Journal of Global Research in Multidisciplinary Studies (JGRMS)
Volume 2 | Issue 11 | November 2026 | Pages 01–07
DOI: https://doi.org/xxxxxxxxxxxxx
e-ISSN: 3049-2963
Review Article OPEN ACCESS Available online: https://saanvipublications.com/journals/index.php/jgrms

A Review of Metadata-Driven Data Engineering for Enterprise Mergers and Acquisitions

Dr. Parth Gautam
Assistant Professor
Department of Computer Sciences and Applications
Mandsaur University, Mandsaur
parth.gautam@meu.edu.in

Corresponding Author: * Parth Gautam (parth.gautam@meu.edu.in)

Received: 2 November 2026 | **Revised:** 5 November 2026 | **Accepted:** 15 November 2026

***Abstract*—Metadata-driven data engineering plays an important role in integrating heterogeneous enterprise data systems during mergers and acquisitions. Differences in database schemas, data formats, platforms, governance policies, and security requirements create significant challenges for achieving consistent and reliable information exchange. This paper reviews metadata-driven data engineering approaches for enterprise integration throughout the merger and acquisition lifecycle. It examines descriptive, structural, administrative, technical, and usage metadata, as well as metadata acquisition, cleaning, verification, and maintenance processes. The study covers intelligent, metadata-driven architectures, centralized metadata control, automated data processing, quality management, governance, monitoring etc. It also examines how to use metadata catalogs, data lineage, data provenance, and access control to enhance data discovery, interoperability, transparency, and regulatory compliance. The enterprise integration and data democratization are discussed under the light of modern data architectures such as data lake houses, data mesh, and Data Vault. The existing research is analyzed and compared, revealing shortcomings of automated metadata discovery, heterogeneous schema matching, semantic integration, and continuous synchronization. The paper identifies scalable and secure, adaptive and intelligent, metadata-driven frameworks for enhancing integration efficiency, maintaining data integrity and facilitating informed decision,making during mergers and acquisitions.**

***Keywords*—Metadata-driven data engineering, enterprise data integration, data interoperability, data governance, data Lakehouse.**

I. INTRODUCTION

The significance of metadata is frequently underestimated. For data consumers, metadata is frequently regarded as an essential burden and a needed component of data generation and dataset dissemination. Metadata experts, a limited cohort including IT, geographic, scientific, and archival sectors, has unique interpretations of an optimal metadata description. As of now, many metadata methodologies include diverse degrees of exertion, intricacy, and profitability. Metadata play a vital role in data-informed decision-making due to the fundamental connection between trustworthy data and the dependability of the decision results. They have a significant position throughout the data acquisition process with respect to dataset traits, provenance, and data reliability [1]. Although they mean different things, the words "merger" and "acquisition" are often used interchangeably [2]. During an acquisition, one company buys either the whole or a fraction of another company. A merger is the formal joining of two or more companies into a single entity. A merger refers to a legal transaction in which two or more companies consolidate, leading to the survival of a single corporation as the legal entity. By definition, a merger involves the consolidation of two or more companies into one, whereas an acquisition is a financially strong corporation purchasing a smaller entity. A merger signifies the unification of many businesses to create one or more organizations. Mergers and acquisitions are processes that include takeovers, corporate realignment, or corporate dominance that transform the ownership arrangement of organizations [3].

The management of enterprise data, or EDM, has become a vital component of improving business performance, especially in the data-driven world of today. EDM can help manage data assets in an organization comprehensively. The main focus of EDM is on procedures, guidelines and equipment that can be used to collect data successfully and efficiently, to store, organize and utilize data [4]. Data estates for enterprises are now spread out in data warehouses, data lakes, data lake houses, software-as-a-service apps, integration platforms, operational databases, analytical pipelines and machine learning systems. As a result, metadata is also fragmented: schemas and constraints reside in databases; transformations are logged in code and orchestration logs; business definitions are captured in glossaries and documents; who is responsible is captured in organizational conventions and there are often different views of a sensitivity label across platforms. Therefore, metadata management is not just documentation, but an operational control plane that enables finding, understanding, trusting and controlling data [5][6].

Copyright © 2026 The Author(s).
This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, adaptation, and reproduction in any medium, provided the original author(s) and the source are properly credited.

J. Glob. Res. Multidiscip. Studies, 2 (11):01-07
DOI: https://doi.org/xxxxxxxxxxxx

A. *Structure of the Paper*

This paper is organized as follows: Section II presents the foundations of metadata-driven data engineering, including metadata types, lifecycle, architecture, and governance. Section III discusses metadata-driven data engineering for enterprise mergers and acquisitions, covering architecture, data integration, modern data architectures, and data democratization. Section IV reviews related studies, comparative analysis, and research gaps. Section V concludes the paper and presents future research directions.

