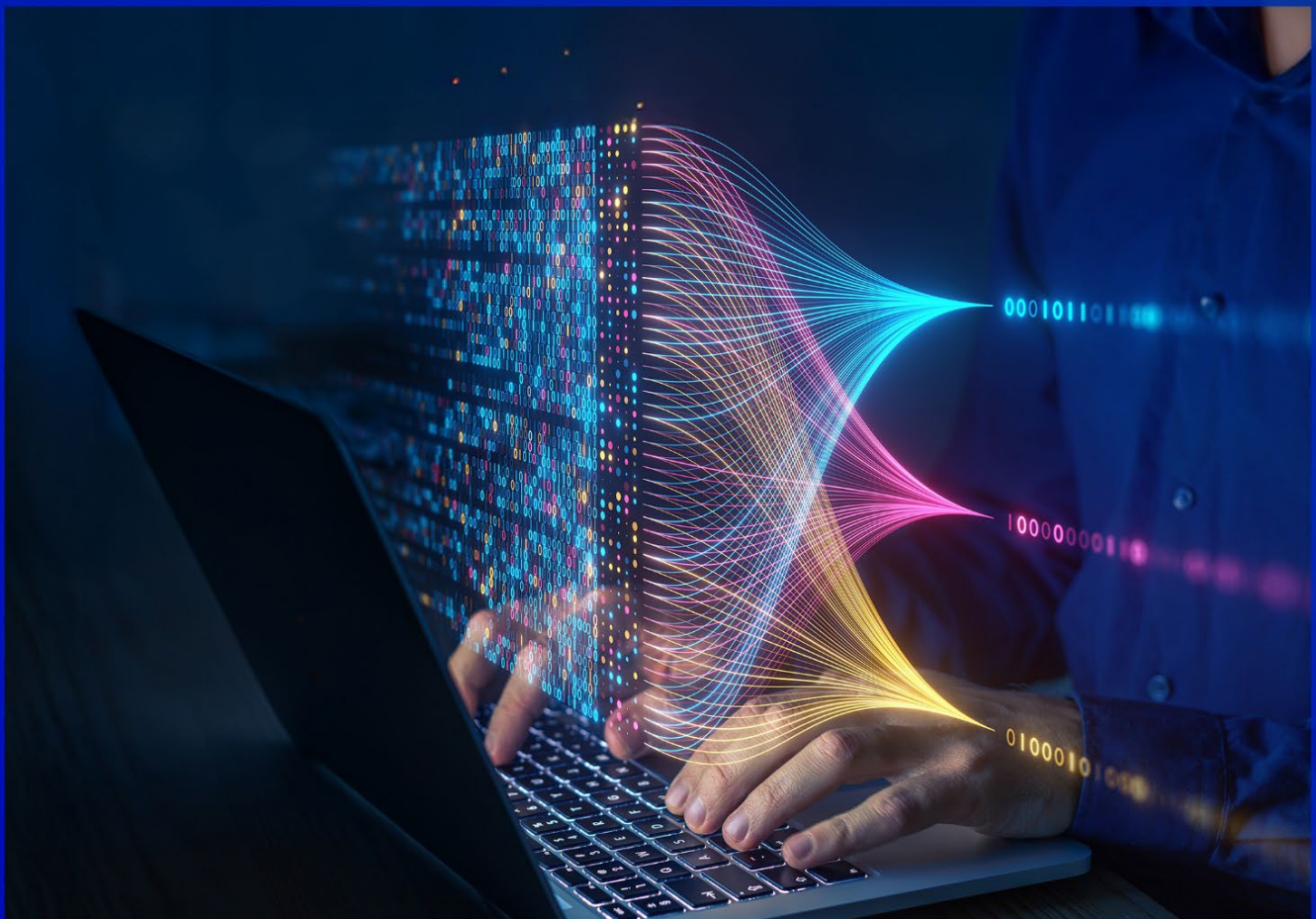




# The DQ Trifecta

Data Quality for AI Success: Emerging Imperatives



A Perspective Report

October 2025

[home.kpmg/in](https://home.kpmg/in)

KPMG. Make the Difference.

