Modern™ WhitePaper

Operationalizing Data Quality in Banking: From Point Checks to Systemic Controls

Table of Contents

Executive Summary

Why Banks Need Systemic Data Quality

Banks continue to struggle with data quality not because the problem is misunderstood but because it is addressed through fragmented checks rather than systemic controls. As regulatory scrutiny intensifies and data is used more directly in risk models, customer decisions, and AI-driven processes, this approach is no longer sufficient.

Most data quality efforts focus on technical validation, including null checks, thresholds, and schema rules. While necessary, these measures do not prevent failures caused by inconsistent enforcement, upstream changes, or missing business context. Data can meet technical standards and still create material risk.

This paper makes the case for a shift toward systemic data quality by treating quality as an enterprise control system embedded across data pipelines and consumption. By operationalizing quality through clear ownership, context-aware controls, and auditable response mechanisms, banks can reduce risk, improve trust, and support analytics and AI at scale without replacing their existing data stack.

Introduction

The financial services sector, particularly banking, operates on a foundation of data. The integrity, accuracy, and timeliness of this data are paramount, as they directly impact regulatory compliance, risk management, and strategic decision-making. Poor data quality is not merely an operational inconvenience; it is a significant financial risk, with many numerous high-profile examples of major financial losses and significant reputational impact. Addressing this challenge requires a comprehensive understanding of the common problems and a commitment to implementing holistic, strategic solutions.

The Full Spectrum of Data Quality

Data Quality is often thrown around as a generic term that oversimplifies the realities. There are 6 dimensions of Data Quality that are industry-accepted standards: Accuracy, Completeness, Consistency, Validity, Uniqueness, and Timeliness. To those, I would add these two as sub-categories: regulatory and risk. All of these dimensions are important, and you should be treating them holistically to maximize the value to your organization. Many data programs in Financial Services will focus on Completeness and Accuracy and overlook the others, resulting in sub-optimal improvements and frustration with the value of the investment into a Data Quality Program. DAMA.org has a number of great resources for implementing these quality dimensions.

Common Data Quality Problems in Banking

Data quality issues in banking are often rooted in a combination of legacy systems, fragmented data architectures, poorly designed business processes, and human error. These problems manifest across the critical dimensions that have been mentioned before, undermining the reliability of information used for everything from customer service to regulatory reporting. A Gartner survey highlights that a significant percentage of banks struggle with data quality and integrity, facing issues like gaps in important data points and incomplete transaction flows.

The following table summarizes the most common data quality problems, their underlying causes, and their potential impact on financial institutions.

Table 1: Most Common Data Quality Problems, Causes, and Impact

Problem Category	Specific Issues	Underlying Causes	Impact
Accuracy & Completeness	Inaccurate data entry, missing payment information, gaps in important data points, and incomplete transaction records.	High reliance on manual data input, lack of standardized data collection protocols, and non-scalable data processes.	Leads to financial mismanagement, flawed risk calculations, and poor customer experience due to incorrect information.
Consistency & Uniqueness	Duplicate customer records, inconsistent data formats across different systems, and discrepancies in customer details (e.g., address) between departments.	Fragmented legacy IT infrastructure, the absence of a central Single Source of Truth (SSOT), and poor data synchronization.	Causes reconciliation issues, increases operational costs, and introduces potential financial and compliance risks.
	Managing a wide variety of data formats, including structured data (spreadsheets) and unstructured data (emails, PDFs), with no consistent schema.	Data collection from multiple, disparate sources with varying technical requirements and formats.	Creates difficulty in data processing and analysis, leading to discrepancies and data quality degradation during integration.

Problem Category	Specific Issues	Underlying Causes	Impact
Accessibility & Timeliness	Data silos that prevent a unified view of the customer, fragmented data systems, and the inability to access transaction data in real-time.	Outdated technology, complex and disparate data sources (e.g., transactions, market data, customer profiles).	Hinders the effective use of advanced analytics and AI, prevents timely and strategic decision-making, and slows down critical regulatory reporting.
Regulatory & Risk	Data that fails to meet stringent regulatory requirements (e.g., BCBS 239, GDPR) or data that compromises the integrity of risk models.	Unclear data ownership, limited understanding of data risks across the enterprise, and frequent changes in regulatory mandates.	Results in substantial fines and sanctions for non-compliance, leading to incorrect risk assessments and increased exposure to market and credit risks.

Strategic Solutions and Best Practices

Addressing these complex data quality challenges requires a holistic strategy that integrates governance, technology, and organizational culture. The solution is not solely a technological one; it must be driven by a top-down commitment to data excellence.

The most effective solutions focus on establishing a robust framework for data management and platforms that embed quality controls across the data lifecycle.

