



Data Quality Foundations: Building Blocks for Financial Integrity

Puneet Matai

Associate Director – Enterprise Data Governance Group, Santander Bank N.A., USA
Email: puneet.matai@gmail.com

Abstract This whitepaper explores the significance of data integrity for risk management, compliance, and decision-making processes. It outlines common data quality challenges faced by financial institutions and proposes solutions such as robust data governance frameworks and regular audits. By prioritizing data quality, organizations can enhance operational efficiency, mitigate risks, and safeguard their reputation. The whitepaper concludes by emphasizing the strategic imperative of continuous data quality improvement for long-term success in the ever-evolving financial landscape.

Keywords Data Quality, Financial Industry, Risk Management, Compliance, Decision-Making, Data-Governance, Reputation, Operational Efficiency

Introduction

"Data Quality Foundations: Building Blocks for Financial Integrity" emphasizes the critical role of data quality in the financial industry. Covering key aspects such as risk management, compliance, and decision-making, the whitepaper underscores the importance of robust data governance frameworks and continuous improvement initiatives for long-term success.

Importance of Data Quality in the Financial Industry

Ensuring high data quality is paramount in the financial industry to avoid detrimental consequences. The development of a data governance framework is foundational in ensuring data quality. Regular audits are essential for evaluating the quality of financial data.

Poor data quality can lead to create inefficiencies in business operations. Moreover, inaccurate or incomplete data can result in errors, delays, and redundant efforts, ultimately impacting productivity and increasing costs. Recent research by Gartner reveals that organizations suffer an average annual loss of \$15 million due to subpar data quality [1]. For instance in banking, insurance and investment companies, data serves as an essential fuel. If it is compromised, it can lead to create catastrophic failures.

Hence, good quality data is an important requirement in financial organizations and maintaining data quality can ensure compliance with regulatory requirements avoiding penalties and legal consequences.

Impact of Poor Data Quality on Risk Management, Compliance and Decision Making

In the digital age, financial organizations rely heavily on data for strategic decision-making, operational optimization, and growth. The quality of data impacts the effectiveness of decision-making in organizations. Poor data quality can lead to flawed insights, misguided strategies, and ultimately, business failure.

The quality of data directly influences the quality of insights derived from analysis. Accurate and complete data leads to create reliable insights. On the other hand, inaccurate data can lead to the release of misguided conclusions. This can have potential financial losses or missed opportunities in the organizations.



Quality is crucial for compliance and risk management, as inaccurate data can lead to non-compliance and impact the privacy of the customers. Quality data ensures accurate risk assessment and efficient risk management. To improve data quality, financial organizations should implement measures such as data validation checks, regular data cleansing, and promoting a data-literate culture within the organization.

Identifying Data Quality Challenges

Common Data Quality Issues in Financial Data

Data quality issues are prevalent in many organizations and can have significant impacts on operations, decision-making, and compliance. When data is moved from one database to another during a transformation or upgrade of digital infrastructure, it can lead to various quality issues. It is evident from the statistics given below:

