ISSN: 2754-6659

Journal of Artificial Intelligence & Cloud Computing

SCIENTIFIC Research and Community

Review Article

Open Access

Data Quality Engineering Frameworks for Regulatory-Grade AML Transaction Monitoring Systems

Mukesh Kumar Mishra

Individual Contributor, Financial Technology & Data Engineering, USA

ABSTRACT

Anti-Money Laundering (AML) systems function within tightly enforced compliance architecture defined by global regulators. The increasing scale and sophistication of financial fraud and money laundering have made robust AML controls critically important. Breakdowns in AML controls can expose institutions to reputational damage, regulatory penalties, and systemic financial risk. These challenges necessitate the adoption of systematic, data-driven approaches for effective risk detection and mitigation. Contemporary AML platforms integrate rule engines with machine learning models to evaluate transactional risk patterns, various risk scoring models demonstrating data integrity checks across multiple stages of data lifecycle. Despite these advancements, financial crime incidents continue to rise, indicating gaps in current system effectiveness. Still a lot needs to be explored in data engineering to help in data analysis and resolution. This paper proposes data quality engineering (DQE) model for regulatory grade AML environments. **DQE framework focusses on 5 data quality dimensions - completeness, accuracy, consistency, timeliness, and lineage integrity**. This architecture uses Microsoft Azure, Azure Databricks and data will be stored in Delta Lake. Technical validations performed using above model also aligns with the regulatory compliances and obligations under the Bank Secrecy Act (BSA), Financial Action Task Force (FATF) recommendations, and Office of Foreign Assets Control (OFAC). This study hypothesizes that DQE implementation can reduce false negatives and can improve financial crime detection.

***Corresponding author**

Mukesh Kumar Mishra, Individual Contributor, Financial Technology & Data Engineering, USA.

Received: April 30, 2026; **Accepted:** May 08, 2026; **Published:** May 15, 2026

Keywords: Data Quality Engineering, AML Transaction Monitoring, Regulatory Compliance, Azure Databricks, Delta Lake, Financial Crime Analytics, Data Lineage, BSA/AML

Introduction

In the current market, effectiveness of AML analytics is directly influenced by the quality of input data used in detection models and reporting pipelines. These regulatory pipelines are designed to be operationally efficient, auditable, and compliant with reporting standards. Regulatory audits consistently highlight deficiencies in data integrity and weak governance as key risk drivers in AML failures.

Despite these concerns, industry efforts have largely prioritized detection algorithm development over foundational data quality engineering which includes machine learning algorithms, graph analytics and Natural Language Processing (NLP) while treating quality of data as a pre-condition for further processing but not as a part of architectural design. This imbalance introduces a critical gap and results in implementation issues, compromised data pipelines, unreliable alerts and incomplete Suspicious Activity report (SAR).

This paper aims to address and illustrate approaches to bridging these gaps. The five quality dimensions model described in the abstract is derived from regulatory guidance and industry data management standards, and is operationalized through engineering controls and can be implemented within modern cloud native data platforms. This engineering model includes -

- Formal approach to Data Quality Dimension with regulatory mapping
- Engineering Control architecture for quality – data ingestion, data curation and analytic engines
- Linked quality control with compliance control
- Alert accurately and timely
- Audit readiness

Regulatory requirement for AML Data quality

Financial regulators enforce strict data quality and reporting obligations on AML monitoring systems for data directly coming from source or direct input by the user. The Bank Secrecy Act (BSA) requires institutions to implement internal controls for identifying and reporting suspicious activity. Similarly, the Office of the Comptroller of the Currency (OCC) and the Financial Crimes Enforcement Network (FinCEN) have issued guidance emphasizing the importance of data completeness and integrity as foundational AML program components. OFAC screening processes rely on real-time data processing, supported by robust lineage tracking and auditability

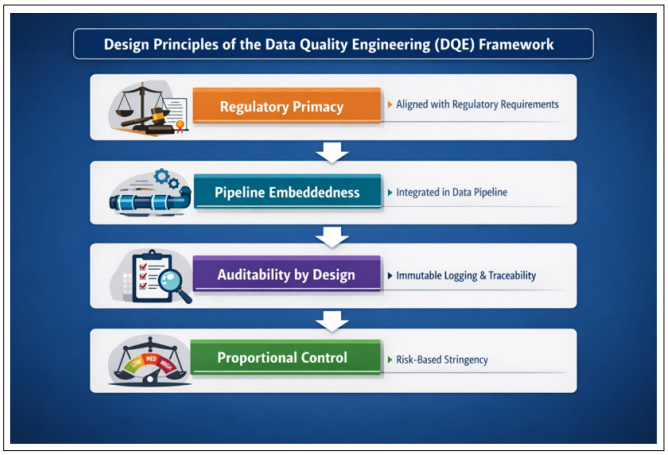
Data Quality framework in Enterprise Data Management (EDM) - Traditional process

