Data cleansing: how buyers keep their master data in shape for the long term

Data cleansing refers to the systematic review, correction and enrichment of data records, in particular material, supplier and procurement master data. In industrial procurement, it is a prerequisite for reliable processes, robust reporting and automated workflows. Without clean data, even modern ERP and e-procurement systems only function to a limited extent. The following article explains methods, KPIs and practical measures for sustainable data cleansing in procurement.

As of October 2026 · amplifa-marketplace.ai/en/glossary/datenbereinigung

Key facts at a glance

01 Data cleansing involves identifying, correcting and deduplicating faulty or duplicate data records.

02 Typical problem areas are material master records, supplier master data, free-text orders and price and condition data.

03 Well-cleansed data frequently reduces process costs in practice and lowers error rates in ordering, goods receipt and invoice verification.

04 Key metrics are completion rate, duplicate rate and the proportion of standardised material classifications.

05 Routine operation beats large projects: six-monthly to annual cleansing cycles are considered standard practice.

06 Clear data governance with defined responsibilities prevents errors from re-emerging after cleansing.

Definition: data cleansing

Data cleansing (data cleansing) is the systematic process of reviewing data records for errors, duplicates, gaps and inconsistencies and correcting them permanently. In the procurement context, it

primarily targets master data that drives ordering, sourcing and reporting processes.

Distinction from data maintenance

Data cleansing is a one-off or cyclical corrective measure with project character, whereas data maintenance refers to ongoing updating in day-to-day operations. The two interact: once cleansed, the data foundation must be kept stable through continuous maintenance. Without clear rules for new records and changes, errors typically return within a few months.

- Cleansing: project-style correction of existing data
- Maintenance: continuous updating in routine operation
- Governance: rules, roles and approvals for data entry

Relevant data records in procurement

The focus is on material master records with descriptions, units of measure and classification keys as well as supplier master data with contacts, payment and rating information. Added to these are condition and price data, spare part assignments and the data behind every free-text order. Documents such as drawings and specifications also belong to this if they are used for procurement decisions.

Methods and approach to data cleansing

An effective approach in practice follows a multi-stage process: analysis, standardisation, deduplication, enrichment and validation. The choice of methods depends on data volume, the system landscape and the error culture in data entry.

Analysis and duplicate checking

At the outset, a data quality report is produced showing completion rates, error rates and inconsistencies per field and data area. Duplicates are identified using fuzzy matching methods that also recognise spelling variants, abbreviations and translations. For supplier master data, additional verification criteria such as VAT identification number and commercial register data help identify truly identical suppliers. The results are then prioritised by business criticality: master data for A-parts and core suppliers is cleansed first.

Standardisation and classification

Material descriptions are structured according to uniform naming conventions, typically in the pattern nomenclature – attribute – value. A recognised classification scheme such as eCl@ss or UNSPSC enables group-level evaluations and supports compare sourcing. Assignment can be partly automated through pattern recognition, but requires expert approval. Defining mandatory fields is important so that new records are captured completely from the start.

Enrichment and validation

After correction, records with missing information are enriched, for example with manufacturer keys,

approval status or supplier ratings. Validation is carried out on a sampling basis by the specialist departments – ideally via a defined procurement approval workflow in which reviewed records are released. Only then is a record considered cleansed and released for automation such as catalogue orders or call-off orders.

Key KPIs for data quality

Data quality must be measurable, otherwise cleansing projects lose support after go-live. Proven metrics can be derived directly from the ERP system and integrated into regular procurement controlling.

Completeness and correctness metrics

The completion rate indicates what percentage of mandatory fields is populated per record. The correctness rate measures how many records pass an expert review without objections. Both values can be evaluated per material group or supplier area and summarised in a periodic data quality report. This makes visible where further refinement is needed.

Structural and process metrics

The duplicate rate shows the proportion of duplicate material or supplier records. The proportion of classified materials and the proportion of catalogue-eligible items reveal how far standardisation has progressed. The share of free-text orders in order volume is also practically relevant: a high value often indicates incomplete master data and hampers statistics and procurement controlling.

Risks and controls in data cleansing

Data cleansing carries its own risks: incorrectly corrected records can have worse consequences than the original errors. Controls, approval levels and documentation are therefore indispensable.

