Alice Dixon, 2024, 12:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

An Integrated Framework for Data Quality, Completeness, and Trustworthiness

Alice Dixon[1]
Research Fellow in Trustworthy Data Systems[1]

Grace Knight[2]
Senior Data Integrity Consultant[2]

Rebecca Walsh[3]
Data Quality Research Specialist[3]

Chaitanya Srinivas[4]
Senior Java Software Developer[4]

Niharika[5]
Lead Engineer[5]

Abstract- In the rapidly evolving digital landscape, organizations increasingly rely on data-driven decision-making to improve operational efficiency, strategic planning, and business performance. However, issues related to data quality, completeness, consistency, and trustworthiness continue to pose significant challenges across enterprise information systems. Inaccurate, incomplete, outdated, or inconsistent data can compromise analytical outcomes, reduce stakeholder confidence, and lead to ineffective decisions. This research proposes an integrated framework for assessing and improving data quality, completeness, and trustworthiness within modern data management environments. The framework combines data profiling, automated validation, anomaly detection, data quality metrics, metadata management, and data governance practices to identify and address potential data deficiencies. It evaluates key dimensions, including accuracy, completeness, consistency, validity, uniqueness, timeliness, and integrity, to provide a comprehensive assessment of data reliability. Furthermore, the framework emphasizes data lineage, provenance tracking, transparency, and accountability to establish confidence in data sources and analytical outputs. Automated monitoring and rule-based validation can support continuous quality assessment, while human oversight enables the verification of critical findings and the correction of complex data issues. The proposed approach is designed to support enterprise databases, data warehouses, cloud platforms, and analytical systems by integrating quality assessment with established governance processes. Its effectiveness can be evaluated through measurable indicators such as error rates, completeness scores, validation accuracy, and the time required to detect and resolve data quality issues. Overall, the framework provides a structured foundation for strengthening data reliability, improving analytical confidence, and promoting trustworthy data management practices across organizational environments.

Keywords— Data Quality, Data Completeness, Data Trustworthiness, Data Quality Management, Data Quality Assessment, Data Quality Assurance, Data Quality Framework, Integrated Data Quality Framework, Data Governance, Data Governance Framework, Data Integrity, Data Accuracy, Data Consistency, Data Validity, Data Uniqueness, Data Timeliness, Data Reliability, Data Availability, Data Accessibility, Data Usability, Data Relevance, Data Conformity, Data Precision, Data Correctness, Data Consistency Validation, Data Completeness Assessment, Data Accuracy Measurement, Data Quality Metrics, Data Quality Dimensions, Data Quality Indicators, Data Quality Rules, Data Quality Standards, Data Quality Monitoring

© 2024 Alice Dixon. his is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited.

I. INTRODUCTION

Data has become a fundamental resource for modern organizations, research institutions, government agencies, and technology-driven industries. The widespread adoption of digital technologies, cloud computing, Internet of Things (IoT) devices, and artificial intelligence has resulted in the generation of large volumes of data from diverse sources. These data are used for decision-making, predictive analytics, operational management, scientific discovery, and automated information processing. However, the usefulness of data depends not only on its availability but also on its quality, completeness, and trustworthiness. Poor-quality data can produce misleading analytical results, while incomplete or unreliable information can reduce the effectiveness of decision-making processes.

Data quality refers to the degree to which data satisfies defined requirements and is suitable for its intended purpose. It encompasses several dimensions, including accuracy, consistency, validity, uniqueness, and timeliness. Data completeness focuses on whether the required attributes, records, and information are present within a dataset. Data trustworthiness concerns the extent to which users can rely on data based on its source, reliability, integrity, transparency, and traceability. Although these characteristics are closely related, they represent different aspects of data reliability and should be evaluated together to obtain a comprehensive assessment.

In many existing data management systems, quality checks concentrate on specific problems such as missing values, duplicate records, invalid entries, or inconsistent formats. These approaches can identify individual defects but may not adequately evaluate the overall reliability of a dataset. For example, a dataset may contain all required fields but still include inaccurate values. Similarly, a dataset may be accurate and consistent while lacking sufficient information about its origin or processing history. Such limitations demonstrate the need for a unified framework that considers multiple data characteristics simultaneously.

This research focuses on developing an integrated framework for evaluating data quality, completeness, and trustworthiness. The framework combines measurement techniques, validation procedures, provenance analysis, and continuous monitoring to support a systematic assessment of data reliability. By examining these dimensions collectively, the proposed approach aims to improve data management practices, reduce uncertainty, and support more dependable analytical outcomes. The framework is intended to provide a foundation for evaluating datasets across different domains and information systems.

1. Background of the Study

The increasing dependence on data-driven technologies has made data quality an important concern in information management. Organizations collect information from databases, online transactions, sensors, enterprise applications, external data providers, and distributed computing systems. These sources frequently differ in structure, format, collection methods, and reliability. As a result, combining information from multiple sources can introduce inconsistencies, missing values, duplication, and uncertainty.