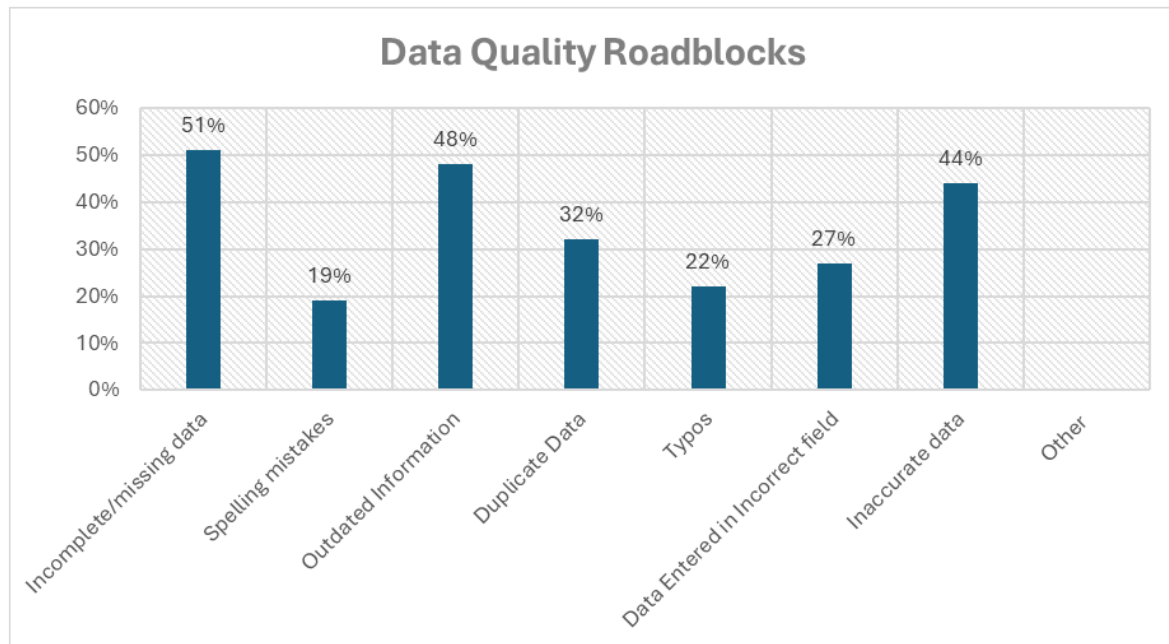


Figure 1: Data Quality Roadblocks [2]

Here are common data quality issues:

- **Incomplete or Inconsistent Data:** Data sets often contain missing or inconsistent information. It affects the data to conclude through accurate analysis with prompt decision-making. For instance, missing customer information in loan application forms can lead to processing delays or inaccurate credit assessments.
- **Duplicate Records:** Redundant data entries can manipulate analysis results and lead to inefficiencies in data processing. Banks may develop multiple entries for the same customer in a CRM system. It can result in creating redundant communications or even different records representing the same customer.
- **Data Decay:** Over time, data can become outdated or irrelevant, impacting the accuracy and relevance of insights derived from it. For example, outdated market data used for investment analysis may result in generating manipulated forecasts or ineffective portfolio management strategies.
- **Ambiguous Data:** Unclear or ambiguous data can lead to misinterpretation and incorrect conclusions and may compromise the entire process. For instance, unclear categorization of transaction types in financial records can lead to misunderstandings and errors in financial reporting.

Importance of Data Quality in Risk Management and Regulatory Compliance

“Data quality is the cornerstone of effective risk management and regulatory compliance”

The ability of financial institutions to effectively manage and govern big data hinges on the reliability and accuracy of their data. Without a sound data quality process in place, the institutions fail to address not only operational challenges but also those related to risk.



Deloitte Survey on Data Quality Management

Deloitte Netherlands conducted a survey which involved representatives from eight mid-sized banks in the Dutch banking industry [3]. The survey focussed on topics related to data quality management, including regulation, awareness, governance, processes, and systems. The key findings from the survey are given below:

- All respondents expressed concerns about the impact of supervisory and regulatory processes on data quality management.
- Banks that are not regulated by the European Central Bank (ECB) need to be aware of the evolving standards set by the Dutch National Bank (DNB) and ensure the implementation of regulatory compliance [3].
- All the respondents indicated a high likelihood of maintaining or scaling up their focus on data quality in the future.
- The respondents also presented their views on implementing end-to-end risk processes. It enables organizations to identify issues earlier and address them proactively

The above survey evidence that data quality is paramount in ensuring effective risk management and regulatory compliance in the financial industry. Overall, data quality serves as the foundation for building a robust risk management framework and ensuring the long-term success of financial institutions.

Data Quality Metrics and KPIs



Figure 2: Data Quality Dimension

For data teams, specifying and measuring data quality is crucial for ensuring the reliability and usability of their data sets. Traditionally, various dimensions have been used to assess data quality, such as accuracy, completeness, consistency etc., (Fig 2). However, without standardization, it can be challenging to effectively communicate and evaluate data quality. To address this issue, organizations often focus on a subset of primary dimensions and KPIs that are most critical for their needs.