# Table of Contents

<b>Foreword</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
<b>2 Emerging imperatives on DQ for AI</b>	<b>5</b>
a) Business-Led, Process-Fed	
b) 360° Integration within Data Teams	
c) Observe, Detect, Act: The Power of Observability	
d) Semantic Intelligence through metadata quality	
e) The Full Picture: beyond Masters	
f) Agnostic by Principle, Flexible by Design	
<b>3 A Strategic Lens for AI Maturity</b>	<b>10</b>
<b>Authors</b>	<b>11</b>
<b>Key Contacts</b>	<b>12</b>

# Foreword

We see it in headlines across all industries – billions are poured into AI startups, and executives and influencers promote AI as the ‘silver bullet’ for productivity, creativity and cost efficiency. But the hard reality is your AI programs are only as good as the data that flows into them.

If data is the critical fuel, then an effective data quality methodology must be the discipline required to ensure trustworthy data. But no surprise, yesterday’s data quality methodologies don’t quite make the grade in today’s overwhelmingly data-driven digital world. If we believe that data is the lifeblood of decision-making, innovation, and trust, what must we do differently in our data quality processes? What new and innovative approaches must we take to pin down the challenges of achieving trusted data for our AI programs?

The ‘DQ Trifecta – Data Quality for AI Success’ cracks the permafrost of the old data quality methodologies to present a more holistic, complete approach to this challenge. The paper does more than suggest changes to existing data quality approaches. It suggests a cultural shift in how organizations should view the curation of arguably one of the most important corporate assets – their data.

The paper compares the traditional dual approach of proactive data quality at source and reactive data quality at consumption to a new, holistic view of data quality observability across the entire data supply chain. To address the demands of clean data to feed AI programs, data quality must start

with those who understand its impact the most, the business, then bring together people, process and technology, executing proactive data quality checks across the entire landscape. It takes a village – all stakeholders impact the quality of the information asset.

Finally, DQ Trifecta emphasises the importance of how data is defined. Metadata is no longer a technical definition but is critical in defining the semantic meaning of data, which is at the heart of trusted data. And not just traditional master data, but all data in the business process, from master data to transaction data to analytics.

DQ Trifecta will open up a new view of approaching data quality. It’s not your grandfather’s data quality anymore. It’s data quality for the age of AI.



**John Bottega**

**President**  
EDM Association

## 1) Introduction

### 2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

360° Integration within Data Teams

Observe, Detect, Act: The Power of Observability

Semantic Intelligence through metadata quality

The Full Picture: beyond Masters

Agnostic by Principle, Flexible by Design

### 3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

### 4) Authors

# 1 Introduction

Data quality issues are among the leading causes of biased AI models and poor performance outcomes. For AI to deliver on its transformative promise, organisations must first build a solid data foundation. Clean, accurate, and reliable data is the fuel that powers effective, trustworthy AI.

Neglecting data quality before model build undermines the entire analytics strategy. Poor-quality data erodes trust in insights, leading to flawed decisions and reputational risk. It inflates costs through rework, delays, and increased maintenance, reducing ROI on AI initiatives. Strategically, it creates a fragile foundation—models become less scalable, harder to govern, and non-compliant with regulatory standards. Moreover, it limits the organisation's ability to operationalize advanced analytics, as unreliable outputs discourage adoption by business stakeholders. Ensuring robust data quality upfront is not just a technical necessity; it's a strategic enabler

for sustainable, trustworthy, and value-driven AI transformation.

This perspective report addresses the persistent 'garbage-in, garbage-out' challenge by highlighting emerging trends in data quality across industries such as finance, life sciences, healthcare, and consumer markets. Drawing from experiential insights, it introduces a structured approach for AI enablement built around three key pillars:

- **Strategic Anchors:** Aligning with purpose and direction of data quality initiatives
- **Operational Levers:** Answers 'How' to operationalize / implement the DQ programs
- **Integration Points:** Defines 'Where' are my critical touchpoint for maximizing DQ impact

By shifting the focus from reactive data fixes to proactive, embracing process-driven quality improvements, organisations can unlock scalable, bias-free AI solutions that drive meaningful business outcomes.

“

Ensuring high-quality and sufficient data is essential not just for successful AI implementation, but also for creating real value within an organisation. Just as data observability helps move us from a reactive approach to a more proactive one focused on building trust, having rich, high-quality data can turn AI from a mere technical tool into a powerful driver of meaningful change. To achieve this, organisations need to establish strong data standards and privacy protections throughout the entire data lifecycle, ensuring that their AI solutions lead to reliable, relevant, and profitable results.

