Hitachi Solutions

Brochure

Beyond "Garbage In, Garbage Out"

Data Quality as the Driving Force of Digital Value in the AI Age

HITACHI

Introduction: Data Quality – The Bedrock of Digital Value Creation

AI and digital transformation are reshaping company operations, affecting people and business processes, alongside the technology. Data quality should be prioritised at the Board level along with AI initiatives, and CEOs must act now to embrace these changes. Any M&A activity as part of a company value creation plan will only increase the need for CEO's and their board to embrace what AI can enable, although this does introduce even more data complexity.

High-quality data is the foundation of any successful digital transformation. In fact, "data is the cornerstone" of modern, AI-driven business change, enabling organisations to build strong data foundations and drive value from their digital initiatives. When data is accurate, consistent, and timely, it empowers better decisions, streamlined processes, and innovative services that create real business value. Conversely, poor data quality can derail even the most ambitious transformation efforts. Studies show that 70–85% of AI projects fail[1], with poor data quality cited as a primary cause (alongside misaligned goals and unrealistic expectations). Business leaders have learned the hard way that you cannot extract meaningful insights or ROI from "garbage" data – the old adage "garbage in, garbage out" is even more applicable in the AI age.

Beyond lost project investments, low data quality carries heavy ongoing costs. It quietly undermines operational efficiency and trust across the enterprise. If executives and frontline employees don't trust the numbers, they will find workarounds, often reverting to manual spreadsheets and driving "grey IT". This behaviour leads to fragmented "multiple versions of the truth." Trust in data is usually the factor that makes or breaks a new data initiative – without it, people create their own off-system analyses, perpetuating errors and inconsistency. In short, poor data quality isn't just an IT issue; it's a business value issue. It can mean the difference between a data-driven organisation and one flying blind on gut feeling.

High-Quality Data Enables…	Poor-Quality Data Causes…
Confident decision-making based on facts and trends	Doubt and second-guessing of reports and KPIs
Successful AI/analytics projects that deliver ROI	Failed AI/analytics initiatives; models learning "wrong" patterns, or providing incorrect advice
Unified customer views and insights	Fragmented, duplicate records ("multiple truths")
Regulatory compliance (e.g. in finance)	Compliance breaches, restatements, and penalties
A deep understanding of business processes and performance	Limited real knowledge and unseen process errors, requiring fire-fighting, and rework to fix

Table: Consequences of Data Quality – good vs. bad data impacts across a business.

[1] Quelle: Ryseff, J.; De Bruhl, B.; and Newberry, S.J. (2024): The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed

"Fit for Purpose" Data Quality – Not All Use Cases Need the Same Level

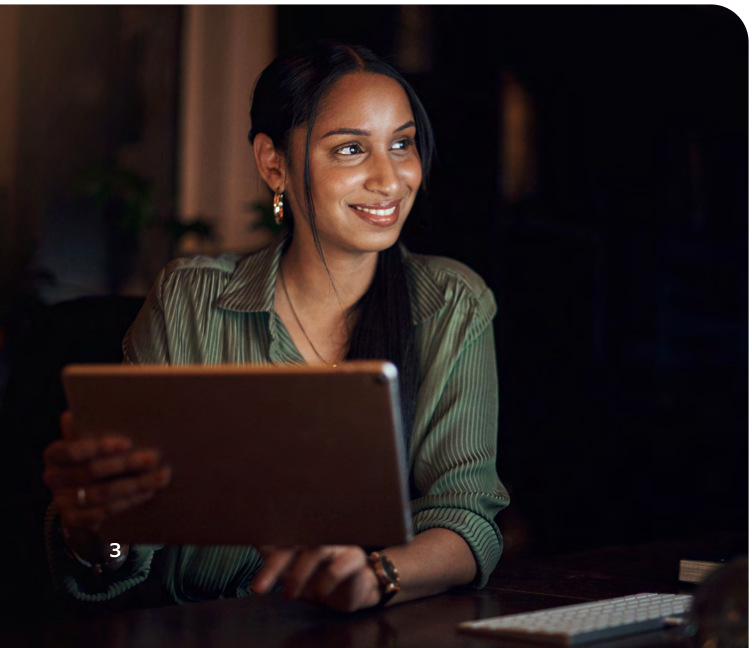
While data quality is critical, the required level of quality can differ by business function and use case. In a balanced digital transformation, companies calibrate their data quality efforts to what's "fit for purpose." A common pitfall is assuming all data must meet an identical gold standard. For example, data used for financial reporting and data for AI-driven sales order automation have distinct requirements and tolerances. An effective data strategy aligns quality investments with the value and risk at hand.