Table 2: Solutions, Best Practices and Benefits

Solution Area	Key Actions & Best Practices	Benefits
Data Governance Framework	Establish a formal, top-down framework that defines clear policies, standards for data collection, quality controls, and accountability for data ownership. Align the strategy with core business and regulatory objectives.	Provides a structured, repeatable approach to managing data, ensures compliance with regulations, and clarifies roles and responsibilities across the organization.
Continuous Data Audits and Monitoring	Implement regular data audits to systematically check data against defined quality dimensions (accuracy, completeness, timeliness). Deploy data quality and observability tools for real-time monitoring and automated detection of anomalies and deviations.	Enables proactive identification and remediation of data issues, ensuring data remains reliable and fit-for-purpose for all consumers.
Technology Modernization and Centralization	Invest in modern data platforms to break down silos and establish a Single Source of Truth (SSOT). Utilize AI/ML-powered tools for automated data cleansing, validation, and statistical analysis to handle high-volume, high-velocity data.	Significantly improves data accessibility and consistency, enabling advanced analytics and faster, more informed decision-making.

Solution Area	Key Actions & Best Practices	Benefits
Culture and Training	Foster a data-focused culture where data quality is a shared responsibility. Provide mandatory data quality training for all employees, especially those involved in data creation and entry, to minimize human-induced errors at the source.	Cultivates a proactive approach to data management, increases data literacy, and embeds quality checks into daily operational workflows.
Data Lineage and Metadata Management	Implement robust metadata management to define and catalog data assets. Track data lineage to understand the data's journey from its source through all transformations to its final use.	Increases transparency and trust in data, simplifies the process of compliance reporting, and significantly speeds up root cause analysis when data issues arise.

Operationalizing Data Quality Solutions

The best practices we just outlined—governance, continuous monitoring, modernization, culture, and lineage—are necessary, but they are not sufficient on their own. **Many data quality programs fail to deliver because "data quality" quickly becomes a narrow set of technical checks: null percentages, value ranges, and basic schema validations.** Those checks matter, but they do not catch many of the issues that create real risk in banking.

To operationalize data quality, quality controls need to be applied with the business and operational context. In practice, that means validating data not only for what it is (a value in a column), but also for what it specifically represents (a customer, a product, a contract, a jurisdiction, a currency) and how it will be used (pricing, eligibility, risk models, regulatory reporting, customer communications). When context is missing, "passing" data can still drive incorrect outcomes.

Operationalizing data quality requires three additional disciplines that should be treated as a continuation of the strategic best practices:

1. **Contract and schema drift controls**

   Banking data pipelines increasingly rely on internal producers and third-party sources. Even small upstream schema changes (type changes, field renames, dropped fields, precision changes) can cause silent failures or incorrect downstream computations. Effective programs treat schema and contract validation as a first-class quality control measure, with clear actions when drift is detected (e.g., quarantine, stop ingestion, notify owners, or roll back to a the last-known-good contract).

2. **Source-aware harmonization controls**

   In global banking environments, the same business concept can arrive in different formats depending on the source system, region, or product. Date formats, account identifiers, categorical codes, and reference data often require harmonization. The important point is that these checks are not always universal; they are often specific to the data source and to the transformation that will be applied. A good operational model validates "fit for transformation" before data is standardized and consumed.

3. **Business-rule validity controls**

   Many of the costliest data quality failures are not "invalid values," but "valid values in the wrong context." Examples include product eligibility logic (e.g., retirement products tied to age constraints), tiering and pricing logic, and cross-currency handling where thresholds and interest rules depend on base currency. These controls require business context and ownership; they also require an escalation path when exceptions occur.

A practical way to implement these disciplines is to treat data quality as a control loop:

- **Detect:** identify anomalies or violations (technical, contractual, harmonization, and business-rule checks)
- **Decide:** classify severity and determine response based on policy (warn, quarantine, block, route)
- **Act:** execute the response and notify accountable owners and stewards
- **Audit:** retain lineage, evidence, and outcomes for compliance, reporting, and continuous improvement

This control-loop approach bridges strategy and execution: governance defines policies and assigns ownership; monitoring detects issues; modernization and centralization ensure consistent controls; culture enables timely remediation and lineage provides auditability and root-cause analysis capabilities.

How A Strategic Partner Solution Can Help

Financial services organizations are not software companies, so the idea of building custom solutions should only be the option if you do not have a robust commercial solution available to you. DataOS from The Modern Data Company provides the operating layer to implement data quality at scale.

Contextual data quality is the requirement, and DataOS is the first data operating systems that makes contextual data quality deployable, governable, and auditable across the enterprise. Instead of treating quality as a set of disconnected checks embedded in code and scattered across teams, DataOS centralizes quality controls, standardizes responses, and ties controls to the business context in which data is produced and consumed.

DataOS as a Data Quality Control Plane for Banking

The shift to cloud-native data platforms like DataOS represents a fundamental strategic solution for the data quality challenges faced by the banking sector. By providing a unified, scalable, and governed platform, DataOS helps financial institutions move beyond the limitations of fragmented legacy systems to establish a robust data foundation.

