecting systematic errors.

Completeness refers to the presence of all required data for a given purpose. In CRM systems, completeness enables the creation of reliable customer profiles and supports accurate analytical outcomes, thereby reducing the risk of biased segmentation or decision-making. Completeness can be measured as the ratio of available values to required values, or the percentage of missing fields within a dataset. Ensuring completeness requires clear definitions of required data, validation at the point of entry, and processes for identifying and addressing gaps in existing data.

Consistency refers to the uniform representation of data across sources and systems, ensuring that values are not in conflict. Inconsistent data can lead to confusion, errors in analysis, and incorrect conclusions. Ensuring consistency requires cross-field and cross-database validation checks, standardized data definitions, and processes for reconciling differences between sources.

Timeliness concerns whether data are sufficiently up to date to support effective decision-making. Real-time data is vital in domains such as stock trading, where rapid changes demand instant responses, while daily or weekly updates may be sufficient for inventory management. Timeliness depends on collection methods, processing speed, and update frequency, requiring organizations to minimize latency and ensure data is current. A related challenge is data decay, the gradual obsolescence of information, which is seen in marketing where up to forty percent of email addresses change every two years. Mitigating decay and maintaining timeliness are crucial for reliable analytics, resource efficiency, and improved decision-making.

Relevance assesses how well data supports specific business needs and decision-making. Data must serve its intended purpose, aligning with organizational objectives and challenges. Relevance is highly context-dependent, varying across different organizational areas. Marketing departments prioritize demographics, purchasing behavior, and market trends to craft effective strategies, while financial teams focus on cash flow, revenue, and expenses for accurate planning. Thus, relevance is not static but shifts with users' goals and requirements. Measuring relevance ensures that

collected information contributes to informed decisions through tracking dataset usage, assessing time to analysis, and gathering user feedback to identify gaps.

Validity ensures that data accurately represent real-world constructs and comply with predefined standards and business rules. It assesses conformity to formats, ranges, and logical relationships, including content validity, construct validity, and criterion-related validity. Invalid data can cause flawed analyses, poor decisions, and operational disruptions, with examples including failed deliveries to incorrect addresses, misleading financial data that affects investors, or inaccurate market research leading to wasted resources. Validity also supports integration and interoperability, as adherence to shared rules and formats facilitates combining data from multiple sources. In the era of big data and AI, validity is increasingly critical because high-quality inputs underpin reliable models and algorithms.

IMPACT AND EVIDENCE: CONSEQUENCES AND CASE STUDIES

Organizations that neglect data quality face serious and multifaceted consequences. Research demonstrates that poor data quality generates quantifiable harm across multiple dimensions, including financial losses, compromised decision-making, operational inefficiency, reputational damage, and missed strategic opportunities.

The financial consequences of poor data quality are substantial and well-documented. Industry analysts estimate annual losses of approximately thirteen million dollars per organization, with U.S. businesses collectively losing over three trillion dollars annually due to data quality issues. Industry-specific studies indicate that organizations can lose up to six percent of their sales due to poor customer data management. In CRM systems, inaccurate or inconsistent records distort insights, inflate costs, and reduce efficiency, often requiring costly cleansing and manual intervention that exponentially exceed the costs of prevention.

Low-quality data leads directly to flawed segmentation, poor targeting, and reduced strategic effectiveness. In CRM systems, outdated contact details or incorrect names undermine marketing efforts and customer relationship strategies. Industry experts estimate that data quality challenges cost companies ten to thirty percent of their annual revenue. Cleaning inaccurate or incomplete data is a significant and resource-intensive task, with manual and automated correction efforts consuming disproportionate organizational resources. In CRM systems, poor validation can erode user trust and compromise adoption, particularly in small and medium-sized enterprises where frequent inconsistencies are typical.

Incorrect interactions or failed campaigns erode customer trust and damage brand image. When organizations deliver poor customer experiences due to inaccurate data, the reputational consequences extend beyond immediate financial losses. Trust, once

damaged, is difficult to restore and has a direct impact on customer loyalty and organizational competitiveness. Poor-quality data also contributes to wasted potential and missed strategic opportunities, with up to forty-five percent of generated leads deemed unusable due to duplication, invalid email addresses, or incomplete records. In healthcare, incomplete patient records can obscure critical patterns, delaying medical discoveries and compromising research progress.