Financial Reporting – Zero Tolerance for Errors

Use Case: Official financial statements, regulatory compliance, board reports.

Quality Needed: Very High. Data must be complete, accurate, and consistent. Even minor errors can lead to incorrect earnings reports or compliance violations.

Why: Finance operates under strict regulations and scrutiny. There is virtually no forgiveness for bad data – "excellence in data quality and compliance" is mandatory. Trustworthy numbers are the basis of investor confidence and strategic decisions.



Generative AI for Sales Order Automation – Agile over Perfect

Use Case: AI-driven extraction of information required for sales order processing from multiple document types and channels.

Quality Needed: Moderate. Data to be extracted (e.g. the amount, company names, order details) should be reasonably clean, but it doesn't need to be pristine. Some noise or gaps are acceptable because AI solutions are capable of filling in the gaps and checking with reference data.

Why: Generative AI models (like large language models) can often tolerate imperfection; they learn patterns and context from vast amounts of data. For example, a few typos or out-of-date entries in thousands of service tickets won't derail an AI in extracting the required information. The priority here is having ample data coverage and context. Minor errors might slightly reduce output quality but usually won't pose serious risk as long as humans review important results. It's more important that the AI has broad exposure to the knowledge domain than 100% accuracy on each input.

Different functions = different data quality benchmarks. An enterprise should identify where "good enough" data is truly sufficient versus where only near-perfection will do. For instance, a marketing department doing trend analysis on social media data might accept lower precision – sentiment scores might be roughly right, and that's okay for campaign tweaks. In contrast, an inventory management system needs highly reliable product data (sizes, quantities, locations) to avoid stock-outs or overstock. The concept of "fit for purpose" means invest the most in quality where it has the most impact. For high-stakes, high-risk data like financials or patient records, you implement rigorous validation and governance. For less critical analyses, you can focus on completeness and let a few inaccuracies slide. This balance ensures effort spent on data quality is itself creating value – mitigating real risks or enabling important use cases – and not just polishing data for its own sake.

Challenges in Merging Disparate Data Sources (and Why Quality Suffers)

Digital transformation often involves consolidating data from many sources – especially after mergers and acquisitions or across different business units. When a company merges several businesses or systems, it inherits a patchwork of data silos and standards. This presents a major challenge: how to unify the data so it's consistent, accurate, and useful for enterprise-wide insights. In practice, merging data sets without a plan can introduce quality issues or even create new ones.

Some common pain points include:

1

Incompatible Data Definitions

Different systems often define the same data in different ways. For example, one system might treat a customer as a contract, while another records customers by individual names. After a merger, this leads to records that overlap or don't match up. If these definitions aren't made consistent, reports that combine this information will be confusing and unreliable.

2

Siloed and Redundant Data

Different systems often define the same data in different ways. For example, one system might treat a customer as a contract, while another records customers by individual names. After a merger, this leads to records that overlap or don't match up. If these definitions aren't made consistent, reports that combine this information will be confusing and unreliable.

3

Cascading Errors

If one source has poor quality data (e.g. outdated customer addresses or missing values), when you merge it with a clean source, the bad data can pollute the good. Determining which source is the "system of record" for each data element is crucial.

4

User Adoption & Trust Issues

A less obvious but critical challenge: after a messy data merger, business users may lose faith in the new system. If a CFO pulls a consolidated report and sees obvious errors (like the same client listed twice or totals that don't add up), confidence in the transformation falters. Users might then stick to their legacy spreadsheets and local databases "until IT figures it out."

Why do these issues matter so much?

Because digital and AI transformation is about seeing the big picture – holistic customer journeys, aggregated risk, end-to-end processes. If the data remains fragmented or unreliable after a merger, the organisation cannot reap the benefits of its transformation initiatives (whether that's a unified CRM, cross-selling strategies, or company-wide analytics). The value creation promised by the merger or system integration gets delayed or lost.