Traditional data quality methods often rely on predefined validation rules, database constraints, and manual inspection. Although these methods remain useful, they may not be sufficient for complex datasets that change continuously or contain information from heterogeneous sources. Furthermore, evaluating completeness and trustworthiness separately from conventional quality dimensions may overlook important relationships between data defects and the reliability of the resulting information.

An integrated approach is therefore needed to connect data quality measurements with completeness analysis and trust evaluation. Such an approach can help organizations identify weaknesses, understand the origin of data problems, and establish appropriate quality improvement procedures.

II. PROBLEM STATEMENT

Existing data management approaches frequently assess accuracy, consistency, completeness, and trustworthiness using separate methods or isolated validation procedures. This separation can result in fragmented quality information and make it difficult to determine whether a dataset is suitable for a particular application. A dataset may satisfy structural validation rules while containing missing values, outdated information, unreliable sources, or insufficient provenance records.

Another challenge is the absence of a consistent assessment process that combines measurable data quality indicators with evidence about data origin, transformation, and reliability. Without a unified approach, organizations may find it difficult to compare datasets, prioritize data quality improvements, and determine the level of confidence that should be placed in analytical results.

The central problem addressed by this research is the need for an integrated framework that systematically evaluates data quality, completeness, and trustworthiness while supporting consistent measurement, validation, and monitoring across different data environments.

3. Research Objectives

The primary objective of this research is to develop an integrated framework for assessing data quality, completeness, and trustworthiness. The specific objectives are to identify the major dimensions affecting data reliability, define suitable indicators for evaluating each dimension, establish a unified assessment process, and investigate how the combined evaluation can support improved data management.

The study also aims to examine the relationships among missing information, data errors, source reliability, and data provenance. Another objective is to establish a systematic procedure for identifying data defects, measuring their severity, and recommending corrective actions. Finally, the research seeks to define an evaluation approach through which the proposed framework can be assessed for consistency, usability, scalability, and effectiveness.

4. Research Questions

The research is guided by the following questions:

- How can data quality, completeness, and trustworthiness be evaluated within a single framework?
- Which metrics are appropriate for measuring the different dimensions of data reliability?
- How does missing or inaccurate information influence the overall trustworthiness of a dataset?
- How can data provenance and source reliability strengthen data trust assessment?
- How can continuous monitoring and automated validation improve the long-term reliability of data?

These questions provide direction for the framework design, methodological development, and evaluation process.

5. Scope of the Study

The study focuses on the assessment and management of structured and semi-structured data in contemporary information systems. It considers data quality dimensions such as accuracy, consistency, validity, uniqueness, and timeliness, together with completeness indicators and trust-related characteristics such as provenance, traceability, and source reliability.

The framework can be considered for applications involving databases, enterprise information systems, cloud platforms, data integration pipelines, big data analytics, and AI-related datasets. Its implementation may be adapted to different domains according to their data requirements, regulatory constraints, and operational needs. The study does not assume that one universal set of metrics will be appropriate for every dataset; instead, the assessment should reflect the intended use and context of the data.

III. LITERATURE REVIEW

Data quality has been widely examined in database management, information systems, data integration, and analytical processing. Previous approaches have considered dimensions such as accuracy, completeness, consistency, timeliness, validity, and uniqueness. These dimensions provide a foundation for evaluating whether data meets organizational or application-specific requirements. Data profiling, integrity constraints, statistical analysis, and rule-based validation are commonly used to detect defects and measure quality.

Research on data completeness has emphasized the identification of missing values, absent records, incomplete attributes, and gaps in dataset coverage. Missing information may arise from collection errors, sensor failures, incomplete user submissions, system integration problems, or restrictions on data access. Depending on the application, missing values can introduce bias, reduce statistical power, or limit the applicability of analytical models. Completeness assessment therefore requires more than simply counting empty fields; it may also involve determining whether the available data adequately represents the required population or phenomenon. Data trustworthiness is related to the reliability and credibility of information. Factors such as source reputation, provenance, data lineage, processing history, and integrity controls can influence whether users consider data dependable. Provenance records provide information about where data originated and how it was transformed. Such records can support auditing, reproducibility, accountability, and the investigation of data defects.

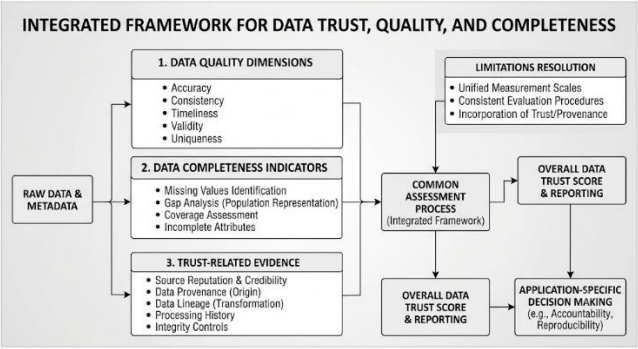
Despite progress in these areas, a key research opportunity is to connect the measurement of quality and completeness with the evidence needed to establish trust. An integrated framework can address this opportunity by bringing together established quality dimensions, completeness indicators, and trust-related evidence in a common assessment process. A rigorous literature review for the final paper should compare relevant studies, identify their limitations, and support the framework design with verified publications from the intended research period.