II. Intelligent Metadata-Driven Data Engineering

The Intelligent Metadata-Driven Data Engineering Framework is a scalable, declarative, and modular approach that transforms how enterprises create, operate, and manage data pipelines. Primarily, it has a metadata-first philosophy, whereby all pipeline behaviour, ingestion logic, transformation rules, data quality constraints, execution parameters, and target definitions are represented in structured metadata rather than by procedural code. This is to ensure uniformity across data environments and to make heterogeneous data environments portable and automated.

A. *Types of Metadata*

Metadata is sourced from several origins and file formats for storage, including pharmaceutical management, patient condition tracking, geographic information, image control, enterprise administration, and social media profiles.

- **Descriptive Metadata:** This type of metadata provides details about the content of a resource, such as its title, author, subject, and keywords[7]. It helps users locate and understand the resource. It is commonly used in digital libraries, academic databases, and search engines.
- **Structural Metadata:** Structural metadata articulates the internal framework of a resource, including its table of contents, pagination, chapters, and sections, and is often included in eBooks, websites, and multimedia materials. It facilitates users in traversing and interpreting the resource's arrangement.
- **Administrative Metadata:** Administrative metadata provides information related to the generation and administration of a resource, such as its file format, date of its creation, copyright status, and storage location. It serves an essential function in archiving, document administration, and content management frameworks.
- **Technical Metadata:** This information Describes the technical attributes of a resource, including resolution, file size, and used software, which is crucial for resource identification, processing, and preservation in domains such as image/video processing and digital asset management.
- **Use Metadata:** Utilise metadata to gather information about the manner in which a resource is used or employed, such as user ratings, access logs, and usage statistics, so facilitating the measurement of resource impact, notably in web analytics, user feedback, and e-commerce platforms.

B. *Metadata Lifecycle*

The metadata lifecycle generally comprises four primary phases: acquisition, cleansing, validation, and upkeep. Every step is essential for maintaining the accuracy, reliability, and value of metadata during its whole existence. Fig. 1 depicts the metadata lifecycle, outlining the progression of data through several phases and emphasising critical tasks and their significance in preserving metadata quality and uniformity.

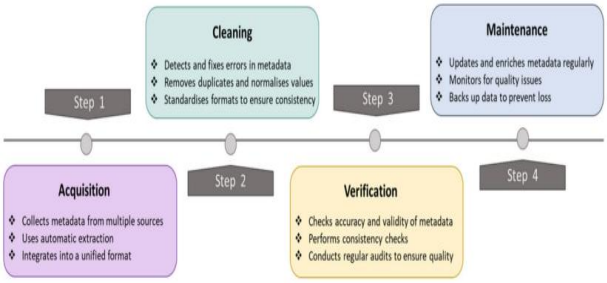

Fig. 1. Metadata Lifecycle Process: Stages and Key Activities

- **Acquisition:** Data acquisition refers to the process of collecting pre-existing data [8]. This stage involves the collection of metadata from several sources, including indexes, abstracts, and bibliographic entries. This data is often generated according to categorization regulations and standards, ensuring uniformity across various datasets.
- **Cleaning:** The process of cleaning metadata include validating the authenticity and accuracy of metadata entries [9]. This procedure includes verifying that metadata fields are completely filled and detecting and rectifying any inaccuracies or inconsistencies..
- **Verification:** The process of metadata verification is concerned with validating the fidelity and entirety of metadata records [10]. This process usually includes executing sample assessments to confirm that the metadata complies with applicable standards and satisfies defined specifications, assuring its trustworthiness.
- **Maintenance:** The quality and continued efficacy of metadata depend on its regular maintenance[11]. This includes protecting metadata against unauthorized use or manipulation, creating backup and recovery processes, and updating metadata information on a regular basis.

C. *Intelligent Metadata-Driven Data Engineering Architecture*

Intelligent Metadata-Driven Data Engineering uses metadata to automate, coordinate, and optimize enterprise data management, integration, and processing activities. The architecture is structured into five logical layers, as shown in Fig. 2. namely, Metadata Configuration Layer, Centralized Metadata Controller, Execution and Processing Layer, Data Quality and Governance Layer and Monitoring and Integration Layer[12]. The layers are weakly coupled and strongly coordinated via metadata-based orchestration, allowing flexibility without compromising control.