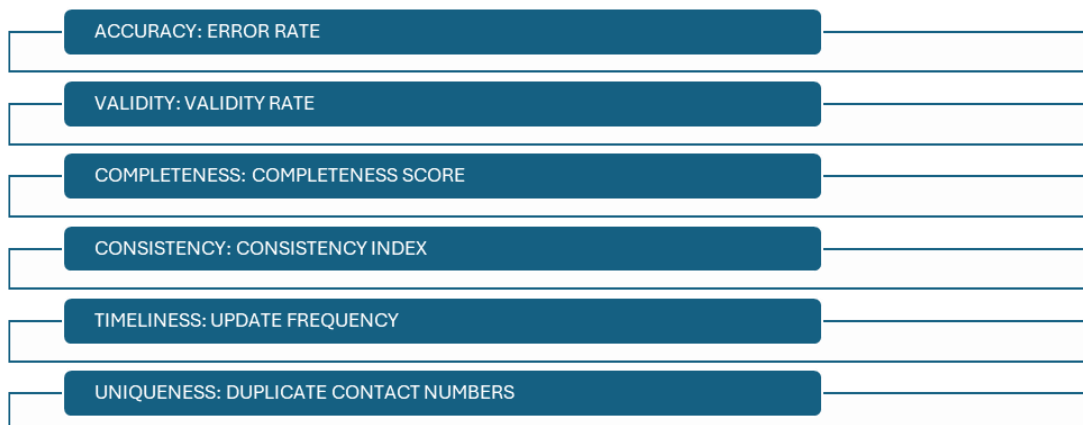


Figure 3: KPIs for Data Quality Dimensions/Metrics



The error rate is a measure of the percentage of inaccuracies or discrepancies found within the data completed to the actual values which helps in depicting **accuracy**. The **validity** rate measures the percentage of valid data entries compared to the total number of data entries [4].

Completeness is an important dimension whose metric is depicted using the completeness score which calculates the data entries compared to the total expected data entries. One important metric of completeness is *minimum occurrence* [5] which assures the number of occurrences to assess the time lag.

The **consistency** index measures the patterns and parameters of data sources over time. Numeric consistency is a metric which quantifies the degree to which numerical data aligns with permissible value ranges.

Timeliness measures the update frequency to ensure data accuracy. It reflects the time elapsed between data updates during data refresh cycles [4]. **Uniqueness** assures the record of a single entry by assessing the duplicate contact numbers or customer data within the dataset.

Here's why implementation of the above KPIs becomes crucial for the improvement of data quality:

- Implementing automated monitoring systems with these metrics can assess the discrepancies automatically, generating notifications or alerts.
- Setting up alerts to notify stakeholders when data falls outside of predefined thresholds ensures improvement.
- When inconsistencies are identified, performing root cause analysis using the KPIs can help in investigating error factors.
- Data profiling technique can analyse the distribution and patterns of data values within the data set.
- Establishing feedback loops as KPIs to encourage open communication and monitoring efforts to refine data quality processes.

Strategies for Data Quality Improvement

Best Practices for Identifying and Rectifying Data Quality Issues

The use of Data Life Cycle Phases (Define, Measure, Analyse, Improve, Control) can help in analysing the best practices of data quality issue management. For each phase, the problem and solutions will be outlined as examples to recommend the best practices:

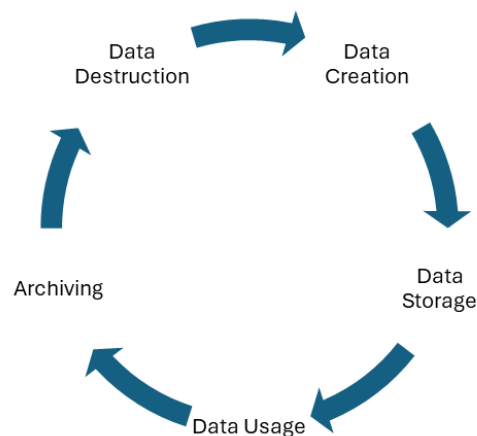


Figure 4: Data Life Cycle Management

Data Creation

- **Problem:** Organizations often overlook the importance of establishing clear processes and tools for data creation, leading to data that is inconsistent.
- **Solution:** Implement processes and tools for data creation ensuring that users have access to tools and are trained to create data in a standardized format.
- **Example:** A financial institution establishes standardized templates and guidelines for customer data entry in their CRM system



Data Storage:

- **Problem:** Without proper planning, organizations may struggle to balance factors like cost, accessibility, security, and compliance.
- **Solution:** Use a mix of storage media such as cloud, disk, and tape based on the specific needs of different data types to consider factors of accessibility and redundancy. Compliance with regulations like the Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry (PCI), Sarbanes-Oxley, Group Data Protection Regulation (GDPR), Personal Data Protection(PDP) and Security Exchange Commission (SEC) rules is essential for organizations in heavily regulated industries.