Real-world cases vividly illustrate the stakes of data quality and the serious consequences of inadequate governance. One of the most striking examples occurred in space exploration, where NASA's Mars Climate Orbiter was destroyed by a mismatch between imperial and metric units, a failure of consistency that resulted in a 125 million dollar loss. This case exemplifies how inadequate data validation processes and insufficient governance can lead to catastrophic losses, underscoring that data quality is not merely a technical concern but a mission-critical issue requiring robust verification across multiple organizational levels.

A striking corporate example is the algorithmic failure at a major technology company, in which poor input data fed into machine learning models produced faulty algorithms that ultimately resulted in 110 million dollars in losses and a significant decline in stock value. The error illustrates how inadequate data quality upstream can propagate through AI systems, leading to significant business failures. This case demonstrates that AI systems are only as reliable as their training data and that governance must begin at the data ingestion stage.

Research into computerized criminal records in the United States reveals that these records are often inaccurate or incomplete, thereby undermining the reliability of the justice system. Poor information quality introduces inefficiencies and errors with serious social consequences, highlighting the need to align process design with information system requirements from the outset. This case demonstrates that even critical public-sector systems remain vulnerable to data quality failures, with organizational accountability and robust process design essential even in mission-critical public functions.

In financial institutions, legacy data silos can perpetuate duplication, inconsistency, and limited access, hampering analytics and decision-making. The decentralized storage and inconsistent data entry practices created systemic risk throughout the organization. Transitioning toward modern paradigms has improved availability and quality across the enterprise, demonstrating that organizational architecture directly impacts data quality outcomes. This case shows that data governance is as much an architectural issue as a technical one, with silos representing a structural vulnerability.

ENSURING DATA QUALITY: METHODS AND BEST PRACTICES

Ensuring high data quality requires a combination of standardized methodologies and continuous validation. Effective data collection is fundamental, with standardized procedures and tools minimizing errors at the source while validation checks confirming that data entries meet predefined criteria.

In CRM adoption among SMEs, robust data collection processes are crucial for capturing accurate and relevant customer information. This involves integrating multiple internal and external sources and ensuring that data is timely and aligned with business objectives. Systematic frameworks and staff training help maintain consistency and reduce errors such as duplicate or incomplete records, which can undermine CRM performance. Best practices include the use of automated data entry systems, regular cleaning processes to eliminate inaccuracies, and validation checks to ensure consistency and accuracy. Continuous monitoring and maintenance are crucial as data sources and systems evolve, ensuring that quality standards are consistently maintained.

Data cleaning and validation are crucial for maintaining data quality, as they involve detecting and correcting inaccuracies, inconsistencies, and missing values. Automated tools can efficiently flag anomalies, but critical production environments often require a balanced approach that integrates manual review to ensure accuracy and reliability. A multicenter clinical database study demonstrated the effectiveness of semi-automated assessments in detecting widespread issues, while minimal manual checks remained necessary to address errors missed by automation. Regular monitoring and quality reviews help sustain reliable datasets. Establishing metrics and periodic audits ensures data remains accurate, timely, and fit for analysis. In CRM systems, poor validation can erode user trust and compromise adoption, particularly in SMEs facing frequent inconsistencies. Preventive controls and improved entry processes further minimize errors.

Continuous monitoring and regular maintenance preserve data integrity by detecting anomalies, cleansing outdated records, and ensuring consistent definitions across systems. Effective collaboration among IT teams, data stewards, and business users sustains reliability throughout the data lifecycle. Advanced methods, including deep learning models, enhance the imputation of mixed-type tabular data, and data fusion improves reliability by integrating heterogeneous sources, thereby reducing uncertainty and providing more comprehensive representations.

Effective governance requires clear roles, responsibilities, and leadership commitment throughout the data lifecycle, including compliance with regulations such as GDPR. An integrated data and AI governance model spans four phases: data sourcing and preparation with ethical acquisition and bias detection, model development with fairness-aware validation, deployment and operations with drift monitoring, and

feedback and iteration to update policies dynamically. Automated checkpoints across these phases link data integrity with model ethics. Responsible governance must extend beyond compliance to embed fairness, accountability, and inclusivity. Ethical lapses can erode trust, reinforce biases, and lead to harmful outcomes across various domains, making the embedding of ethical principles within governance frameworks essential for evaluating data quality not only for efficiency but also for societal impact.