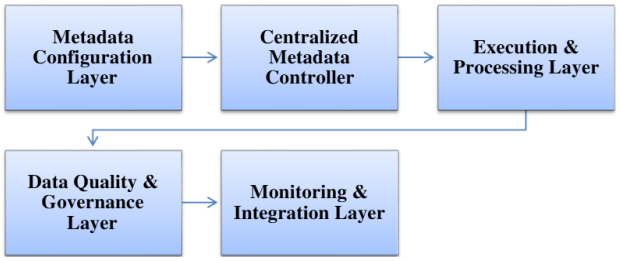

Fig. 2. Intelligent Metadata-Driven Data Engineering Framework

DOI: https://doi.org/xxxxxxxxxxxx

- **Metadata Configuration Layer:** The Metadata Configuration Layer provides all the definitions of pipelines in a standardized, human-readable format. YAML is used because of its readability, trustworthiness in version control and compatibility with CI/CD workflows.
- **Centralized Metadata Controller:** The main system of the structure is the Centralized Metadata Controller, which is the brain of the orchestration.
- **Execution and Processing Layer:** The Execution Layer uses Apache Spark as its distributed processing environment, with its flexibility and fault-tolerance to work with various data formats, such as Parquet, CSV, Excel, ORC, and JSON.
- **Data Quality and Governance Layer:** Metadata-defined rules are applied on the pipeline to ensure data quality, eliminating the need to execute a separate validation process.
- **Automated Exception Handling and Ticketing:** The framework has been integrated with the incident management frameworks, including ServiceNow, to support enterprise-level operations. In cases where data quality breaches or execution failures exceed predetermined limits, incident tickets are automatically created, including the source system, the violated rule and the number of records.
- **Monitoring, Logging, and Observability:** The framework has been integrated with the incident management frameworks, including ServiceNow, to support enterprise-level operations. In cases where data quality breaches or execution failures exceed predetermined limits, incident tickets are automatically created, including the source system, the violated rule and the number of records.
- **CI/CD and Version Control Integration:** The metadata is automatically updated to implement validation, testing and deployment pipelines via CI/CD pipelines with a minimal amount of manual control and configuration error. The strategy supports sound version rollback processes, allowing teams to roll back to the sound versions in case of an issue.

D. Role of Metadata Management in Data Governance

Metadata management is integral to data governance, supporting several critical functions that enhance the value and usability of data assets.

1) Data Discovery and Cataloging

Metadata enables the creation of data catalogs, centralized repositories that index and organize data assets [13]. By associating data with descriptive metadata, such as keywords, classifications, and tags, catalogs empower users to efficiently search for and access relevant data. This capability is particularly vital in large organizations with heterogeneous data environments.

2) Data Lineage and Provenance

Metadata tracks the origin, movement, and transformation of data across systems, providing a comprehensive view of data flows. Data lineage offers a visual representation of these flows, enabling organizations to ensure data integrity, troubleshoot errors, and meet audit requirements. Provenance metadata establishes the authenticity and reliability of data sources, enhancing trust in data assets.

3) Data Quality Management

Metadata includes quality metrics like completeness, accuracy, consistency, and timeliness, crucial for tracking and enhancing data reliability. By embedding these metrics in metadata repositories, organizations can proactively identify and address quality issues, ensuring that data meets stakeholders' needs.

4) Compliance and Access Control

Metadata is essential for implementing security and privacy regulations by specifying access rights and labelling sensitive material, including personally identifiable information (PII). This feature guarantees compliance to rules like GDPR, CCPA, and HIPAA, protecting the organization's standing and reducing legal and academic responsibilities.

III. Metadata-Driven Data Engineering for Mergers and Acquisitions

Metadata-based data engineering for mergers and acquisitions (M&A) is defined as a process that utilizes the use of metadata to discover, structure, integrate and manage enterprise data. This approach allows for data cataloging, schema matching, data provenance, data quality checking, and semantic interoperability between data sources. The method bases its identification of data inconsistencies on technical, business, operational and semantic metadata, enhancing information visibility, aiding due diligence, and helping system integration following a merger or acquisition.

A. Metadata-Driven Architecture

The platform organizes functionality into five layers, each with distinct responsibilities.[14]. This follows established architectural decomposition principles, as shown in Fig. 3. to enable independent evolution of each tier.