Data Usage:

- **Problem:** Inadequate management of data usage leads to issues such as data silos, inefficient access controls, and limited availability of data for analysis.
- **Solution:** Utilize data analysis techniques like Pareto Analysis to identify patterns and trends contributing to data quality problems or silos.
- **Example:** A mutual fund house analyses records to identify inconsistencies in portfolio management coding, leading to discrepancies in client investments. Here, root cause analysis reveals insufficient training for staff on coding practices.

Archiving:

- **Problem:** Failure to properly archive data results in unnecessary storage costs.
- **Solution:** Establish policies for data retention and archiving. Further, prioritize active data for storage on high-performance media while storing the remaining in archives for cost-effective long-term storage.

Data Destruction:

- **Problem:** Lack of a systematic approach to data destruction can lead to data breaches, privacy violations, and non-compliance with regulatory requirements
- **Solution:** Implement procedures for secure and compliant data destruction, including proper verification of data erasure.

Approaches for Enhancing Data Quality in Legacy System

Data quality in legacy systems refers to the accuracy, completeness, consistency, and reliability of data stored in outdated technologies. Due to the age and complexity of these systems, data quality issues are common and can lead to inaccuracies.

Solutions for Enhancing Data Quality in Legacy Systems [2]:

- **Data Extraction and Backups:** Before any migration process begins, it is important to safely extract data from the legacy systems and create backups.
- **Data Mapping and Standardization:** This involves identifying data fields, relationships, and dependencies to ensure compatibility with the new system. It helps in standardizing data formats, and conventions, and improves consistency.
- **Data Cleaning Optimization:** The use of data cleansing tools and techniques can remove irrelevant or outdated data stored in the database affecting overall quality over a long period.
- **Validation and Testing:** In legacy systems, validate and test the migrated data to ensure the accuracy, completeness, and consistency in the new system. The implementation of end-to-end testing identifies and resolves discrepancies and errors.
- **Backup and Recovery Planning:** Develop recovery plans with backup schedules, disaster recovery protocols, and offsite storage solutions to mitigate risks and ensure data resilience in the event of system failures.

Enhancing Data Quality Risk Management**Strategies to Improve Data Quality for Effective Risk Management**

The essential strategies and techniques which aim to improve data quality for effective risk management are:



- **Data Standardization:** Implement data profiling tools such as Talent Data Quality, IBM InfoSphere Information Analyzer etc., to analyse data structure.
- **Data Validation:** The use of automated validation tools such as Trifacta to identify errors and inconsistencies in data can be beneficial. Data quality monitoring tools such as SAS Data Quality and Informatica Data Quality can continuously assess data quality.
- **Data Integration:** The use of data mapping and transformation techniques can reconcile data from different sources and formats. The use of data virtualization platforms such as Cisco Data Virtualization, Denodo etc., can be used to integrate data in real-time without physical data movement.
- **Data Governance:** Data governance software platforms such as Collibra and Informatica Axon to automate data governance processes. It is important to define data stewardship roles and responsibilities to oversee data quality initiatives.

Role of High-Quality Data in Predictive Analytics and Decision-Making Process

High-quality data plays a fundamental role in predictive analytics and decision-making processes, embodying the adage "*Garbage In, Garbage Out*" (*GIGO*) [6]. In predictive analytics, the accuracy and reliability of insights are directly proportional to the quality of input data. If the data used to train predictive models is inaccurate, incomplete, or inconsistent, the resulting predictions will be flawed and unreliable.

The concept of *GIGO* refers to poor data quality due to inaccuracies and incompleteness. For instance, if a customer information in bank shows underestimated or overestimated demand, it can lead to inefficient resource allocation depicted as "garbage in". Now, if a machine learning algorithm is trained using the 'garbage in', it may produce biased predictions and initiate 'garbage out' to create flawed predictions from the flawed data input.

Similarly, in decision-making processes, the outcomes derived from data-driven insights are only as good as the data upon which they are based. High-quality data ensures that decisions are well-informed, reliable, and aligned with organizational objectives.

Moreover, high-quality data ensures confidence among decision-makers. It allows them to make informed choices confidently. Ensuring the feeding of high-quality data can mitigate the risks associated with *GIGO* and enable organizations to harness their full potential.