1. Limitations of Existing Approaches

Many traditional data quality approaches focus on detecting specific errors through predefined rules. For instance, database constraints can identify invalid values, while profiling methods can reveal missing fields and unexpected distributions. However, passing these checks does not necessarily establish that data is trustworthy or appropriate for a particular purpose.

A second limitation is that different quality dimensions may be measured using inconsistent scales or evaluation procedures. This makes it difficult to compare datasets or combine the results into a single assessment. Furthermore, data provenance and source reliability are not always included in conventional quality scoring methods, even though they may be important in applications that require accountability or reproducibility.

The proposed framework addresses these limitations by organizing the evaluation into related components and establishing a consistent process for measuring, interpreting, and reporting the results.



IV. CONCEPTUAL FOUNDATIONS

1. Data Quality

Data quality represents the extent to which a dataset satisfies specified requirements and supports its intended use. Accuracy measures how closely data values correspond to the relevant real-world facts or accepted reference values. Consistency examines

whether values agree across records, systems, and business rules. Validity determines whether values conform to defined formats, ranges, and constraints. Uniqueness assesses whether duplicate records or entities are present, while timeliness measures whether data is sufficiently current for the task being performed.

These dimensions are related but not interchangeable. A dataset can be valid according to its schema while containing inaccurate values. Similarly, a dataset may be accurate when collected but become outdated as the underlying conditions change. Effective quality assessment must therefore consider the dimensions relevant to the intended application instead of relying on a single validation test.

2. Data Completeness

Data completeness measures whether the necessary information is available for the intended task. It can be assessed at the attribute, record, dataset, or population level. Attribute completeness examines the presence of required fields, while record completeness evaluates whether individual records contain the necessary information. Dataset completeness considers the overall availability of required records, and population completeness examines whether the dataset adequately covers the relevant entities or observations.

A basic completeness measure can be defined as:

$$C = \frac{N_{\text{present}}}{N_{\text{required}}} \times 100$$

where N_{present}represents the number of required data elements that are present and N_{required}represents the total number of required data elements.

This measure is useful when all required elements are equally important and the expected set of elements is known. More advanced applications may require weighted completeness measures, rules for optional fields, or coverage measures that account for missing entities. Completeness should therefore be interpreted in relation to the purpose and requirements of the dataset.

3. Data Trustworthiness

Data trustworthiness describes the extent to which data can be relied upon for a specific decision or application. It may depend on the quality of the information, the credibility of its source, the integrity of its processing history, and the availability of evidence supporting its origin. Trustworthiness also involves transparency and traceability, which allow users to investigate how data was collected, transformed, and maintained.

Trustworthiness cannot be established solely by the absence of missing values or by a high quality score. For example, a complete dataset from an unknown or unreliable source may still require additional verification. Likewise, data with strong provenance records may remain unsuitable if it contains serious accuracy or consistency problems. A comprehensive trust assessment should therefore combine direct quality measurements with evidence about data origin, handling, and verification.

4. Relationship Between Quality, Completeness, and Trustworthiness

Data quality, completeness, and trustworthiness are distinct but interdependent concepts. Completeness contributes to quality by indicating whether required information is present.

Quality contributes to trustworthiness by providing evidence about accuracy, consistency, validity, and timeliness. Provenance and traceability strengthen trustworthiness by making data history and transformations more transparent.

The relationship is not always linear. A dataset with high completeness may still be inaccurate, while a dataset with some missing values may remain suitable for a specific analysis.

Similarly, a trustworthy source does not guarantee that every record is correct. The integrated framework must therefore preserve the individual measurements of each dimension while providing an overall interpretation of their combined significance.

IV. PROPOSED INTEGRATED FRAMEWORK

1. Framework Overview

The proposed framework organizes data evaluation into a sequence of connected stages: data acquisition, profiling, quality assessment, completeness analysis, trustworthiness evaluation, integrated scoring, corrective action, and continuous monitoring. Each stage produces information that supports the subsequent stage, allowing data defects and reliability concerns to be identified systematically.

The framework begins by defining the intended use of the data and the requirements that the dataset must satisfy. It then examines the dataset's structure, values, metadata, and available provenance information. Quality and completeness metrics are calculated, and trustworthiness is evaluated using evidence such as source identity, processing history, and validation records. The results are combined into an assessment report that identifies strengths, weaknesses, and recommended actions.