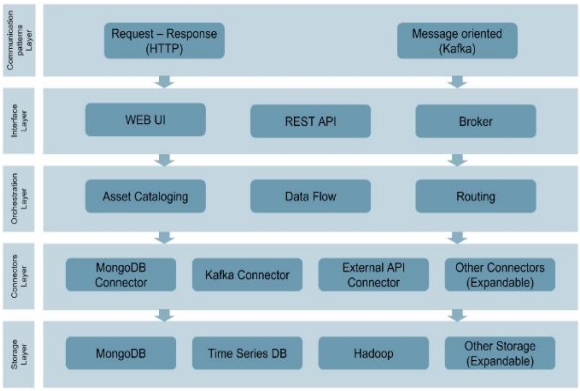

Fig. 3. Five-Layer Platform Architecture

1) Communication Layer

The topmost layer is the Communication Layer which creates two communication paradigms according to different data interaction patterns. Request-response communication using HTTP is suitable for synchronously performing operations such as data retrieval, metadata queries, and configuration management. Asynchronous, high-volume streams of data are processed by using message-oriented communication through Apache Kafka, which enables producers and consumers to be independent from one another. This two-fold solution can support both interactive user operations and continuous IoT data flow.

2) Interface Layer

The interface layer exposes platform capabilities through two mechanisms. A RESTful API provides programmatic

DOI: https://doi.org/xxxxxxxxxxxx

access for external systems, allowing automated integration with analytics tools, visualization platforms, or other services [15]. The endpoints follow REST principles, using standard HTTP methods (GET, POST, DELETE) and returning data in JSON format. A web-based user interface built on this API lets operators manage assets, configure connectors, create data flows, and monitor operations without programming.

3) Asset Cataloging Layer

This layer implements the metadata registry that drives runtime integration: metadata stored here is an executable specification, not merely descriptive documentation. Asset Cataloging maintains a MongoDB database storing metadata for all data sources, connectors, and data flows. The data format rules, connection parameters, and connector identifier are then sent to the user by the registry, which is used to dynamically generate the integration call.

4) Connector Layer

Integration adapters resides in the Connector Layer to facilitate translation between a platform's unified interface and specific storage technologies. Each connector provides an out-of-the-box interface of four operations (retrieval, file retrieval, store, file upload) and hides the technology-specific details of the connection, authentication, and data format handling. The platform dynamically invoke connectors based on metadata specifications, instead of hardcoded logic.

5) Storage Layer

This layer is the bottom layer, which contains heterogeneous data repositories used by pilot sites: MongoDB databases for flexible document storage, Kafka topics for event streaming, Hadoop HDFS for distributed large-scale file storage, and time-series databases for sensor data. The platform also connects to other enterprise systems through RESTful API connectors allowing you to access distributed data sources without changing or replacing infrastructure.

B. Importance of Effective Data Integration



Fig. 4. Assessment Phase

Mergers and acquisitions have become a popular strategy for organizations planning to expand their market share, gain access to new technologies, or achieve economies of scale.. The first thing data engineers need to do after a merger is to thoroughly check out the current data situation [16]. This means, as in Fig. 4. identifying which data sources, formats, and systems both companies use, and understanding the rules and security measures they have in place for handling data.

- **Platforms:** Different companies often use various data management platforms. One might rely on Azure to handle large databases, while another might use GCP.
- **Data Models:** Comparison of similarities and differences of actual data is a crucial step. Data models define how data is structured and related. Comparing data models helps identify similarities and differences and what changes are needed to integrate them.
- **Licenses and Costs:** Data management solutions usually come with licensing fees, which could be based on the number of users or on a subscription model. Comparing these costs helps in deciding which option is the most cost-effective while meeting the merged company's needs.
- **Servers and Applications:** Companies might use different servers and applications. For instance, one may use Microsoft servers and the other Linux systems. Data engineers must integrate these server environments to ensure smooth functionality.
- **Maintenance:** The server and applications need to be maintained constantly. This includes the regular maintenance of software, hardware replacement and data backup. Maintaining things properly helps in keeping everything working smoothly and prevents downtimes.
- **Dependencies and Supporting Vendors:** During the integration process, it's important to identify all dependencies and coordinate with supporting vendors. Dependencies may be software licenses, or third party software or contracts that require management.
- **Resources:** Hardware, software tools, and skilled personnel are essential for managing and integrating data. Their adequate availability supports a smooth and efficient integration process.
- **Security and Compliance:** Keeping data safe and following rules are crucial. This means implementing strong security measures, such as encryption and access controls, to protect sensitive information [17].
- **Business Continuity:** Business continuity is all about making sure an organization can keep running smoothly, even when unexpected disruptions occur during the data transition phase.

