



Why data quality for agentic AI requires semantic context

High data quality for agentic AI is fundamentally different from traditional data cleansing because autonomous agents require rich, relational context to make dependable decisions. In an agentic architecture, clean rows and columns in isolated tables are not enough; agents need semantically connected data that explains how materials, suppliers, assets, and operational constraints relate across the enterprise. Without this deep contextual web, autonomous agents hallucinate, misroute purchase orders, or miscalculate inventory risks, turning automated workflows into operational liabilities.

To deploy agentic AI successfully, organizations must move away from querying static records and begin navigating interconnected relationships. When enterprise systems model physical entities as semantic digital twins, an enterprise data space naturally emerges. This semantic foundation gives autonomous agents the precise situational awareness required to negotiate contracts, reallocate inventory, and resolve supply disruptions without human micromanagement.

What makes data quality for agentic AI different from traditional data hygiene?

Traditional data hygiene focuses on surface-level syntax. For decades, IT organizations evaluated data quality through completeness, uniqueness, and adherence to rigid schema rules. In that legacy paradigm, if a material record in an enterprise resource planning (ERP) system contained an active part number, a short text description, and a valid unit of measure, the system marked it as high quality.

Autonomous agents, however, do not merely read text fields; they act on the operational implications behind that text. When an AI agent is tasked with mitigating a plant shutdown by procuring an alternative industrial pump, simple syntactic validity does not help it determine whether a replacement in another facility matches the necessary mechanical tolerances. Traditional data pipelines strip away domain relationships, leaving the agent with isolated fragments of information.

Table 1 highlights the operational differences between legacy data cleansing and the requirements of agentic AI.

When data lacks semantic connections, agentic systems fail. For example, why supply chain visibility dashboards miss bottlenecks traces back to identical physical components being recorded under different part numbers and local descriptions across regional business units. If a deterministic dashboard cannot spot these bottlenecks, an autonomous procurement agent relying on that same fragmented data will make flawed purchasing decisions. Real data quality for agentic AI requires data to be machine-legible, richly categorized, and fully contextualized.

What is a semantic digital twin for material master data?

A semantic digital twin for material master data is an AI-enriched digital representation of a physical item, part, or service that captures its real-world properties, technical attributes, and operational relationships. Instead of treating a spare part as an arbitrary eight-digit string with a truncated 40-character description, a semantic digital twin defines what the part actually is, where it is used, how it operates, and which suppliers can deliver it.

In industrial supply chains, identical physical materials are routinely logged under completely disparate naming conventions. A maintenance technician in Germany might record an item in German using technical jargon, while a plant engineer in Brazil records the exact same item in Portuguese using manufacturer part numbers. Traditional database tools treat these as distinct records, inflating inventory costs and obscuring spend visibility.

Building a semantic digital twin resolves this fragmentation by establishing five core dimensions:

1. **Taxonomic classification:** The item is classified into a rigorous industrial hierarchy, mapping automatically to global taxonomies such as UNSPSC or internal spend categories.
2. **Technical attribute extraction:** Essential operating specifications (such as outer diameter, voltage, material composition, and pressure rating) are extracted from unstructured text and normalized into standardized key-value pairs.
3. **Multilingual normalization:** As explored in

mapping multilingual spend data with AI techniques, language barriers disappear when deep learning models identify the underlying engineering concept rather than performing literal dictionary lookups.

4. **Relational linking:** The material is connected to parent assets, production lines, approved suppliers, historical contract prices, and alternative interchangeable parts.
5. **Continuous synchronization:** As field engineers update maintenance logs or suppliers update specifications, the digital twin updates continuously to reflect real-world status.

Once an organization builds semantic digital twins for its materials, assets, and suppliers, it transforms passive data into an active knowledge graph. AI agents can then inspect this graph to evaluate whether two items with different internal codes are functionally identical, allowing them to consolidate purchases or redeploy surplus stock across global plants.

Enterprise data space vs traditional data lakes for procurement

Many enterprise leaders assume that dumping all procurement, inventory, and supplier data into a cloud data lake creates an AI-ready data foundation. In practice, traditional data lakes often become unmanageable data swamps when applied to autonomous procurement workflows.

A data lake stores raw, disparate tables in centralized object storage. While this architecture allows data scientists to run bulk analytical queries, it strips away business logic and fails to preserve the functional links between systems. An AI agent browsing a traditional data lake sees trillions of disconnected data points with no intrinsic sense of hierarchy or operational meaning. It cannot easily infer whether an invoice in one ERP system relates to a material master record in another.

In contrast, an enterprise data space (EDS) is not a standalone storage repository that you purchase off the shelf. Instead, an enterprise data space is the semantic environment that emerges naturally once all digital twins across materials, suppliers, assets, and spend are

built and interlinked.

By utilizing standardized communication protocols such as the Model Context Protocol (MCP) and open application programming interfaces (APIs), the enterprise data space exposes the underlying knowledge graph directly to external software. Rather than writing brittle SQL joins across disjointed database tables, AI agents navigate an interconnected web of verified enterprise entities. The enterprise data space guarantees that whenever an agent queries an entity, it receives the complete contextual picture, including contractual terms, physical inventory levels, supplier certifications, and machine criticality ratings.