Case Study

Wells Fargo's fake account scandal was one of the severe crises in 2016 when it was discovered that employees had engaged in multiple fraudulent practices to meet sales targets. Wells Fargo, one of the largest banks in the US, instituted aggressive cross-selling goals for its employees [7]. This led to creating a toxic sales culture where employees felt pressured to meet unrealistic demands at any cost.

Several challenges emerged from the company's operations:

- **Unethical Practices:** Employees involved in opening millions of unauthorized savings accounts for customers without their knowledge or consent.
- **Data Integrity Issues:** The fraudulent practices highlighted broader data integrity issues within Wells Fargo's systems. Customer information was inaccurately represented or manipulated which led to discrepancies [7].
- **Lack of Oversight:** Their system lacked centralized oversight which contributed to a culture of impunity.

Responses, outcomes, and lessons learned:

The senior executive of Wells Fargo resigned in the wake of the scandal. The bank faced investigations from regulatory agencies. Wells Fargo agreed to pay \$3 billion in settlement to resolve civil and criminal probes into the fraudulent practices [8]. The bank prioritized data governance initiatives to improve data integrity and transparency:

- The scandal tarnished the company's reputation and eroded trust among customers. Thus, organizations must prioritize data quality to uphold their reputation and ensure customer loyalty.



- The fraudulent practices at Wells Fargo resulted in mismanagement of customer accounts and financial discrepancies. Thus, ensuring data accuracy and consistency is essential for optimizing organizational performance.

Conclusion

Foundational data quality is paramount for financial integrity, as it serves as the basis of effective risk management, regulatory compliance, and decision-making processes.

To address data quality challenges and mitigate risks, financial institutions must prioritize the development of robust data governance frameworks, implement regular audits and monitoring processes, and invest in data quality improvement initiatives.

The case of Wells Fargo's fake account scandal discovers the critical importance of data quality in the financial industry. The case study scandal revealed the detrimental consequences that can arise from compromised data integrity.

By adopting the outlined strategies in the whitepaper, financial organizations can enhance operational efficiency, mitigate risks, and safeguard their reputation and trustworthiness in the eyes of customers and stakeholders.

Ultimately, prioritizing data quality is *not just a "best practice, but it is a strategic imperative"* for long-term success and sustainability in the ever-evolving financial landscape

References

- [1]. Gartner, "How to Create a Business Case for Data Quality Improvement," *gartner.com*, Jun. 19, 2018. <https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement> (accessed Oct. 28, 2021)
- [2]. Data Ladder, "how legacy systems and bad data quality hinders a digital transformation plan," *JURNAL TEKNIK HIDRAULIK*, vol. 10, no. 1, p. 1, Jul. 2020, doi: <https://doi.org/10.32679/jth.v10i1.62>. (Accessed Oct. 28, 2021)
- [3]. Deloitte, "How does the Dutch Mid-sized Banking industry manage data quality," 2021. Available: <https://www2.deloitte.com/nl/nl/pages/financial-services/articles/how-does-the-dutch-mid-sized-banking-industry-manage-data-quality.html> (accessed Nov. 22, 2021)
- [4]. D. Loshin, "Chapter 5 - Data Quality and MDM," *ScienceDirect*, Jan. 01, 2009. <https://www.sciencedirect.com/science/article/abs/pii/B9780123742254000059> (accessed Nov. 21, 2021).
- [5]. D. Plotkin, "Chapter 3 - Data Stewardship Roles and Responsibilities," *ScienceDirect*, Jan. 01, 2014. <https://www.sciencedirect.com/science/article/abs/pii/B9780124103894000039> (accessed Nov. 21, 2021).
- [6]. Deloitte, "Too Big to Ignore Big Data 2.0," 2013. Available: <https://www2.deloitte.com/content/dam/insights/us/collections/Issue-12-MASTER.pdf> (accessed Nov. 28, 2021)
- [7]. Suraj Srinivasan & Jonah S. Goldberg, "Recovering Trust After Corporate Misconduct at Wells Fargo," *Harvard Business School*, Jun. 2020. <https://www.hbs.edu/faculty/Pages/item.aspx?num=58343> (accessed Nov. 28, 2021)
- [8]. J. Kelly, "Wells Fargo Forced To Pay \$3 Billion For The Bank's Fake Account Scandal," *Forbes*, 2020. <https://www.forbes.com/sites/jackkelly/2020/02/24/wells-fargo-forced-to-pay-3-billion-for-the-banks-fake-account-scandal/?sh=5dae075b42d2> (accessed Nov. 28, 2021)