C. Modern Data Architectures

Enterprise data management increasingly relies on scalable and flexible approaches to organize, integrate, and analyze data. Among these approaches, the lakehouse combines the advantages of data lakes and data warehouses, while data mesh enables decentralized data ownership and domain-driven management. Data Vault supports effective auditing and historical data management.

- **Data Lakehouse:** The Lakehouse paradigm represents a significant evolution in enterprise data architecture, combining the flexibility of data lakes with the reliability of traditional warehouses [18].
- **Data Mesh:** This architecture has emerged as a transformative approach to enterprise data management, particularly in organizations dealing with complex, distributed data environments.
- **Data Vault:** This approach provides a robust framework for enterprise data warehousing, particularly in scenarios requiring extensive historical tracking and adaptability.

D. Data Democratization in the Modern Enterprise

In modern enterprises, data democratization enables analysts, domain experts, and business users, beyond data

DOI: https://doi.org/xxxxxxxxxxxx

engineers, to leverage data for decision-making, experimentation, and AI adoption [19]. It represents an organizational shift toward making data, analytical insights, and advanced analytics capabilities accessible to a diverse range of business users rather than limiting them to dedicated data specialists.

1) Metadata Catalogs for Discoverability

These platforms serve as searchable directories, allowing users to easily locate relevant datasets, understand their context, and assess their suitability for various analytical tasks. Metadata catalogs typically include descriptions, ownership information, and usage guidelines, making data assets more visible and accessible across the organization.

2) Lineage and Quality Indicators for Trust

Data lineage tools track the origin and transformations of data, helping users understand how it has evolved and whether it can be trusted for specific purposes. Quality indicators, such as completeness, accuracy, and timeliness metrics, further build confidence in data-driven decision-making by providing transparency into the reliability of data assets.

3) Role-Based and Attribute-Based Controls for Secure Access

These security mechanisms ensure that sensitive information remains protected by granting access based on users' roles, responsibilities, or specific attributes. This enables organizations to share data widely while maintaining compliance with privacy and regulatory requirements.

4) No-Code or Low-Code Analytics Tools for Self-Service

By removing the need for advanced programming or technical expertise, these platforms enable business users, analysts, and domain experts to perform complex analyses, visualize data, and generate reports independently. This self-service approach accelerates insights and promotes experimentation throughout the enterprise.

IV. Literature Review

This section reviews related studies on metadata-driven data engineering for enterprise mergers and acquisitions, focusing on intelligent metadata management, data interoperability, post-merger integration, cybersecurity, digital transformation, and workflow optimization to improve data consistency, metadata quality, integration efficiency, and informed decision-making throughout the M&A lifecycle.

J. Bobba et.al. (2026) suggested an AI-enabled Smart Metadata-Driven Enterprise Data Engineering (AIMS-MDE) framework that was comprised of 5 components: semantic metadata intelligence, adaptive metadata graph construction, AI-based workflow recommendation, predictive workflow optimization, autonomous orchestration and continuous metadata knowledge refinement. This proposed framework aims to address the above limitations by transforming enterprise metadata into "smart" metadata through knowledge representation and dynamically generating adaptable and efficient workflows for execution. In enterprise workflow applications, workflow efficiency, metadata quality, recommendation accuracy and resource use efficiency were greatly enhanced during the experimental analysis. The proposed framework achieved a 99.3% accuracy level in automation, 99.1% accuracy level in decision making, 99.0% accuracy level in recommendations, 98.8% accuracy level in workflow and an overall accuracy of 99.2%, which was better than the conventional metadata engineering approaches [20].

R. Zhao, Z. Yan, and L. Zhao (2026) examine the impact of digital transformation on green M&A among Chinese A-share listed enterprises from 2010 to 2023, with an in-depth analysis of the moderating role of CEO career concerns in this process. The results show that digital transformation significantly promotes green M&A, particularly when CEOs face external and prospective concerns, whereas internal career concerns weaken this effect. Heterogeneity analysis reveals that the digital–green relationship varies by CEO characteristics, industry type, and enterprise characteristics [21].

B. Babajanov et.al. (2025) investigate a metadata-driven data interoperability approach (MDDDI) to solving technical, organizational, ethical, and regulatory obstacles to data interoperability, revealing that existing solutions do not fully address these issues. By defining key metadata types and outlining interactions among actors, the domain-agnostic MDDDI reference schema, demonstrated with a synthetic smart greenhouse example, helps to bridge isolated systems by embedding all necessary integration details in the metadata [22].

