How to find duplicate materials with AI in global inventory

To find duplicate materials with AI and reduce inventory duplicate data, supply chain directors must transition from basic data cleansing to building a semantically connected digital twin. Traditional manual cleaning cannot scale across fragmented global ERP systems, leaving duplicate parts hidden under differing names, languages, and part numbers. Implementing a semantic digital twin allows autonomous procurement agents to accurately identify MRO inventory duplicates, optimize warehouse spend, and establish continuous ERP spend data governance.

For procurement organizations, redundant inventory ties up working capital and inflates warehousing costs. When your goal is to reduce inventory value, relying on legacy databases to spot duplicate parts is no longer sufficient. This guide provides an actionable master data readiness checklist to evaluate if your material master data is harmonized enough to support agentic AI.

Why basic data cleansing fails autonomous procurement agents

Basic data cleansing relies on rigid taxonomic rules, string matching, and manual interventions. While this approach can fix simple spelling mistakes or standard formatting issues, it fails to handle the complex realities of global supply chains. When multiple facilities operate on different ERP instances, they inevitably describe identical physical parts in completely different ways.

Autonomous procurement agents, which use artificial intelligence to automate purchasing decisions, require machine-legible data to function. If your agentic AI cannot recognize that two separate database entries represent the exact same physical item, it will continue to order redundant stock.

Traditional cleansing methods fail to support these autonomous agents for several reasons:

- **The limitation of string matching:** If one plant records a part as “Ball bearing 6204-2RSH” and another records it as “Brg 6204 2RS”, standard database queries will treat them as different items.

- **The challenge of multilingual data:** Global organizations often have descriptions written in German, English, Spanish, or Chinese across different localized systems. Standard cleansing tools cannot reconcile these descriptions without manual, slow, and expensive translation projects.
- **Lack of contextual understanding:** Traditional databases look at fields in isolation. They lack the semantic context to understand that a manufacturer part number, a supplier ID, and an internal material code are all interconnected.

To make your data machine-legible, you must move beyond basic cleansing toward deep master data harmonization. This means creating a unified semantic framework where the system understands the actual meaning and physical attributes of every material, regardless of how it was entered into an ERP.

The master data readiness checklist for agentic AI

Before deploying autonomous procurement agents to manage your inventory and purchasing, you must assess whether your master data is ready. This checklist helps supply chain directors and master data managers evaluate their current data quality and structure.

- **Multilingual normalization.** Are material descriptions in different languages automatically translated and standardized into a single, uniform format? This prevents duplicate purchases caused by language barriers across international plants.
- **Technical attribute extraction.** Can your system automatically extract physical dimensions, materials, and tolerances from unstructured free-text descriptions? This allows agents to compare functional equivalents, even when manufacturer part numbers are missing.
- **Active knowledge graph integration.** Is your master data connected to a live semantic knowledge graph that maps relationships between materials, suppliers, and equipment? This enables agents to understand the context of each part, including where it is used and who supplies it.
- **Point-of-entry data validation.** Does your system prevent the creation of new duplicate records at the moment a user requests a new material code? This ensures master data remains clean over time, stopping the growth of MRO inventory duplicates.
- **Cross-ERP alignment.** Are material records across different legacy ERP instances mapped to a single virtual record without requiring a massive IT migration? This allows global visibility of inventory, enabling plants to share stock instead of buying new parts.

How a semantic digital twin eliminates MRO inventory duplicates

A semantic digital twin is a virtual representation of your physical assets, materials, suppliers, and equipment, all connected by their real-world meaning rather than just database identifiers. Instead of forcing every local business unit to adopt a single, rigid description standard, the semantic digital twin acts as an intelligent translation and mapping layer.

This digital twin uses artificial intelligence to analyze unstructured text, find patterns, and establish relationships. For example, if your plant in Germany stocks a “Keilriemen XPZ 1000” and your plant in the United States stocks an “XPZ1000 wedge belt”, the semantic digital twin recognizes that these are functionally identical items.

By mapping both entries to the same virtual node in a digital twin, the system immediately identifies the duplicate. This capability transforms how supply chain directors manage MRO inventory duplicates by allowing you to:

- **Redistribute existing stock:** Instead of purchasing a new part for the German plant, the autonomous agent can identify that the US plant has surplus stock of the exact same item.
- **Consolidate supplier spend:** When you identify that different plants buy the same part from different local distributors under varying names, you can leverage your global buying power to negotiate a single contract with the primary manufacturer.
- **Eliminate redundant safety stock:** You no longer need to hold safety stock for five seemingly different parts when they are actually the same physical item.

Step-by-step process to find duplicate materials with AI

If you want to find duplicate materials with AI across your global systems, you should follow a structured methodology that prioritizes semantic understanding over simple text matching.