”

**Srilatha Kulkarni**

Director, Data Analytics  
Merck KGaA

1) Introduction

2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

360° Integration within Data Teams

Observe, Detect, Act: The Power of Observability

Semantic Intelligence through metadata quality

The Full Picture: beyond Masters

Agnostic by Principle, Flexible by Design

3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

4) Authors

## 2 Emerging imperatives on DQ for AI

### a) Business-Led, Process-Fed

A significant shift on the ownership of data quality efforts is underway—**business teams are increasingly leading data quality efforts**. This change stems from the realisation that data is a strategic asset, and its value is best understood by those who use it to drive decisions. When business units take ownership of data quality, they align it directly with operational goals, customer outcomes, and performance metrics. This approach ensures that data is not only technically correct but also contextually relevant and actionable. One of the key advantages of this approach is that **Business-led** data quality fosters **greater accountability**, as teams responsible for outcomes are also responsible for the data that informs them.

Moreover, Data quality improvement projects are increasingly shifting focus from merely cleansing data at the source to **enhancing data quality within key business processes**.

This approach recognizes that poor data often originates not just from external systems but from flawed internal workflows, inconsistent data handling, and lack of governance during operational execution. By embedding data quality checks and controls directly into core processes (such as KYC, customer onboarding, supply chain management, or financial reporting) organisations ensure data remains accurate, complete, and timely throughout its lifecycle. This proactive strategy **reduces downstream issues, improves decision-making, and enhances compliance**.

In contrast, source-level improvements, while essential, often treat symptoms rather than root causes. They may correct data after it's already been misused or misinterpreted. Process-centric improvements, however, foster accountability, align data standards with business objectives, and create a culture of data stewardship.

“

In my experience, data quality programs fail when framed as IT exercises. They succeed when business leaders hardwire definitions into workflows, tie quality to revenue and risk, and treat exceptions like operational defects. That shift turns governance into advantage.

**Chiranjoy Chowdhuri**  
Head of AI and Data  
Dr. Reddy's Laboratories

”

1) Introduction

2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

**360° Integration within Data Teams**

Observe, Detect, Act: The Power of Observability

Semantic Intelligence through metadata quality

The Full Picture: beyond Masters

Agnostic by Principle, Flexible by Design

3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

4) Authors

## 2 Emerging imperatives on DQ for AI

### b) 360° Integration within Data Teams

As enterprises accelerate their journey toward AI-powered decision-making, a pivotal shift is underway: **data quality is no longer the sole domain of governance teams—it's becoming a shared responsibility across the entire data organisation.** The emerging trend of **360° integration of data quality** within data engineering, data science, and analytics teams is redefining how businesses build trust in their data. It suggests a coming together or alignment of data quality practices and technicalities across these teams to ensure high-quality data and reliable insights.

This integration spans **people, process, and technology.** On the people front, data engineers are embedding quality checks directly into pipelines, while data scientists are validating input datasets for model reliability. From a process standpoint, agile DQ practices—such as continuous profiling, anomaly detection, and feedback loops—are being woven into daily

workflows. Technologically, organisations are leveraging existing tools—ETL platforms, orchestration frameworks, and even dashboarding solutions—to implement DQ without waiting for centralized tooling.

*AI's promise of rapid market entry is realized only when underpinned by robust data quality!* High-quality data improves model training, performance, and interpretability, reducing bias and drift while enhancing reliability and trust in AI outputs. Integrating Data Quality within the larger data landscape ensures reliable, timely, and accurate data, reducing operational risks and enabling faster, data-driven decision-making across business functions. In essence, the future of data quality is collaborative, embedded, and continuous. Organisations that embrace this integrated mindset can not only unlock the full potential of their data but also lead the way in AI-driven business success.

“

AI is as good as the data it learns from, hence a governance first data strategy that embeds strong data quality across the data lifecycle including data engineering, data visualization and data science teams is important to drive success from AI initiatives’.

**Shireen Ali**  
Head - ESG Data and Capability Services  
HSBC

”

1) Introduction

2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

360° Integration within Data Teams

**Observe, Detect, Act: The Power of Observability**

Semantic Intelligence through metadata quality

The Full Picture: beyond Masters

Agnostic by Principle, Flexible by Design

3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

4) Authors

## 2 Emerging imperatives on DQ for AI

### c) Observe, Detect, Act: The Power of Observability

Traditionally, organisations have relied on a dual approach to Data Quality: **Proactive DQ** at source systems using generic validation rules (e.g., null checks, format checks, type enforcement), and **Reactive DQ** at the consumption layer through custom SQL scripts and ad-hoc solutions. While effective in isolated scenarios, this fragmented model often leads to delayed issue detection, redundant rework, and limited traceability across the data lifecycle.