2. Data Acquisition and Profiling

The first stage involves collecting data from relevant sources and recording available information about its origin, format, collection time, and processing history. Data may originate from databases, files, APIs, sensors, enterprise systems, or external providers. Recording source metadata at this stage helps preserve information needed for later traceability and trust evaluation.

Data profiling is then used to understand the dataset's structure and characteristics. Typical activities include examining data types, value distributions, missing-value frequencies, uniqueness, valid ranges, and relationships among fields. Profiling helps establish an initial baseline and identifies areas requiring more detailed analysis. The results can also guide the selection of validation rules and metrics.

3. Quality Assessment Module

The quality assessment module evaluates accuracy, consistency, validity, uniqueness, and timeliness according to the requirements of the target application. Validation rules can identify values outside acceptable ranges, violations of relationships between fields, inconsistent representations, and duplicate records. Where suitable reference data is available, accuracy can be assessed by comparing observed values with verified values.

Each quality dimension should be measured independently before an overall quality score is calculated. This makes the assessment more transparent and helps identify the specific reasons for a low score. For example, a dataset with high consistency but low timeliness requires a different corrective action from a dataset with high timeliness but poor accuracy.

4. Completeness Assessment Module

The completeness module identifies missing values, absent records, incomplete attributes, and potential gaps in coverage. The assessment begins by defining which elements are required, optional, or conditionally required. This distinction prevents legitimate empty values from being incorrectly classified as quality defects.

The module calculates completeness indicators at appropriate levels, such as individual attributes, records, or the entire dataset. It can also identify patterns of missingness and determine whether missing data is concentrated in specific sources, time periods, or groups. Where appropriate, the system can recommend data collection, source reconciliation, or imputation techniques. Imputed values should be clearly distinguished from observed values to preserve transparency and prevent artificial completeness from being mistaken for verified information.

5. Trustworthiness Assessment Module

The trustworthiness module evaluates the evidence supporting confidence in the data. It examines source identity, provenance availability, lineage completeness, validation history, and the reliability of the processes through which the data was obtained and transformed. Depending on the application, it may also consider source verification,

audit trails, access controls, and documented data ownership.

Trustworthiness assessment should distinguish between verified evidence and assumptions. For example, a documented source does not automatically mean that every value is correct, and the presence of a lineage record does not prove that every transformation was appropriate. The module should report the evidence available, the checks performed, and any remaining uncertainty so that users can make informed decisions.

6. Integrated Assessment and Reporting

The integration stage brings together the results of the quality, completeness, and trustworthiness modules. Rather than treating the combined score as a substitute for detailed measurements, the framework retains individual dimension scores and uses the overall result as a summary indicator.

A simple weighted model may be defined as:

$$S = w_q Q + w_c C + w_t T$$

where:

- $S$is the overall assessment score.
- $Q$is the normalized data quality score.
- $C$is the normalized completeness score.
- $T$is the normalized trustworthiness score.
- w_q, w_c, and $w_t$are the corresponding weights, with $w_q + w_c + w_t = 1$.

The scores should be normalized to a common scale, such as 0–100, before combination. The weights should be justified by the application requirements, expert judgment, or empirical evaluation rather than selected arbitrarily. In high-risk applications, a weighted average alone may be misleading because a very low score in one critical dimension can be concealed by high scores in other dimensions. Such cases may require minimum thresholds, mandatory validation rules, or explicit risk flags.

The final report presents the overall score, dimension-level results, detected defects, provenance evidence, and recommended corrective actions. This provides users with a clearer understanding of both the strengths and limitations of the dataset.

V. RESEARCH METHODOLOGY

1. Research Design

The study can adopt a framework-development and evaluation methodology. The first phase identifies the relevant dimensions of data quality, completeness, and trustworthiness through a structured review of existing research and established data management practices. The second phase defines the framework components, metrics, and evaluation rules. The third phase implements the framework using a suitable dataset or prototype, and the final phase evaluates its performance against defined criteria.

This design supports the development of a transparent and reproducible assessment method. The methodology should specify the datasets, preprocessing procedures, quality rules, scoring techniques, and evaluation criteria used in the study. These details are necessary for other researchers to reproduce the results and assess the applicability of the framework.

2. Data Collection and Preparation

The framework may be evaluated using real-world datasets, publicly available benchmark datasets, or controlled datasets containing known quality defects. The selected data should reflect the types of problems the framework is intended to address, such as missing values, duplicates, invalid entries, inconsistencies, or incomplete provenance information.

During preparation, the data is examined for structural errors, incompatible formats, missing attributes, and potential anomalies. Any preprocessing should be documented so that it does not conceal the original quality problems. If artificial defects are introduced for testing purposes, the defect types, locations, and rates should be recorded to establish a reference against which detection performance can be measured.