DATA QUALITY IN THE AI ERA

The emergence of artificial intelligence and machine learning as dominant paradigms for data-driven decision-making has fundamentally altered the landscape of data quality. Traditional frameworks, developed when data primarily served human analysts, must now accommodate systems in which algorithms themselves are the primary consumers of data.

Traditional definitions of data quality center on the concept of fitness for use, whereby data are evaluated in relation to their capacity to support human decision-making processes. However, the rise of AI introduces a fundamental shift: data must now be fit not only for human interpretation but also for algorithmic processing and training machine learning models. This dual requirement expands the scope of data quality beyond conventional dimensions. AI systems impose specific requirements that extend beyond traditional quality dimensions, including AI-readiness, algorithmic fairness, and model explainability and traceability.

International standards such as ISO 8000 and ISO/IEC 25012 provide foundational frameworks for data quality management, codifying dimensions such as accuracy, completeness, consistency, timeliness, and validity. However, these frameworks were developed in an era preceding widespread AI adoption and therefore do not fully address the unique governance challenges introduced by algorithmic decision-making systems. The regulatory landscape for AI governance is rapidly evolving, with comprehensive frameworks that mandate systematic approaches to AI risk management and quality assurance.

The integration of ethical considerations into data quality frameworks represents one of the most significant paradigm shifts in the era of AI. Traditional data quality management focused primarily on technical accuracy and operational efficiency. In contrast, contemporary AI governance recognizes that data quality cannot be divorced from ethical principles such as fairness, accountability, transparency, and equity. Biased training data lead directly to biased algorithms, which in turn produce discriminatory outcomes across healthcare, criminal justice, lending, hiring, and other high-stakes domains. Assessing and mitigating bias is no longer an optional ethical enhancement but a core data quality requirement.

The relationship between AI and data quality is bidirectional: AI systems both depend on high-quality data and offer powerful tools for assuring data quality at scale. However, current practices remain imbalanced, with greater emphasis on ensuring data quality for AI than on using AI to improve data quality. The AI era demands a fundamental reconceptualization of data quality, with governance frameworks that address lifecycle integration, ethical embeddedness, continuous adaptation, and sociotechnical approaches to quality management.

DISCUSSION AND CONCLUSIONS

This review confirms the persistence of inconsistencies in defining and applying data quality dimensions across various sectors. The hierarchical model offers a promising path by harmonizing terminology and adding new dimensions relevant to modern contexts. At the same time, AI and machine learning are transforming governance by automating profiling, anomaly detection, and metadata management, reducing manual intervention and improving compliance. However, most tools focus on data quality for AI, while few exploit AI for data quality, highlighting a gap between current market offerings and the need for intelligent automated data quality management.

The FAIR principles provide a widely recognized foundation for transparent management, ensuring that datasets are discoverable through rich metadata, accessible with clear usage conditions, interoperable through standardized formats, and reusable across different contexts. FAIR should be viewed as a flexible, community-driven framework for creating AI-ready digital objects that machines can process with minimal human intervention, underscoring that alignment with FAIR principles requires sustained organizational commitment beyond technical implementation.

Semantic-driven technologies, including ontologies and knowledge graphs, emphasize semantics as a standalone dimension of data quality. Ethical principles are increasingly embedded directly in AI systems through privacy-preserving algorithms such as federated learning and homomorphic encryption, which minimize risks by training or computing on distributed, encrypted data. This reflects a shift toward privacy-by-design architectures, making governance an integral part of technical systems.

Despite progress, key gaps remain. The organizational dimension of information quality in public administration is underexplored, as governance structures, staff skills, and administrative processes strongly influence reliability. Promising theoretical tools require testing in broader contexts. Measurement standards also lack harmonization, with sector-specific standards remaining fragmented. SMEs remain underexplored, with deficits in technical skills and resources limiting the adoption of robust practices.