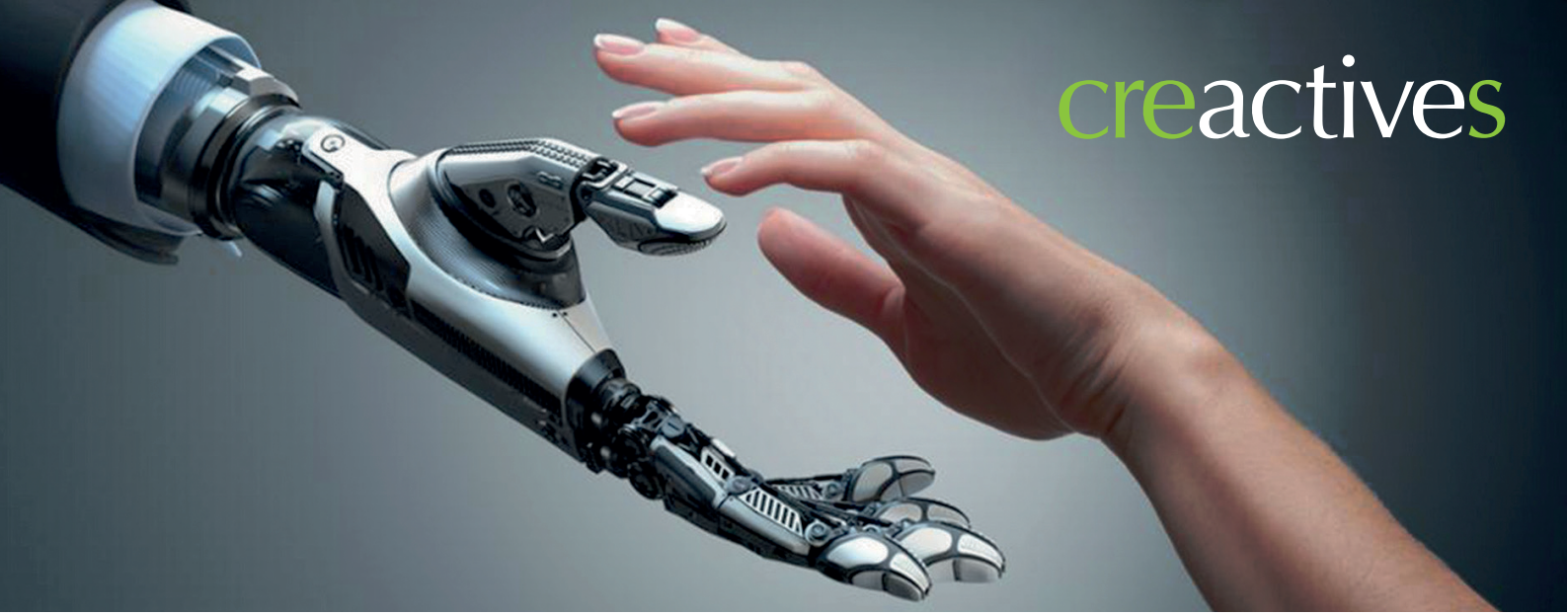
1. **Ingest fragmented ERP master and spend data.** Gather material master records and spend data from all your global ERP systems, legacy databases, and warehousing platforms. Do not worry about cleaning the data before ingestion: the AI is designed to handle dirty, unstructured, and multilingual text.
2. **Apply semantic classification.** The AI scans your material descriptions and classifies them into a unified taxonomy. This step groups similar items together, such as placing all bearings, valves, or electrical components into their respective logical categories, regardless of how they were originally named.
3. **Extract and normalize technical attributes.** Once the materials are classified, the AI extracts key physical attributes from the free-text fields. This includes measurements, materials, pressure ratings, and temperatures. The system normalizes these attributes into standard units of measure, such as converting “10mm”, “10 mm”, and “0.39 inches” into a single, searchable value.
4. **Build the active knowledge graph.** The normalized data is then mapped into an active knowledge graph. This graph connects the material to its manufacturers, suppliers, and the specific machines or assets where the part is used. This step creates the semantic digital twin of your supply chain.
5. **Identify and resolve duplicate clusters.** With the knowledge graph in place, the AI identifies clusters of identical or highly compatible materials. The system flags these duplicates for your master data team or autonomous procurement agents, allowing you to archive redundant codes and update your inventory strategies.

Driving continuous master data governance with Creactives

Achieving clean data is not a one-time project. As soon as a data cleansing initiative ends, new duplicate parts begin to creep back into your ERP systems as users request new material codes. To prevent this, global manufacturing and utility enterprises require continuous master data governance.

Creactives provides enterprise software that builds semantic digital twins of procurement data, linking materials, suppliers, assets, and equipment into a unified space. This semantic foundation enables vertical AI agents and analytics to optimize spend and maintain continuous master data quality.

Rather than relying on slow, manual cleaning processes, the proprietary AI developed by Creactives harmonizes master data twelve times faster than traditional methods, even during complex corporate acquisitions or S/4HANA migrations. By validating material data at the point of entry and translating multilingual descriptions automatically, the software ensures your material master data remains permanently clean, machine-legible, and fully optimized for autonomous procurement agents: https://www.creactives.com/



creactives

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

Creactives Group SPA
Via Fermi, 4 – 37135 Verona – Italy

creactives

creactives.com
linkedin.com/company/creactives-spa