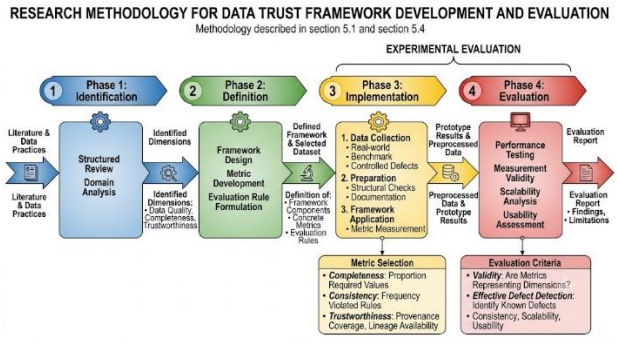
3. Metric Selection

Metrics should be selected according to the framework's objectives and the characteristics of the target dataset. Completeness can be measured through the proportion of required values present. Consistency can be measured through the frequency of violated rules, while validity can be assessed through the proportion of values satisfying specified constraints. Accuracy requires reference values, verified labels, or another defensible source of comparison.

Trustworthiness metrics may include provenance coverage, source verification rate, lineage availability, and the proportion of records with sufficient supporting evidence. These measures should not be treated as universal substitutes for trust. Instead, they provide observable indicators that contribute to a context-specific evaluation.

4. Evaluation Criteria

The proposed framework can be evaluated using several criteria. Measurement validity examines whether the selected metrics represent the intended dimensions. Consistency assesses whether the framework produces stable results when the same data and rules are used. Detection effectiveness measures the ability to identify known defects. Scalability examines performance as dataset size increases, while usability evaluates whether the assessment reports are understandable and actionable.



Where a baseline method is available, the integrated framework can be compared with separate quality checks or existing data profiling approaches. The evaluation should report the experimental setup, assumptions, observed results, and limitations. Claims of improved performance should be supported by measured results rather than inferred from the framework design alone.

VI. IMPLEMENTATION STRATEGY

1. System Architecture

A practical implementation can be organized into modular components for data ingestion, metadata management, profiling, validation, completeness analysis, trust evaluation, score calculation, and reporting. This modular structure allows individual components to be modified or replaced without redesigning the entire system.

The ingestion component accepts datasets and associated metadata. The profiling component extracts structural and statistical information, while the validation component applies defined rules. The completeness component identifies missing information, and the trust component examines provenance and source evidence. Finally, the reporting component combines the results and presents detailed findings to users.

2. Automated Data Quality Rules

Automated rules can be used to detect common data problems. Examples include checking whether dates follow an accepted format, whether numerical values fall within valid ranges, whether required fields are present, and whether unique identifiers are duplicated. Cross-field rules can verify relationships such as whether a transaction date falls within an acceptable period or whether a total matches the sum of its component values.

Rules should be documented and version-controlled because data requirements may change over time. The framework should also distinguish between errors that require correction and warnings that need human review. This prevents potentially valid but unusual data from being automatically rejected.

3. Data Quality Dashboard and Reporting

A dashboard can summarize the assessment results through dimension-level scores, missing-value rates, defect counts, provenance coverage, and changes over time. Users should be able to identify the source

of each issue and examine the rules responsible for flagging it.

Reporting should prioritize transparency rather than presenting only a single overall score. A high-level indicator can support comparisons, but detailed metrics are necessary for diagnosing problems and selecting corrective actions. Reports should also communicate limitations, unresolved issues, and the time at which the assessment was performed.

VII. RESULTS AND DISCUSSION

This section should be completed using the actual results obtained from implementing and evaluating the proposed framework. If the research is still at the proposal stage, it should be presented as an evaluation plan rather than as a report of completed experiments.

1. Data Quality Assessment Results

The results should describe the accuracy, consistency, validity, uniqueness, and timeliness measurements obtained from the selected datasets. The analysis should identify which dimensions produced the greatest number of detected defects and whether the problems were concentrated in particular attributes or sources.

The discussion should explain the practical implications of the findings. For example, frequent violations of consistency rules may indicate differences in data entry standards or integration procedures. Low timeliness may indicate that data refresh schedules are unsuitable for the intended use. Each interpretation should be supported by observed evidence from the evaluation.

2. Completeness Assessment Results

The completeness results should report the proportion of required values and records that were available before and, where applicable, after corrective actions. The analysis should distinguish missing values from intentionally empty or inapplicable fields. It should also identify whether incompleteness was concentrated in particular variables, collection periods, or source systems.

Where imputation or data recovery is applied, the evaluation should determine whether these procedures improve usability without introducing unacceptable bias or uncertainty. Missing values should not be considered successfully resolved merely because they have been replaced with estimated values; the reliability of the replacement method must also be examined.

3. Trustworthiness Assessment Results

The trustworthiness analysis should examine the availability and quality of provenance records, source verification evidence, processing histories, and validation records. It should explain how these indicators influence confidence in the dataset and identify cases in which insufficient evidence prevents a strong trustworthiness assessment.