For practitioners, embedding quality mechanisms directly into business process models offers a sustainable approach to governance, with validation and monitoring built into

workflows rather than treated as external add-ons. For policymakers and standard setters, these insights underscore the importance of explicitly validating data quality in process modeling guidelines to mitigate systemic risks. Embedding FAIR principles remains a practical pathway toward reproducibility and accountability in research.

This review demonstrates that data quality is a multidimensional concept that extends beyond technical accuracy to include contextual, representational, and accessibility dimensions, as well as emerging attributes such as governance and semantics. Ensuring data quality is therefore not only a technical task but also a strategic and socio-organizational imperative for scientific reproducibility and organizational trust. The literature consistently demonstrates that poor data quality undermines decision-making, wastes resources, erodes trust, and yields harmful outcomes. Effective practices cannot rely solely on post-hoc cleaning but must embed validation and governance across the entire data lifecycle.

Key challenges remain, including the proliferation of heterogeneous data sources, the velocity of big data, and entrenched organizational silos. While AI-driven validation tools and governance frameworks offer promising solutions, cost-benefit trade-offs often prevent systematic adoption. For practitioners, high-quality data should be treated as a strategic asset. For researchers and policymakers, stronger integration with data ethics and greater attention to underexplored contexts are urgently needed. Finally, aligning practices with FAIR principles reinforces transparency, reproducibility, and long-term societal impact.

REFERENCES

- [1] Jaladi, D. S., & Mallemapati, A. (2022). A Scalable Full-Stack. NET Enterprise Application Framework for Data Integration with Master Data Management Systems. *International Journal of AI, BigData, Computational and Management Studies*, 3(2), 169-177.
- [2] Batini, C., & Scannapieco, M. (2016). *Data and Information Quality: Dimensions, Principles and Techniques*. Springer.
- [3] 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.
- [4] Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218.
- [5] Aluri, Y. S. (2021). Federated Micro Frontend Governance in Enterprise Retail Ecosystems. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, 2(2), 114-125.
- [6] Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103–110.
- [7] Mallemapati, A., & Jaladi, D. S. (2021). A Scalable Master Data Management Architecture for Enterprise Data Integration and Governance in Full-Stack Application

- Environments. *International Journal of Emerging Research in Engineering and Technology*, 2(1), 101-110.
- [8] Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. *Frontiers in Big Data*, *5*, 850611.
- [9] Aluri, Y. S. (2022). Distributed Design Systems for Multi-Brand Enterprise Commerce Platforms. *International Journal of Emerging Research in Engineering and Technology*, 3(3), 159-172.
- [10] Haug, A., & Zachariassen, F. (2016). Data quality in CRM systems. *Industrial Management & Data Systems*, *116*(9), 1896–1914.
- [11] ISO 8000-1:2022. *Data quality — Part 1: Overview*. International Organization for Standardization.
- [12] ISO/IEC 25012:2008. *Software engineering — Software product Quality Requirements and Evaluation (SQuARE) — Data quality model*. International Organization for Standardization.
- [13] Mallemapati, A., & Jaladi, D. S. (2022). An API-Driven Master Data Management Framework for Distributed Enterprise Application Integration. *International Journal of Emerging Research in Engineering and Technology*, 3(2), 151-160.
- [14] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.
- [15] Yuvaraj, N., & Kumar, M. S. (2021). From Governed Data to Customer Health Signals: Integrating Telemetry with Enterprise Data Quality Controls. *International Journal of Emerging Trends in Computer Science and Information Technology*, 2(4), 115-125
- [16] Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., ... & Wyborn, L. (2019). Make scientific data FAIR. *Nature*, *570*(7759), 27–29.
- [17] Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3), 1–52.
- [18] Loshin, D. (2011). *The Practitioner's Guide to Data Quality Improvement*. Morgan Kaufmann.
- [19] Kumar, M. S., & Yuvaraj, N. (2020). Building a Privacy-Aware Customer Data Foundation: A Governance-First Approach to Digital Service Systems. *International Journal of Emerging Research in Engineering and Technology*, 1(4), 55-68.
- [20] Redman, T. C. (2008). *Data Driven: Profiting from Your Most Important Business Asset*. Harvard Business Press.
- [21] Lee, Y. W., Pipino, L. L., Funk, J. D., & Wang, R. Y. (2006). *Journey to Data Quality*. MIT Press.
- [22] Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data Quality and Record Linkage Techniques*. Springer.