Data Management Body of Knowledge (DMBOK), published by DAMA International, identifies multiple foundational dimensions of data quality: completeness, consistency, conformity, accuracy, integrity, and timeliness. Wang and Strong further expand data

quality concepts by incorporating contextual and accessibility dimensions. However, traditional methods were not designed to include specific financial compliance data rules, requirements of audit trail management, real time or streaming data transactions monitoring. Traditional approaches often overlook data pipeline quality as an independent engineering concern [1-4].

Proposed Data Quality Engineering Framework for AML systems
The proposed Data Quality Engineering model has 4 design pillars which is derived from the regulatory practice and traditional process of DQE [5-9].

Figure:1



AML Data Quality Dimensions
The model defines 5 quality dimensions for AML transactions. Fig-2 shows general details about the 5 data quality dimensions. Fig-3 shows the relation between the regulators and Engineering Controls [10-14].

Figure:2

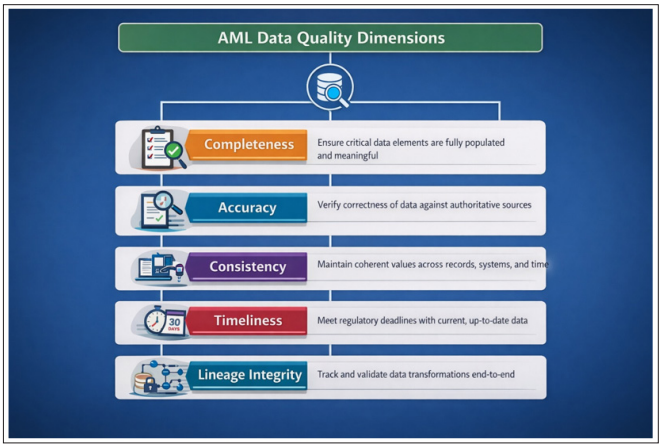
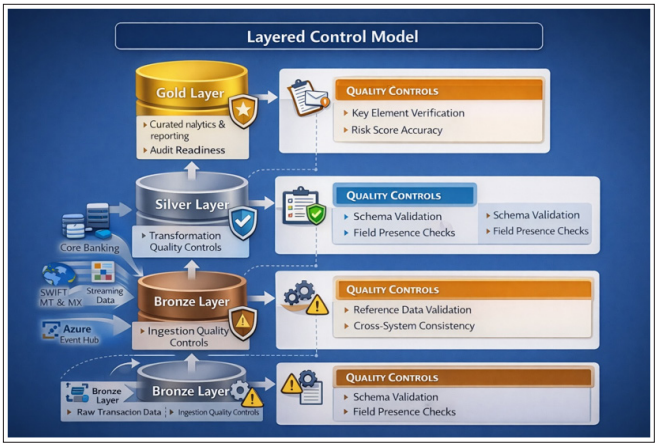


Figure:3

Quality Dimension	Primary Regulatory Reference	Engineering Control Layer
Completeness	BSA 31 C.F.R. §1020.320; FinCEN SAR Filing Requirements	Bronze ingestion validation; Silver schema enforcement
Accuracy	OFAC SDN Screening; FATF Recommendation 10	Silver reference matching; Gold cross-validation
Consistency	OCC BSA/AML Examination Procedures §2000	Silver reconciliation controls; cross-system sync checks
Timeliness	BSA 30-day SAR filing; OFAC real-time screening mandate	Event Hub latency monitoring; streaming SLA enforcement
Lineage Integrity	FATF Recommendation 11; EU AMLD6 Article 40	Delta Lake transaction logs; Unity Catalog lineage graph

DQE framework – Architecture
The DQE model implements quality controls across a three-layer architectural model aligned with the Medallion architecture pattern.