The discussion should emphasize that trustworthiness is not necessarily a directly observable property that can be established through one universal numerical measure. Instead, the framework provides a structured evaluation of the evidence supporting the use of data for a defined purpose.

4. Comparative Analysis

If the framework is compared with existing methods, the results should show whether integration improves defect detection, assessment coverage, reporting consistency, or operational efficiency. The comparison should use the same datasets and clearly defined evaluation criteria wherever possible. A meaningful comparison should also consider implementation cost, processing time, maintenance requirements, and interpretability. An integrated framework may provide more comprehensive reporting but require additional metadata, more validation rules, or greater computational resources. These trade-offs should be discussed objectively.

Challenges and Limitations

The development of an integrated framework presents several challenges. First, data quality requirements vary across applications, making it difficult to establish universal metrics or scoring thresholds. A dataset suitable for exploratory research may not meet the stricter requirements of financial reporting or clinical decision-making.

Second, completeness depends on knowing which data elements are required. Without an appropriate reference schema, expected population, or collection specification, missing information may be difficult to distinguish from information that was never required. Third, trustworthiness often depends on evidence that may not be available, including source verification records, collection methods, and transformation histories.

Another limitation concerns the combination of different metrics into an overall score. Weighting methods can introduce subjectivity, and aggregate scores may hide critical weaknesses in individual dimensions. The framework must therefore retain dimension-level results and apply mandatory thresholds when necessary. Finally, implementation across heterogeneous data systems may require integration with existing databases, metadata repositories, governance policies, and monitoring tools.

These limitations suggest that the framework should be adaptable, transparent, and evaluated within clearly defined application contexts.

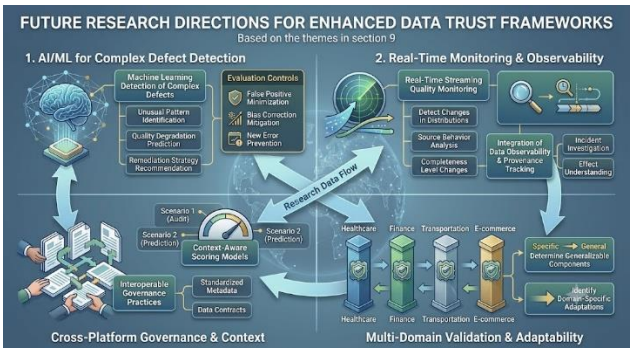
Future Research Directions

Future research can investigate the use of machine learning to detect complex data quality defects that are difficult to identify through fixed validation rules. Such techniques may help identify unusual patterns, predict quality degradation, or recommend appropriate remediation strategies. However, automated detection should be evaluated carefully to avoid false positives, biased corrections, and the introduction of new errors.

Another direction involves real-time quality monitoring in streaming and distributed data environments. Continuous monitoring could identify changes in data distributions, source behavior, or completeness levels before these problems significantly affect downstream applications. The integration of data observability techniques with provenance tracking may also improve the ability to investigate data incidents and understand their effects.

Further studies may explore context-aware scoring models that adjust evaluation criteria according to the intended use of the data. Research could also investigate the role of standardized metadata, data contracts, and interoperable governance practices in improving cross-platform quality assessment. Finally, the framework should be tested across multiple domains to determine which components can be generalized and which require domain-specific adaptation.

Future research should also focus on developing standardized evaluation metrics to compare the performance of different data quality assessment frameworks. The scalability of the proposed framework can be examined using large-scale datasets and distributed computing environments. In addition, explainable artificial intelligence techniques may help users understand why particular datasets receive specific quality and trustworthiness scores. Further investigation into privacy-preserving data quality assessment could enable organizations to evaluate sensitive data without exposing confidential information. Longitudinal studies may also be conducted to assess how the framework performs over time and how continuous improvements in data governance influence overall data reliability and trust.



VIII. CONCLUSION

Data quality, completeness, and trustworthiness are closely connected characteristics that influence the reliability and usefulness of information in modern data-driven environments. Assessing these characteristics independently can provide valuable information, but it may not fully capture the

combined factors that determine whether a dataset is suitable for a particular purpose. An integrated approach offers a structured way to connect quality measurements, completeness indicators, and evidence about data origin and processing history.

This paper has outlined a framework that combines data profiling, quality assessment, completeness analysis, trustworthiness evaluation, integrated reporting, and continuous monitoring. It also described a methodology for selecting metrics, implementing the framework, and evaluating its effectiveness. By retaining individual dimension scores alongside an overall assessment, the framework can provide a more transparent view of data reliability and help users identify appropriate corrective actions.

The proposed framework provides a basis for further implementation and empirical validation. Its effectiveness should ultimately be established through experiments involving representative datasets, clearly defined evaluation criteria, and comparisons with relevant existing methods. Future work can extend the approach through automation, real-time monitoring, and context-aware assessment to support more dependable data management and decision-making.