The emerging paradigm of **DQ Observability** is transforming this landscape. By embedding DQ checks across the entire data pipeline—coupled with **data lineage repository creation and awareness**—teams can now proactively monitor, trace, and resolve quality issues in real time. This also provides opportunity to consolidate and rationalize DQ

rules and provide transparency in DQ scoring across the lifecycle for all stakeholders. This shift enables **end-to-end visibility**, allowing data engineers and scientists to pinpoint root causes faster, assess downstream impacts, and ensure trust in data products.

Observability platforms integrate seamlessly with orchestration tools and metadata catalogs, enabling scalable, automated, and context-aware DQ enforcement.

This evolution not only enhances operational efficiency but also strengthens the foundation for reliable AI/ML models and analytics for accurate, unbiased and reliable AI responses and improves contextual AI explainability and traceability, ensuring that business decisions are powered by high-integrity data from ingestion to insight.

“

Data observability isn't just an evolution of data quality; it's the shift from firefighting bad data to proactively engineering. For our clients, this means replacing reactive damage control with the foresight needed, revealing the full context of data health across the data flow, for truly resilient and profitable data-driven decisions.

**Sankara Subramanian**  
Partner & COO, Digital Lighthouse  
KPMG in India

”

1) Introduction

2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

360° Integration within Data Teams

Observe, Detect, Act: The Power of Observability

**Semantic Intelligence through metadata quality**

The Full Picture: beyond Masters

Agnostic by Principle, Flexible by Design

3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

4) Authors

## 2 Emerging imperatives on DQ for AI

### d) Semantic Intelligence through metadata quality

As data ecosystems scale, the spotlight is shifting from raw data validation to **metadata quality as a strategic enabler of semantic intelligence**. Metadata—describing the structure, lineage, context, and meaning of data—is no longer just a technical artifact; it's the foundation for intelligent data operations. By implementing robust metadata quality strategies, organisations can ensure completeness, consistency, and semantic alignment across data assets.

This approach enables systems to interpret *what data means*, not just *what it is*. Through **enriched metadata**, teams can automate schema validations, detect semantic drift, and enforce contextual rules across pipelines. **Metadata Quality** involves assessing completeness, consistency, and accuracy of

metadata fields like schema definitions, data types, lineage, and business descriptions. Organisations need to enrich and augment metadata by *integrating semantic tags, usage context, and lineage information*, enabling intelligent search, traceability, and rule enforcement across pipelines; supported by catalogs, lineage tools, and observability platforms. This creates an unified, intelligent data fabric that adapts to change and scales with business needs.

**High-quality metadata** improves model transparency, feature relevance, and bias mitigation—critical for trustworthy AI. In turn, **semantic intelligence** guides AI Models/LLMs for enterprise-relevant grounded responses and transforms data into a context-rich asset for better decision-making, innovation and growth.

“

As the revolution in AI unfolds, the imperative to ensure that the data content feeding it, is of the highest governance and data quality standards has become vital. Who would have known it, the application of high-grade Data Governance controls and procedures is now at the forefront of driving confident AI based product solutions.

**Chris Skinner**  
Data Governance Director  
LSEG (London Stock Exchange Group)

”

1) Introduction

2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

360° Integration within Data Teams

Observe, Detect, Act: The Power of Observability

Semantic Intelligence through metadata quality

**The Full Picture: beyond Masters**

**Agnostic by Principle, Flexible by Design**

3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

4) Authors

## 2 Emerging imperatives on DQ for AI

### e) The Full Picture: beyond Masters

In today's data-driven enterprises, the **traditional focus on master data quality**—customer, product, vendor—has evolved. Forward-thinking organisations are now extending their data quality initiatives to **transactional and analytical data**, recognizing that business success in the age of AI hinges on the integrity of *all* data. This shift marks a critical inflection point.

**Transactional data** fuels operational decisions, while analytical data powers strategic insights. When these datasets are riddled with inconsistencies, duplicates, or missing data, the consequences ripple across forecasting, prediction, personalization, compliance, and automation.