In banking specifically, the value of DataOS extends beyond basic data checks. DataOS acts as a control plane for data quality by enabling institutions to define quality expectations once, enforce them consistently across pipelines and products, and operationalize response and auditability. This is where business context becomes practical: controls can be tied to specific sources, regions, products, contracts, and downstream use cases.

A control-plane approach is built on four capabilities:

- Centralized, repeatable controls: quality rules are defined centrally and reused across the lifecycle (ingestion, transformation, and consumption), reducing fragmentation and inconsistency.
- Context-aware validation: rules can be applied based on source system, region, dataset, product, and other metadata signals so that checks reflect real-world operating conditions.
- Monitoring: issues are detected during ingestion, and can trigger notifications and escalations to data owners/stewards.
- Evidence and auditability: governance metadata and lineage provide traceability from source to consumption for both internal risk management and regulatory reporting.

How DataOS Solves Banking Data Quality Issues

The DataOS approach directly tackles the root causes of poor data quality, such as data silos, inconsistent controls across teams, the lack of automated quality checks as well as accountability and ownership. The following table maps common banking data quality problems to the specific DataOS capabilities that address them.

Table 3: DataOS Capabilities to Address Data Quality Problems

Data Quality Problem	DataOS Capability	How it Helps
Consistency & Uniqueness (data silos, duplicate records)	**Single, Centralized Platform (Data Cloud)**	By consolidating data from disparate sources into a single, centralized platform, DataOS eliminates data silos and establishes a true Single Source of Truth (SSOT). This inherently resolves inconsistencies and duplicate records that arise from fragmented legacy systems.
Accuracy & Completeness (missing data, incorrect values)	**Data Quality Framework and Data Metric Functions (DMFs)**	DataOS's native Data Quality Framework uses Data Metric Functions (DMFs) for automated, continuous quality checks. These checks (e.g., null value checks, range validation, uniqueness constraints) are applied at the point of ingestion and transformation, ensuring data meets pre-defined standards before it is used for critical functions.

Data Quality Problem	DataOS Capability	How it Helps
Contract & Schema Drift (upstream schema changes, type/precision changes, breakages)	**Contract Validation and Schema Monitoring**	Detects schema drift and contract mismatches early, enabling the institution to take actions such as stopping ingestion or quarantining data and notify owners before downstream systems and analytics inherit incorrect or silently degraded data.
Harmonization Issues (cross-region formats, reference data mismatches)	**Context-Aware Rules and Transformation Controls**	Enables source-aware checks that validate “fit for transformation” and enforce harmonization rules (e.g., region-specific formats) prior to standardization and use, reducing integration-time degradation.
Business-Rule Validity (eligibility, tiering, pricing, cross-currency logic)	**Policy-Driven Controls and Operational Workflows**	Applies business-context validation where “valid values” can still be wrong for purpose; routes exceptions into stewardship workflows rather than leaving issues to be discovered downstream in reports or customer-facing processes.
Regulatory & Risk (auditing, sensitive data controls, model integrity)	**Governance Capabilities (Tags, Policies, Data Lineage)**	The platform provides robust data governance tools, including tags for classifying sensitive data (e.g., PII, BCBS 239-relevant data) and automated data lineage tracking. This simplifies compliance and auditing by providing an immutable, auditable record of data provenance and transformation.
Accessibility & Timeliness (slow access, fragmented systems, delayed reporting)	**Cloud-Native Architecture and Secure Data Sharing**	DataOS's cloud-native architecture provides the scalability and performance required for real-time data access and advanced analytics, replacing slow, fragmented legacy systems. Secure Data Sharing enables seamless, governed data flow across internal departments and with external partners, ensuring timely access to most current information.
Format & Structure (data variety across structured and semi/unstructured sources)	**Support for Diverse Data Types**	The platform natively handles a wide variety of data formats, including structured, semi-structured (JSON, XML), and unstructured data. This simplifies the integration of complex data sources (like transaction logs and data feeds) without the need for extensive, error-prone pre-processing, thus maintaining structural integrity.

The operational impact is tangible:

DataOS reduces the fragmentation of quality checks, improves consistency of enforcement, and makes it easier to incorporate business context into how quality is defined and acted on. This shift moves responsibility from manual, reactive processes to automated, proactive controls tied directly to governance policy and operational workflows.

Conclusion

Data quality is the bedrock of modern banking. The transition from fragmented, siloed data to a unified, high-quality data ecosystem is essential for managing risk, ensuring regulatory compliance, and unlocking the potential of advanced technologies like AI. By implementing a strategic data governance framework, investing in a data operating layer to implement data quality at scale, and fostering a culture of data responsibility, financial institutions can transform their data from a liability into a powerful strategic asset.

The DataOS platform reinforces these best practices by turning data quality into a repeatable control loop: detect, decide, act, and audit, while enabling the business context that makes quality meaningful in day-to-day banking use cases.

Written by: David Harberson, Albion Group

Contact Us

Learn more: **themoderndatacompany.com**
Schedule a demo: **info@tmdc.io**

Modern™