REFERENCES

1. Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management Science, 31*(2), 150–162.
2. Reddy BasiReddy, S. (2016). Java-centric workflow orchestration for enhancing telecom service provisioning and CRM operations. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 1(3), 111–119.
3. Vollem, S. (2024). From deterministic pipelines to intelligent orchestration: A transformer-driven framework for LLM-augmented DevOps automation. International Journal of Research Publications in Engineering, Technology and Management, 7(1), 9964–9975.
4. Thota, M. R. (2016). Resilient data engineering: The evolution of database and big data administration in cloud native platforms. European Journal of Advances in Engineering and Technology, 3(12), 63–69.
5. Nagender, Y. (2018). Reimagining master data management as a foundational enterprise capability across business domains. International Journal of Science, Engineering and Technology, 6(2).
6. Fox, C., Levitin, A., & Redman, T. (1994). The notion of data and its quality dimensions. *Information Processing & Management, 30*(1), 9–19.
7. Seetala, S. R. (2016). Architectural evolution in enterprise data modeling: From dimensional leadership to hybrid integration frameworks. International Journal of Technology, Management and Humanities, 2(1), 52–66.
8. Nanchari, N. (2020). Remote Patient Monitoring in Healthcare: Leveraging Iot for Continuous Care. In International Journal of Science, Engineering and Technology (Vol. 8, Number 4). Zenodo.
9. Parepalli, S. (2016). Cloud aligned ETL framework architectures for enterprise data modernization at scale. International Journal of Technology, Management and Humanities, 2(1), 36–51.
10. Ghanta, S. (2016). Designing high-reliability enterprise Java systems through modular architecture and resilience patterns. International Journal of Scientific Research in Science and Technology, 2(1), 291–306.
11. Boddupally, H. L. (2019). Designing end-to-end observability architectures for high-reliability .NET cloud applications in production environments. International Journal of Scientific Research & Engineering Trends, 5(6).
12. Vankayala, S. C. (2019). Predictive defect governance and decision optimization in mortgage underwriting platforms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(1), 382–398.
13. 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.

14. Menda, J. R. (2018). A hybrid log-driven and event-time streaming pipeline: Integrating Kafka Streams with Apache Flink for real-time financial transaction processing. Journal of Scientific and Engineering Research, 5(1), 284–292.
15. Teegala, R. (2017). Architectural design and governance of multi-tenant microservices for regulated financial institutions. International Journal of Scientific Research in Science, Engineering and Technology, 3(5), 895–913.
16. BasiReddy, S. R. (2017). Data hygiene and batch optimization in enterprise CRM: A 2017 framework for scalable, high-quality customer data integration. Journal of Scientific and Engineering Research, 4(11), 272–280.
17. Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM, 40*(5), 103–110.
18. Nagender, Y. (2020). Leading the end-to-end modernization of enterprise master data platforms using TIBCO EBX within Elavon's core data ecosystem. European Journal of Advances in Engineering and Technology, 7(1), 82–94.
19. Vollem, S. (2022). Architecting high-throughput transaction processing in distributed microservices systems: Principles, coordination mechanisms, and performance optimization. International Journal of Scientific Research & Engineering Trends, 8(3).
20. Seetala, S. R. (2017). Architecting trust in enterprise data warehouses: A structured framework for profiling, validation, and lifecycle quality management. Journal of Scientific and Engineering Research, 4(1), 193–203.
21. Nanchari, N. (2021). IoT and chronic disease management. International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), 8(1), 378–381.
22. Thota, M. R. (2018). Transforming database leadership in the era of cloud-native automation and resilient operations. International Journal of Technology, Management and Humanities, 4(2), 25–43.
23. Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering, 7*(4), 623–640.
24. Boddupally, H. L. (2020). Model driven engineering of robust data pipelines: Leveraging Entity Framework constructs with SQL Server execution layers. European Journal of Advances in Engineering and Technology, 7(2), 83–94.
25. Vankayala, S. C., & Devaraju, S. (2021). Cloud-native banking quality assurance frameworks powered by Azure DevOps and intelligent observability. ESP Journal of Engineering & Technology Advancements, 1(2), 305–310.
26. Ghanta, S. (2017). Operationalizing event-driven architecture in enterprise Java systems using Spring Cloud Stream. Journal of Scientific and Engineering Research, 4(2), 164–171.
27. Parepalli, S. (2016). Data hygiene and batch optimization in enterprise CRM: A framework for scalable, high-quality customer data integration. Journal of Scientific and Engineering Research, 3(5), 285–292.
28. Menda, J. R. (2019). Engineering secure financial microservices through end-to-end encryption, zero trust API governance, and multi-layered cybersecurity controls. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(2), 1389–1405.
29. Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218.
30. BasiReddy, S. R. (2018). Modernizing CRM data pipelines through parallel processing and cloud-native orchestration. International Journal of Scientific Research & Engineering Trends, 4(2). Zenodo.
31. Teegala, R. (2018). Cloud-native transaction platforms in financial systems: Architecture, resilience, and regulatory alignment. International Journal of Science, Engineering and Technology, 6(1).
32. Nagender, Y. (2021). Implementing high-performance data integration pipelines for analytics and reporting in complex enterprise landscapes. International Journal of Scientific Research & Engineering Trends, 7(5).
33. Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information & Management, 40*(2), 133–146.