How autonomous AI agents in procurement navigate semantic relationships

Autonomous enterprise agents do not simply generate text; they make decisions and execute actions across transactional software. In procurement, vertical AI agents handle tasks ranging from automated spot buying and contract leakage detection to predictive replenishment and inventory balancing.

The guiding operational principle for modern procurement automation is simple: stop querying records, start navigating relationships.

Consider an agent assigned to manage spend visibility and rogue purchasing. In a typical legacy setup, an agent searching for spending anomalies might query a purchase order database for off-contract transactions. However, if the purchase descriptions are vague or the supplier names vary by country, the agent cannot tell whether a purchase violated a contract.

When operating on top of a semantically connected enterprise data space, the agent executes a structured relational traversal:

- **Entity resolution:** The agent identifies the item in the transaction and links it directly to its semantic digital twin, ignoring discrepancies in local descriptions or languages.
- **Contract discovery:** The agent traverses the relationship graph from the material twin to the global category level, identifying

existing master service agreements or frame contracts negotiated by other business units.

- **Supplier evaluation:** It checks whether the supplying vendor is an affiliate of an approved global partner or a non-compliant third party.
- **Action execution:** If the purchase is off-contract, the agent flags the leakage, alerts category managers, and updates the procurement routing engine so future requests route to the preferred supplier automatically.

Without semantic data quality, vertical AI agents are blind to these operational connections. By grounding agents in a verified web of digital twins, organizations give their autonomous systems the factual scaffolding necessary to perform end-to-end automation safely.

Continuous master data governance: keeping agentic AI grounded in reality

Data quality is not a one-time migration task; it is an ongoing operational condition. If master data deteriorates over time, the performance of autonomous agents will degrade alongside it. Maintaining data quality for agentic AI requires continuous master data governance that blends machine intelligence with human domain expertise.

Many organizations encounter governance friction when modernizing core enterprise systems. As detailed in the guide on AI procurement master data governance for SAP environments, classic governance frameworks excel at enforcing approval workflows, but they frequently lack the native semantic intelligence to interpret unstructured industrial descriptions or prevent duplicate entries at the point of creation.

Effective continuous governance incorporates three vital elements:

- **Real-time prevention at entry:** When an engineer or buyer requests a new material code, AI engines parse the text instantly, suggest the exact category, extract technical attributes, and prevent duplicates before the record enters the ERP system.
- **Integration of human field knowledge:** Warehouse operators, maintenance

technicians, and field engineers possess crucial practical context that rarely makes its way into enterprise databases. A resilient governance framework captures this frontline expertise and feeds it directly into the digital twin, ensuring that field adjustments update the global knowledge graph.

- **Harmonized spend analytics:** Continuous governance must link directly to spend classification and contract matching. When material masters remain pristine, spend analysis functions as a natural extension of governance, providing real-time visibility into global procurement volumes.

When human oversight and automated semantic governance work in tandem, your enterprise data foundation remains accurate, dependable, and ready for autonomous agentic workflows.

Transforming your procurement architecture with Creatives

Deploying agentic AI without fixing semantic data quality is a recipe for operational disruption. True enterprise automation requires a proven architectural foundation where every material, supplier, and asset is represented by an enriched, machine-legible digital model.

Creatives provides the enterprise software foundation that makes this transformation possible. By building semantic digital twins of complex procurement data and interlinking them into an accessible enterprise data space, Creatives enables global organizations across manufacturing, energy, utilities, and oil and gas to turn fragmented data into an active knowledge graph.

Through advanced AI master data governance, multilingual normalization, and technical attribute management, Creatives ensures that your enterprise systems speak a coherent, machine-actionable language. By establishing reliable semantic digital twins first, you create the trusted data environment where vertical AI agents can safely optimize spend, eliminate warehouse duplication, and automate complex procurement operations across your global enterprise.

Table 1: Traditional data hygiene vs semantic data quality for agentic AI

Quality dimension	Traditional data hygiene	Semantic data quality for agentic AI
Primary focus	Field formatting, deduplication, and schema validation	Relational context, entity resolution, and functional equivalence
Data representation	Relational tables, rows, columns, and isolated records	Knowledge graphs and semantic digital twins
System boundaries	Trapped within specific application databases (ERP, MDM)	Connected dynamically across an enterprise data space
Handling ambiguity	Flags missing fields or rejected transactions	Resolves multilingual terms and maps domain-specific attributes
Agent utility	Low; agents lack real-world context and hallucinate	High; agents navigate explicit operational dependencies

Creactives builds a trusted Data Foundation that creates a Digital Twin of your Master and Spend Data. Our AI-powered solutions eliminate errors, optimize legacy systems, and unlock a wide range of use cases across Procurement & Supply Chain.



SPEND ANALYTICS

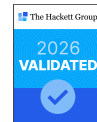


**MATERIALS & SERVICES
MASTER DATA GOVERNANCE**



**PROCUREMENT GUIDANCE
AUTOMATION**

CREACTIVES: SOLUTIONS SELECTED OVER YEARS BY OPINION LEADERS



"Creactives is a best-of-breed tool that offers standard spend analytics features, such as data cleansing and analysis. In addition, it specializes in materials MDM and is built for automated parts and material classification."

Abigail Ommen, Research Analyst & Analyst Production Manager