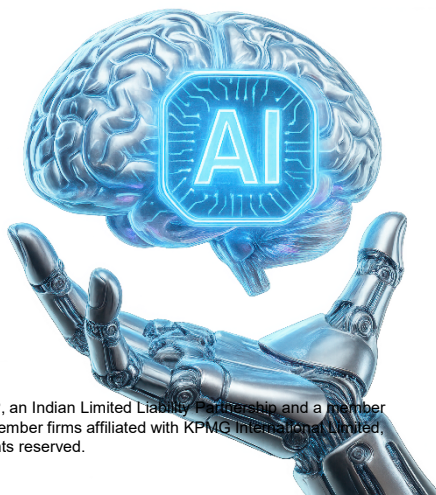
AI is only as strong as the *weakest link in your data chain*. Master data sets the foundation, but transactional and analytics data fuels the underlying intelligence. Both must be governed with equal rigor to realize **AI's full business value**. From use-cases ranging from fraud detection, personalization, supply chain optimisation, risk assessment and customer insights; transactional data is the bedrock of a robust AI engine. In summary, ensuring high-quality data across the entire ecosystem is no longer optional—it's a **strategic imperative**. The result is more trustworthy insights, faster innovation cycles, and resilient AI systems that drive competitive advantage.

### f) Agnostic by Principle, Flexible by Design

Medium-Large Enterprises often tend to fall into the trap of obsessing over the 'perfect' data management platform or the most sophisticated data quality tool. But the emerging trend among agile enterprises is **refreshingly pragmatic: start where you are, with what you have**. Whether it's an ETL pipeline, a dashboarding tool, or a simple scripting framework!

This tool-agnostic approach democratizes data quality. It empowers domain teams to take ownership of their data without waiting for centralized tooling rollouts. More importantly, it aligns perfectly with modern architectural paradigms like **data mesh** and **data fabric**, where decentralized data stewardship and interoperability are key. In a data mesh, quality becomes a shared responsibility across federated domains. Teams use the tools native to their workflows to enforce standards, validate schemas, and monitor anomalies. Similarly, data fabric architectures thrive on metadata-driven automation and real-time observability—both of which benefit from embedded DQ practices across diverse platforms.

Why does this matter? Because AI doesn't just need data—it needs *trusted* data. Models trained on inconsistent or low-quality inputs will inevitably produce flawed outputs, eroding business confidence and value. Tool-agnostic DQ approach/mindset ensures that every data product, regardless of origin or format, is AI-ready. Further, AI can automate the fundamentals of data quality, quietly establishing the **neutral ground** on which every effective AI must/can stand. In essence, the future of data quality is not about chasing the next shiny platform—it's about embedding trust into every layer of your data ecosystem. Organisations that embrace this mindset will unlock faster innovation, scalable AI adoption, and sustainable business success.



1) Introduction

2) Emerging imperatives on DQ for AI

Business-Led, Process-Fed

360° Integration within Data Teams

Observe, Detect, Act: The Power of Observability

Semantic Intelligence through metadata quality

The Full Picture: beyond Masters

Agnostic by Principle, Flexible by Design

3) The Data Quality Trifecta: A Strategic Lens for AI Maturity

4) Authors

# 3 The Data Quality Trifecta: A Strategic Lens for AI Maturity

In an era where Data-Cloud-AI is the backbone of innovation and decision-making, the **Data Quality Trifecta** emerges as a powerful approach to navigate the complexity of modern data ecosystems and evolving business and compliance needs.

*Synthesized from above six key trends shaping the present and future of data quality for AI, this approach offers a structured way to assess and elevate AI maturity across organisations.*

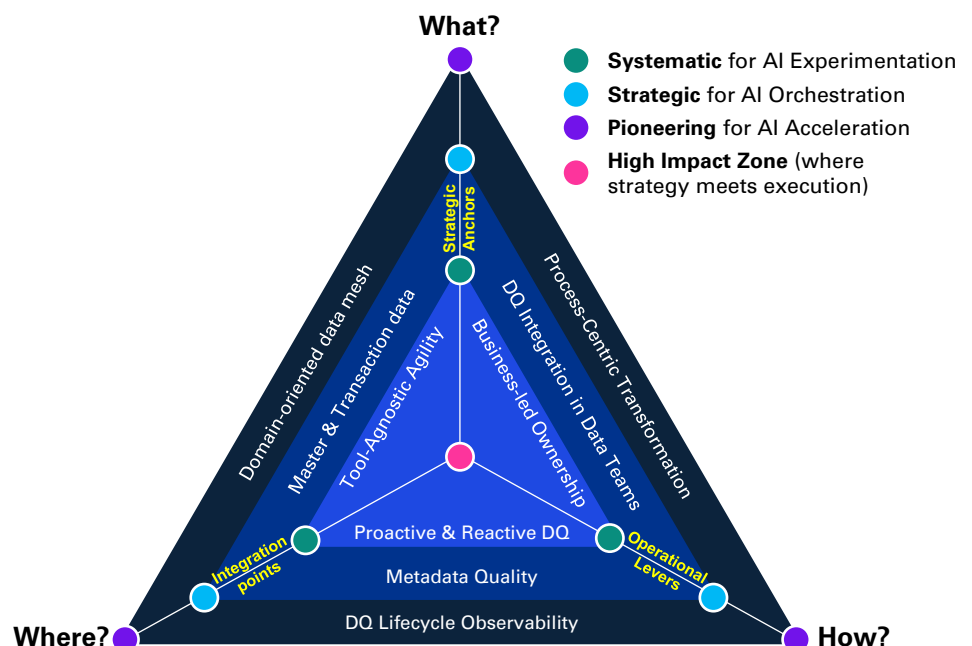
At its core, the Trifecta is built on **three dimensions**:

- 1. Strategic Anchors (What?)** – These define the *purpose and direction* of data quality initiatives. Components emphasize that data quality must be driven by business needs and outcomes and embedded within operational IT, Data and Business processes.
- 2. Operational Levers (How?)** – These are the *execution mechanism* to implement and sustain data quality to bring the data quality strategy to life. This encompasses aspects enabling organisations to prevent issues before they occur and respond swiftly when anomalies arise. Central to this is lineage-based observability, which brings

real-time visibility into data flows, allowing teams to detect, diagnose, and act on quality issues dynamically. Additionally, a strong focus on quality of metadata to ensure that data is not only clean but also contextually rich—making it more usable, trustworthy, and AI-ready.

- 3. Integration Points (Where?)** – These define the *architectural and tooling touchpoints* where quality aspects must be embedded within the Organisation for purpose, flexibility and scale. This involves components advocating for embedding DQ across existing tool stacks—ETL, dashboards, or scripting—without waiting for a monolithic platform; and slowly converging these components into a single data and technology mesh

The Trifecta’s concurrent triangle placement of nine factors enables organisations to map their **AI maturity stages**—from **Experimentation** to **Orchestration** to **Acceleration**. It’s not just an approach; it’s a strategic compass guiding enterprises toward scalable, resilient, and AI-ready data ecosystems.



---

1) Introduction

---

2) Emerging imperatives  
on DQ for AI

Business-Led, Process-Fed

360° Integration within  
Data Teams

Observe, Detect, Act: The  
Power of Observability

Semantic Intelligence  
through metadata quality

The Full Picture: beyond  
Masters

Agnostic by Principle,  
Flexible by Design

---

3) The Data Quality  
Trifecta: A Strategic  
Lens for AI Maturity

---

**4) Authors**

---

## 4 Authors

### **Ganapathy Subramanian**

Partner  
Digital Lighthouse  
KPMG in India

### **Rahul Sharma**

Director  
Digital Lighthouse  
KPMG in India

### **Soumya Rajan**

Associate Director  
Digital Lighthouse  
KPMG in India

# KPMG in India contacts:

**Akhilesh Tuteja**  
Partner and Head  
Client and Markets  
E: [atuteja@kpmg.com](mailto:atuteja@kpmg.com)

**Sankara Subramanian**  
Partner & COO  
Digital Lighthouse  
E: [tsubramanian@kpmg.com](mailto:tsubramanian@kpmg.com)

**Kamesh Srinivasan**  
Partner  
Digital Lighthouse  
E: [kamesh@kpmg.com](mailto:kamesh@kpmg.com)

[kpmg.com/in](https://kpmg.com/in)



Access our latest insights  
on KPMG Insights Edge

Follow us on:  
[kpmg.com/in/socialmedia](https://kpmg.com/in/socialmedia)



The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.

The views and opinions expressed herein are those of the quoted third parties and do not necessarily represent the views and opinions of KPMG in India.

KPMG Assurance and Consulting Services LLP, Lodha Excelus, Apollo Mills Compound, NM Joshi Marg, Mahalaxmi, Mumbai-400 011  
Phone: +91 22 3989 6000, Fax: +91 22 3983 6000.

© 2025 KPMG Assurance and Consulting Services LLP, an Indian Limited Liability Partnership and a member firm of the KPMG global organization of independent member firms affiliated with KPMG International Limited, a private English company limited by guarantee. All rights reserved.

The KPMG name and logo are trademarks used under license by the independent member firms of the KPMG global organization.

This document is for e-communication only. (TL\_1025\_AC)