34. Nanchari, N. (2021). IoT in emergency medical services (EMS). International Journal of Science, Engineering and Technology, 9(4).
35. Seetala, S. R. (2017). Advancing enterprise data governance and data quality management through comprehensive metadata-centric frameworks for modern data ecosystems. International Journal of Technology, Management and Humanities, 3(3), 18–34.
36. Vollem, S. (2022). Streaming-first enterprise decision systems: Architectural evolution from batch dataflows to stateful, exactly-once real-time processing. International Journal of Engineering & Extended Technologies Research (IJEETR), 4(1), 4326–4335.
37. Ghanta S. Quality-Driven Microservice Refactoring of Legacy Java Systems: Patterns, Automation, and Migration Challenges. J Artif Intell Mach Learn & Data Sci 2018 1(1), 3197-3202. DOI:
38. Vankayala, S. C. (2022). Tail latency oriented quality assurance for microservices: A system aware, SLO driven approach. International Journal of Science, Engineering and Technology, 10(5).
39. Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM, 45*(4), 184–192.
40. Boddupally, H. L. (2022). Architectural-driven intelligent refactoring for resilient cloud-native .NET systems. European Journal of Advances in Engineering and Technology, 9(1), 95–104.
41. Menda, J. R. (2019). A distributed identity orchestration framework for secure authentication automation leveraging Keycloak, OAuth 2.0 grant types, and adaptive access policies. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(4), 364–381.
42. Thota, M. R. (2020). Architecting secure and compliant hybrid cloud database systems: Frameworks, cryptography, and big data platforms. International Journal of Scientific Research & Engineering Trends, 6(5). Zenodo.
43. Parepalli, S. (2017). Evolving enterprise reconciliation: From deterministic validation to AI-supported high-integrity data assurance. Journal of Scientific and Engineering Research, 4(6), 242–252.
44. BasiReddy, S. R. (2019). Designing cloud-native CRM platforms for next-generation telecom operations. European Journal of Advances in Engineering and Technology, 6(3), 130–138.
45. Buneman, P., Khanna, S., & Tan, W.-C. (2001). Why and where: A characterization of data provenance. In J. Van den Bussche and V. Vianu (Eds.), *Database theory—ICDT 2001* (pp. 316–330). Springer.
46. Nagender, Y. (2022). Strengthening enterprise data integrity through intelligent matching and deduplication in EBX. European Journal of Advances in Engineering and Technology, 9(11), 163–177.
47. Teegala, R. (2019). Observability-driven engineering in distributed systems. International Journal of Science, Engineering and Technology, 7(3).
48. Nanchari, N. (2022). Integrating IoT with electronic health records (EHRs). Journal of Scientific and Engineering Research, 9(2), 186–188.
49. Seetala, S. R. (2018). A comprehensive framework for cloud migration of enterprise data warehouses: Architectural transformation, performance optimization, and governance considerations. International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), 4(1), 1861–1878.
50. Vollem, S. (2020). Leveraging infrastructure-as-code automation to establish standardized, reliable, and reproducible cloud infrastructure across modern cloud ecosystems. European Journal of Advances in Engineering and Technology, 7(9), 109–122.
51. Vankayala, S. C. (2022). Intelligent failure prediction in CI/CD pipelines using machine learning models for enterprise quality assurance. International Journal of Scientific Research in Science and Technology, 9(6), 820–832.
52. Ghanta S. SAGA and CQRS Implementation Techniques for Distributed Transaction Management. J Artif Intell Mach Learn & Data Sci 2018 1(1), 3203-3208. DOI:
53. Parepalli, S. (2017). Metadata intelligence for automated data lineage in distributed enterprise

systems. International Journal of Science, Engineering and Technology, 5(5).
54. Menda, J. R. (2020). Designing an intelligent framework for automated governance and enterprise risk management through machine learning-driven signals and predictive analytics. International Journal of Science, Engineering and Technology, 8(6).
55. Teegala, R. (2020). Robust digital foundation architectures supporting enterprise Java and Spring engineering. International Journal of Scientific Research & Engineering Trends, 6(4).
56. Thota, M. R. (2021). From autonomic computing to self-driving databases: AI-driven autonomous operations in cloud environments. International Journal of Research and Applied Innovations.
57. Boddupally, H. L. (2021). Quality forecasting and reliability modeling in expansive .NET application landscapes. European Journal of Advances in Engineering and Technology, 8(1), 157–168.