Ensuring Quality: Best Practices to Sustain High-Value Data

Maintaining data quality is an ongoing process, not a one-time fix. Leading organisations treat data as a key asset, with frameworks and practices to preserve and enhance its quality throughout the data lifecycle.

Here are some best practices and strategies executives should consider to keep data quality high and aligned with business needs:

Establish Strong Data Governance

Data governance is the umbrella of policies, processes, and roles that ensure data is managed as an enterprise asset. This includes setting data standards, defining data owners/stewards, and establishing quality controls. In practice, this means having governance bodies that routinely review data quality metrics, approve changes to data definitions, and champion data initiatives. When governance is in place, data quality issues are caught and resolved proactively rather than ad hoc.

Define Clear Data Quality Metrics and Targets

You can't improve what you don't measure[2]. Organisations should define what "quality" means in measurable terms – typical data quality dimensions include Accuracy, Completeness, Consistency, Timeliness, Uniqueness. For each critical dataset, set target thresholds (e.g., 98% of addresses geocodable, customer records 100% filled in core fields, data loaded within 24 hours of generation, etc.). Implement dashboards to monitor these. For instance, you might track completeness (% of records without null values in mandatory fields) or accuracy (error rates found in samples or via reconciliation with trusted sources). By quantifying data quality, you turn it into a managed KPI like any other business objective.

Implement Data Quality Tools and Processes

Leverage technology to automate quality checks. Modern data integration platforms such as Microsoft Fabric often have built-in data profiling and cleansing tools. These can automatically flag anomalies, standardise formats and even apply corrections using reference data. Setting up rules for validation (like "a date field must be a valid date in the past or present") helps catch errors at the point of entry or ingestion. Tools and processes can help to improve data quality, but teams using them need to understand the role they play in enabling organisations to become more data driven.

---

[2] Quote attributed to Peter Drucker, although it is worth pointing out that nothing becomes more important just because you can measure it. It simply becomes more measurable.

Adopt a “Single Source of Truth” Architecture

Simplify your data and AI landscape to reduce inconsistencies. Where possible, establish one primary system of record for each type of data (customer, product, employee, etc.), and have other systems consume data from there rather than maintaining separate copies. Cloud-based data lakes or warehouses are often used to consolidate data from many sources into a single analytics repository – but remember, aggregation alone doesn’t ensure quality. The processes feeding the repository should standardise and reconcile the data as discussed. The end goal is that whenever someone in the company asks, “What’s the current number of active customers?” – the answer comes from one authoritative source, not three different spreadsheets. Achieving that greatly improves confidence in data-driven discussions.