J. Liu et.al. (2024) contribute to the current literature by providing data on theory–topic co-occurrence and its evolution in cross-border M&As, conducting theory–topic co-occurrence analysis that identifies popular research trends and provides new insights into how to integrate M&A practices with theories, identifying ten subareas via bibliometric analysis and BER Topic modeling, and identifying key theories and emerging themes to be examined in future studies [23].

P. Nandi et.al. (2023) explain how M&A can be impacted when there is a failure in cybersecurity assessment and why the organization faces cybersecurity breaches during M&A. Analysis of cybersecurity risk in the merger and acquisition life cycle over time. This paper is a step-by-step guide to the key cybersecurity framework, Secure, Vigilant, and Resilient, for pre-merger to post-merger M&A, explained in subsequent phases. Also, the traditional way for mergers and acquisitions is outdated, and adopting the modern cybersecurity assessment is mandatory during mergers and acquisitions [24].

H. P. Bomma (2022) investigates the key factors and best practices for data engineers when merging data systems in the post-merger phase. It looks into data integration, data integrity maintenance, and the utilization of data for strategic advantage. The following sections will examine and look at the steps to achieve data integration and maximize the benefits of the merger [25].

Research Gap: Several studies have addressed metadata-driven data engineering, data interoperability, cybersecurity assessment, data digital transformation, and data integration after mergers. Despite this, there are still many research gaps pertaining to the integration of these capabilities into a comprehensive approach to enterprise M&A. Existing methods offer a limited support for automated discovery of metadata, schema matching for heterogeneous schemas, semantic data integration and automatic metadata synchronization between merging organizations. In addition, further research is needed in using AI for adaptive workflow optimization and predictive integration planning. Lack of real-

DOI: https://doi.org/xxxxxxxxxxxx

world validation, scalability, data governance, and security concerns are other areas of concern that are not addressed. Future research is needed to create an intelligent metadata-driven framework for reliable, scalable, secure, and efficient enterprise data integration for the entire life cycle of M&A.

Table I presents a comparative analysis of existing studies on metadata-driven data engineering for enterprise mergers and acquisitions. It summarizes the research focus, methodologies, key findings, limitations, and future directions, highlighting metadata intelligence, data interoperability, post-merger integration, cybersecurity, digital transformation, and workflow optimization to identify research gaps and opportunities for developing efficient, adaptive, and intelligent enterprise data integration systems.

TABLE I. COMPARATIVE ANALYSIS OF METADATA-DRIVEN DATA ENGINEERING AND MERGERS AND ACQUISITIONS STUDIES

Reference	Study Focus	Approach / Methodology	Key Findings	Challenges / Limitations	Future Directions
J. Bobba et al. (2026)	Intelligent metadata-driven enterprise data engineering	AI-enabled metadata framework with adaptive workflows	Improved automation, recommendations, decision-making, and workflow efficiency	High dependence on metadata quality and adaptability	Validate scalability and real-world applicability
R. Zhao, Z. Yan, and L. Zhao (2026)	Digital transformation and green M&A	Empirical analysis of Chinese listed enterprises (2010–2023)	Digital transformation promotes green M&A under specific CEO career conditions	Limited geographical and contextual scope	Examine other countries and organizational settings
B. Babajanov et al. (2025)	Metadata-driven data interoperability	Domain-agnostic reference schema and synthetic case study	Improved metadata-based integration across isolated systems	Limited practical validation in real enterprises	Test across diverse domains and large-scale systems
J. Liu et al. (2024)	Cross-border M&A research trends	Bibliometric analysis and BERTopic modeling	Identified major research themes, theories, and emerging topics	Dependence on available publications and research coverage	Explore emerging themes and underrepresented areas
P. Nandi et al. (2023)	Cybersecurity throughout the M&A lifecycle	Lifecycle-based cybersecurity risk assessment framework	Emphasized security assessment from pre-merger to post-merger stages	Limited empirical validation of the proposed practices	Develop automated security assessment and monitoring
H. P. Bomma (2022)	Post-merger data integration and engineering	Analysis of data integration practices and best practices	Highlighted data integrity, system integration, and strategic data use	Limited detail on automated and metadata-driven integration	Investigate intelligent integration and automated data governance

V. CONCLUSION AND FUTURE WORK

Metadata-driven data engineering provides an approach to managing heterogeneous data during mergers and acquisitions. The reviewed studies demonstrate the value of metadata management, data interoperability, cybersecurity assessment, and post-merger integration in improving data consistency and supporting decisions. Metadata catalogs, lineage tracking, quality rules, access controls, and standardized integration layers can help organizations discover assets, understand transformations, protect sensitive information, and coordinate data movement across different platforms. Modern architectures, including data lake houses, data mesh, and Data Vault, support scalable and adaptable enterprise data management. However, existing approaches remain fragmented, and challenges persist in automated metadata discovery, schema matching, semantic alignment, continuous synchronization, and real-world validation. Differences in governance policies, infrastructure, and regulatory requirements also complicate integration. Addressing these limitations requires a unified framework that combines metadata intelligence, automated workflows, governance, and continuous monitoring. Such a framework could make enterprise integration more reliable, secure, and efficient throughout the merger and acquisition lifecycle.