Figure:4



Key Findings -Proposed Dqe Framework

- It is expected that institutions adopting structured DQE frameworks will observe reductions in false-negative alert rates
- Improved data accuracy may contribute to reduced false-positive alerts due to improved counterparty identification, reducing misclassification of legitimate transactions as suspicious
- Enhanced lineage tracking can potentially shorten audit remediation cycles, reflecting the operational value of the lineage integrity dimension
- The model is designed to minimize SAR filing errors through improved data validation controls

Conclusion

This paper has introduced a comprehensive Data Quality Engineering framework for regulatory-grade AML transaction monitoring systems, addressing a significant gap in the existing literature between detection methodology research and data engineering practice. As regulator monitoring continues to grow, Financial institutions that treat data quality engineering as a core compliance function rather than an operational convenience will be better positioned to demonstrate program effectiveness and withstand regulatory examination. Future work will extend the model to address cross-institutional data reliability challenges in correspondent banking networks [15-18].

References

1. Financial Action Task Force (2023) 'International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation—The FATF Recommendations,' FATF/OECD, Paris https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html.
2. (2023) Financial Crimes Enforcement Network (FinCEN), 'Bank Secrecy Act/Anti-Money Laundering Examination Manual,' U.S. Department of the Treasury https://www.occ.treas.gov/news-issuances/bulletins/2023/bulletin-2023-39.html.
3. (2022) Office of the Comptroller of the Currency (OCC), 'BSA/AML Comptroller's Handbook,' OCC, Washington DC https://www.occ.treas.gov/topics/supervision-and-examination/bsa/index-bsa.html.
4. (2017) DAMA International, 'DAMA-DMBOK: Data Management Body of Knowledge,' 2nd ed., Technics Publications https://dama.org/learning-resources/dama-data-management-body-of-knowledge-dmbok/.
5. Wang R, Strong D (1996) Beyond Accuracy: What Data Quality Means to Data Consumers,' Journal of Management Information Systems 12: 5-33.
6. ISO (2022) ISO 8000-2: Data Quality—Vocabulary,' International Organization for Standardization, Geneva https://www.iso.org/obp/ui/en/#iso:std:iso:8000:-2:ed-5:v1:en.
7. Weber M, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, et al. (2019) Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics,' IEEE Transactions on Big Data https://arxiv.org/abs/1908.02591.
8. Kou Y, Chang-Tien Lu, Sirwongwattana S, Yo Ping Huang (2004) Survey of Fraud Detection Techniques,' IEEE International Conference on Networking, Sensing and Control https://ui.adsabs.harvard.edu/abs/2004icns....2...19Y/abstract.
9. Databricks (2020) Delta Lake: High-Performance ACID Table Storage over Cloud Object Stores,' VLDB Proceedings https://ir.cwi.nl/pub/32924.
10. (2024) Microsoft Azure Architecture Center, 'Big Data Analytics in Financial Services,' Microsoft Corporation https://www.sprintzeal.com/blog/azure-architecture.
11. Alharbi NM (2021) 'Machine Learning Approaches for Financial Fraud Detection,' IEEE Access https://www.ijert.org/an-analytical-review-of-ai-and-machine-learning-techniques-for-financial-fraud-detection-ijertconv14is020042.
12. (2019) Office of Foreign Assets Control (OFAC), 'A Framework for OFAC Compliance Commitments,' U.S. Department of the Treasury https://ofac.treasury.gov/recent-actions/20190502_33.
13. (2018) European Parliament, 'Directive (EU) 2018/1673 of the European Parliament—Sixth Anti-Money Laundering Directive (AMLD6),' Official Journal of the European Union https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L1673.
14. Singh S (2024) Real-Time AML Analytics Using Distributed Streaming Architectures,' IEEE Big Data Conference Proceedings http://eprints.umsida.ac.id/16354/1/AIML-Powered%20Anti-Money%20Laundering%20Pipelines.pdf.
15. NIST (2023) AI Risk Management Framework (AI RMF 1.0),' National Institute of Standards and Technology https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.
16. Gartner (2024) 'Magic Quadrant for Cloud Database Management Systems,' Gartner Research https://www.gartner.com/en/documents/7195930.
17. McKinsey Company, 'The Fight Against Financial Crime: A Global Regulatory Update,' McKinsey Global Institute https://www.mckinsey.com/~/media/mckinsey/business%20functions/risk/our%20insights/mckinsey%20on%20risk%20number%2014/mckinsey%20on%20risk%20and%20resilience%20issue%2014.pdf.
18. (2020) Basel Committee on Banking Supervision, 'Sound Management of Risks Related to Money Laundering and Financing of Terrorism,' Bank for International Settlements https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Risk-Based-Approach-Banking-Sector.pdf.coredownload.pdf.