Integrate Quality into the Development Lifecycle

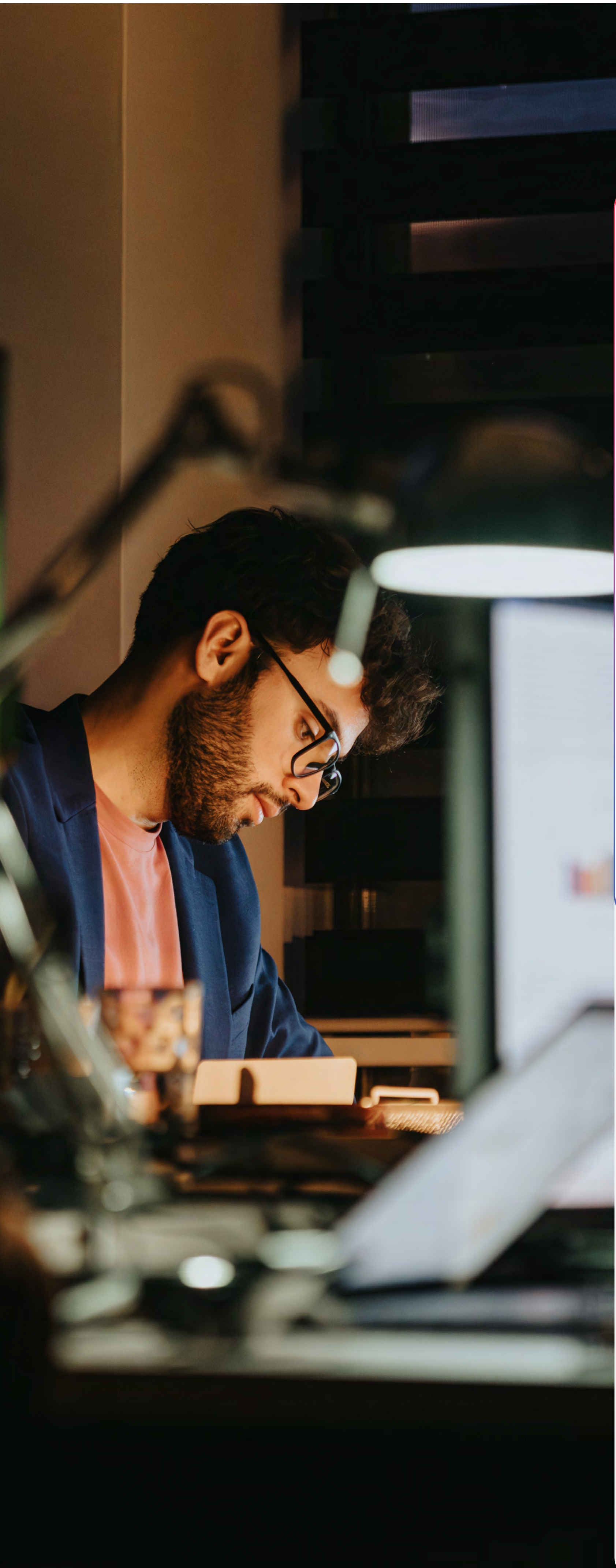
Whenever new digital solutions or transformations are being implemented (like migrating to a new ERP, or launching a new AI-enhanced service), make data quality a key part of the project plan. For example, include data quality criteria in User Acceptance Testing (UAT): users should sign off not just that the software works, but that the data in the new system is correct. Perform dry-run migrations and reconcile totals (e.g., total balances in old system vs new system) to verify nothing was lost or altered. This up-front diligence pays off by preventing bad data from tainting a go-live. It’s cheaper and easier to “get it right at the source” than to fix issues after deployment.

Cultivate a Data Quality Culture

Ultimately, technology and process can only go so far. It’s people who input data, manage data, and make decisions from data. Building a culture that values data quality is crucial. This means training employees on the importance of accurate data entry (and how it ties to business outcomes), giving feedback when errors are found, and perhaps incentivising good data practices.

By implementing these practices, data quality becomes an embedded part of your digital and AI operations, much like cybersecurity or operational excellence. The payoff is substantial: high-quality data lowers operational risks, builds customer and employee trust, improves decision confidence, and enables advanced analytics and AI to flourish. It creates a virtuous cycle – when people trust data, they use it more, which leads to better decisions and performance, which reinforces the value of maintaining data quality.

In summary, sustaining high data quality is an ongoing journey that combines governance, process, people, and technology. It is an integral part of digital and AI transformation success – ensuring that all the shiny new tools, algorithms and services rest on a solid bedrock of reliable data. Organisations that invest in this journey find that their digital transformation projects not only “go live” but deliver the expected value, because the insights and decisions drawn from the systems are based on truth. Those that neglect it will face the opposite: modern systems, but muddy outputs.



Conclusion

Digital and AI transformation isn't just about adopting cutting-edge technologies—it's about harnessing the power of technology to drive meaningful change. At the core of every successful initiative lies high-quality data: the quiet enabler of smarter decisions, sharper insights, and seamless operations. When data is trustworthy, technology delivers on its promise. When it's not, even the most advanced AI tools falter.

True transformation happens when organisations treat data as a strategic asset. It's not about perfection—it's about purpose. Reliable data empowers teams to innovate with confidence, navigate complexity, and unlock value across the enterprise. In a world defined by information, data quality is the backbone of resilience, agility, and competitive edge.

Contact

For further information please contact:

Robert Worsley
Head of Data, Analytics & AI
rworsley@hitachisolutions.com

Rob Evans
PE Alliance Director Europe
robevans@hitachisolutions.com

www.hitachi-solutions.co.uk