Future research should develop metadata-driven frameworks that automate asset discovery, schema matching, semantic alignment, and synchronization across heterogeneous enterprise systems. Large-scale evaluations are needed to assess scalability, accuracy, operational cost, and reliability. Research should strengthen privacy, access control, regulatory compliance, and adaptive workflow optimization throughout the merger and acquisition lifecycle.

- **Funding:** This research received no external funding.
- **Institutional Review Board Statement:** Not applicable.
- **Informed Consent Statement:** Not applicable.
- **Data Availability Statement:** The data supporting the findings of this study are available from the corresponding author upon reasonable request.
- **Conflicts of Interest:** The author declares no conflict of interest.

How to cite this article:
P. Gautam, "A Review of Metadata-Driven Data Engineering for Enterprise Mergers and Acquisitions," *Journal of Global Research in Multidisciplinary Studies* , vol. 2, no. 11, pp. 1–7, Nov. 2026. Doi: XXXX

REFERENCES

[1] T. Řezník *et al.*, "Improving the documentation and findability of data services and repositories: A review of (meta)data management approaches," *Comput. Geosci.*, vol. 169, p. 105194, Dec. 2022, doi: 10.1016/j.cageo.2022.105194.

[2] A. Guruprakash, "The Integration Tax: A Practitioner's Framework for Technology Mergers and Acquisitions," 2026. doi: 10.2139/ssrn.7283078.

[3] M. F. Malik, M. A. Anuar, S. Khan, and F. Khan, "Mergers and Acquisitions: A Conceptual Review," *Int. J. Account. Financ. Report.*, vol. 1, no. 1, p. 520, Nov. 2014, doi: 10.5296/ijafr.v4i2.6623.

[4] K. Ngcobo, S. Bhengu, A. Mudau, B. Thango (Y2-rated Researcher), and L. Matshaka, "From Single Shot to Structure: End-to-End Network based Deflectometry for Specular Free-Form Surface Reconstruction," Sep. 25, 2024. doi: 10.20944/preprints202409.1913.v1.

[5] U. Momin, "Agentic AI for Autonomous Metadata Management and Enterprise Data Catalogues: Automating Metadata Discovery, Classification, Semantic Enrichment, Data Lineage and Stewardship under DAMA and NDMO Frameworks," *Glob. Acad. J. Econ. Bus.*, vol. 8, no. 05, pp. 1289–1299, Sep. 2026, doi: 10.36348/gajeb.2026.v08i05.021.

[6] S. Kilaru, "Automated ETL Intelligence: Metadata-Orchestrated Framework with Rule-Based Heuristics for Monitoring and Reporting," *Int. J. Inf. Electron. Eng.*, vol. 3, no. 6, p. 14, 2013.

[7] W. Yang, R. Fu, M. B. Amin, and B. Kang, "The Impact of Modern AI in Metadata Management," *Human-Centric Intell. Syst.*, vol. 5, no. 3, pp. 323–350, Jul. 2025, doi: 10.1007/s44230-025-00106-5.

[8] K. Lyko, M. Nitzschke, and A.-C. Ngonga Ngomo, "Big Data Acquisition," in *New Horizons for a Data-Driven Economy*, J. M. Cavanillas, E. Curry, and W. Wahlster, Eds., Cham: Springer International Publishing, 2016, pp. 39–61. doi: 10.1007/978-3-319-21569-3_4.

[9] M. Mahdavi, F. Neutatz, L. Visengeriyeva, and Z. Abedjan, "Towards Automated Data Cleaning Workflows," 2019, pp. 1–10.

[10] J. Pepper, A. Senin, D. Jebbia, D. Breen, and J. Greenberg, "Metadata Verification: A Workflow for Computational Archival Science," in *2022 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2022, pp. 2565–2571. doi: 10.1109/BigData55660.2022.10020340.

[11] N. F. Mosha and P. Ngulube, "Metadata Standard for Continuous Preservation, Discovery, and Reuse of Research Data in Repositories by Higher Education Institutions: A Systematic Review," *Information*, vol. 14, no. 8, p. 427, Jul. 2023, doi: 10.3390/info14080427.

[12] V. Sikarwar, "Intelligent Metadata-Driven Data Engineering: Accelerating Standardized, Scalable Data Pipelines," *Int. J. Res. Appl. Innov.*, vol. 8, no. 1, pp. 11690–11699, 2025, doi: 10.15662/IJRAI.2025.0801006.

[13] A. Teymurova, "Metadata Management In Data Governance: A Comprehensive Scientific Analysis," 2025.

[14] M. Koniaris, D. Spentzou, M. Friedemann, H. Mischo, J. Soldatos, and G. Kouzas, "A Metadata-Driven Execution Model for Unified Integration and Management of Heterogeneous IoT Data Sources," *IoT*, vol. 7, no. 3, p. 58, Jul. 2026, doi: 10.3390/iot7030058.

[15] C. Rodríguez *et al.*, "REST APIs: A Large-Scale Analysis of Compliance with Principles and Best Practices," in *International conference on web engineering*, 2016, pp. 21–39. doi: 10.1007/978-3-319-38791-8_2.

[16] F. T. Lohrke, C. Frownfelter-Lohrke, and D. J. Ketchen, "The role of information technology systems in the performance of mergers and acquisitions," *Bus. Horiz.*, vol. 59, no. 1, pp. 7–12, Jan. 2016, doi: 10.1016/j.bushor.2015.09.006.

[17] M. H. Hamza Afzal, "Securing AI Systems Against Adversarial Attacks: A Framework for Building Robust and Trustworthy Machine Learning Models," *Comput. Fraud Secur.*, no. 5, pp. 65–74, May 2024, doi: 10.52710/cfs.867.

[18] S. Chilakala, "Enterprise Data Architectures : A Comprehensive Analysis Of Modern Solutions , Market Trends , And Mplementation Frameworks," *Int. J. Res. Comput. Appl. Inf. Technol.*, vol. 8, no. 1, pp. 3388–3399, Feb. 2025, doi: 10.34218/IJRCAIT_08_01_243.

[19] B. R. Nida, "The Evolution of Enterprise Data Architectures: Unifying Warehouses, Lakes, Lakehouses, and Mesh for AI-Ready Analytics," *J. Front. Multidiscip. Res.*, vol. 6, no. 2, pp. 507–512, 2025, doi: 10.54660/.JFMR.2025.6.2.507-512.

[20] J. Bobba, A. Srivastava, P. S. Bhadoria, V. K. Shukla, A. J. Christy, and S. V. V. Prasad, "An AI Automated Metadata-Driven Approach for Enterprise Data Engineering Workflows," in *2026 International Conference on Science Technology Engineering and Management (ICSTEM)*, IEEE, Aug. 2026, pp. 1–8. doi: 10.1109/ICSTEM68144.2026.11735511.

[21] R. Zhao, Z. Yan, and L. Zhao, "Digital Transformation, CEO Career Concerns, and Enterprise Green Mergers and Acquisitions," *IEEE Trans. Eng. Manag.*, vol. 73, pp. 936–948, 2026, doi: 10.1109/TEM.2025.3647692.

[22] B. Babajanov, L. Hein, F. Kodirov, M. C. Tank, A. Jalilian, and C. Schönberg, "Metadata-Driven Data Interoperability," in *2025 IEEE 26th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, IEEE, Jun. 2025, pp. 2120–2125. doi: 10.1109/EDM65517.2025.11096744.

[23] J. Liu, Q. Meng, F.-M. Tseng, and Y. Sun, "Three Decades of Cross-Border Mergers and Acquisitions Research: Mapping the Field and Paths Forward," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 5847–5868, 2024, doi: 10.1109/TEM.2024.3371055.

[24] P. Nandi, A. Fatma, P. Pathak, and V. Bhatt, "Securing Business Intelligence during Mergers and Acquisition: Critical Cybersecurity Imperatives," in *2023 4th International Conference on Data Analytics for Business and Industry (ICDABI)*, IEEE, Oct. 2023, pp. 315–320. doi: 10.1109/ICDABI60145.2023.10629281.

[25] H. P. Bomma, "Navigating Data Integrations Post Mergers & Acquisitions A Data Engineer ' s Perspective," *Int. J. Innov. Res. Creat. Technol.*, vol. 8, no. 6, pp. 6–10, 2022, doi: 10.5281/zenodo.14787277.