Technical and legal risks

Merging duplicates can lose historical order and contract relationships if change histories are not migrated along with them. For supplier master data, bank-related data is particularly sensitive: errors or manipulations here can lead to mispayments, which is why changes should run through a secure procurement approval workflow with a four-eyes principle. Legal requirements on data protection and retention must be examined case by case.

Organisational controls

A clear allocation of roles has proven effective: who may create materials, who approves, who deletes? Change processes for master data should be structured like changes to components – including documented approval, similar to a change order. A temporary freeze on new records during cleansing prevents new duplicates from being created in parallel. After the project, a firmly defined master data team organisation takes over ongoing monitoring.

Typical sources of error

Errors often arise at the point of entry: free-text descriptions from urgent orders, local designations used by individual plants or unchecked imports from legacy systems. Interfaces to supplier portals and data exchange, such as the transfer of technical documentation, can also cause inconsistencies if mapping rules are missing. Addressing these causes is more sustainable than merely cleansing symptoms.

Current developments: automation and AI-supported cleansing

Data cleansing is increasingly supported by software: matching algorithms, language models and classification tools take over routine tasks. However, human expert approval remains central, especially for technical master data.

Automated classification and matching

Modern tools automatically suggest classifications for unclassified materials and recognise duplicates even across site boundaries. In practice, such suggestions achieve high hit rates for simple materials but still require expert review for critical parts. The benefit lies less in saving entry time than in making it possible to cleanse large legacy volumes at all within a reasonable timeframe.

Data quality as a foundation for automation

Agile procurement approaches, automatic purchase requisitions and AI-supported supplier selection only work with reliable data. The same applies to evaluations such as plan-actual inventory comparison or the derivation of order quantities via the economic order quantity (EOQ): incorrect master data distorts every optimisation. Data cleansing is therefore not an isolated IT project but the foundation of the digital procurement strategy.

Typical metrics

Metric	Typical value
Mandatory field completion rate	> 98% *Proportion of fully populated mandatory fields per record, often used as a target value for critical material groups*
Duplicate rate	< 1% *Proportion of duplicate records; typically significantly higher in uncleansed legacy data*
Proportion of classified materials	> 95% *Proportion of materials with a valid classification key, e.g. eCl@ss*

Proportion of free-text orders	< 10% *A high level indicates incomplete master data and insufficient catalogue coverage*

Practical example

A mid-sized mechanical engineering company with several sites discovered that the same ball bearing was managed under four different material numbers – with different descriptions and prices. A duplicate cleansing exercise across all plants reduced the number of material masters many times over and revealed bundling potential. In parallel, the company introduced a data quality report showing completion rates and duplicate rates on a quarterly basis.

- Duplicate check of all material master records across site boundaries
- Introduction of binding naming conventions and mandatory fields for new records
- Approval process for master data changes based on the four-eyes principle
- Quarterly data quality report as a procurement controlling instrument

Conclusion

In industrial procurement, data cleansing is not a one-off project but an ongoing task with measurable benefits. Clean material and supplier master data enable bundling, automated processes and reliable KPIs. What matters is the combination of methodical cleansing, clear governance and continuous maintenance in routine operation. Companies that systematically prepare their data foundation create the basis for sourcing strategies, automation and reliable reporting. The effort typically pays for itself through reduced process costs and better decision-making foundations.

Frequently asked questions

? What does data cleansing mean in procurement?

Data cleansing refers to the systematic review and correction of procurement data such as material, supplier and condition master records. This includes removing duplicates, completing missing fields and standardising descriptions. The goal is a data foundation on which ordering processes, reporting and automation can reliably build.

? Why are duplicate material master records problematic?

Duplicates fragment procurement volumes: the same item is sourced under multiple numbers from different suppliers or at different prices. As a result, bundling potential is lost, inventory data is distorted and reporting provides a skewed picture. Warehouse stocks and materials planning also suffer from unclear material identities.

? How often should procurement master data be cleansed?

In practice, six-monthly to annual cleansing cycles for critical data areas have proven effective, supplemented by continuous maintenance in day-to-day operations. More important than the cadence is preventing new errors through clear data entry rules and approval processes. An ongoing data quality report shows where corrective action is needed.

? Which KPIs measure data quality in procurement?

Proven metrics include mandatory field completion rate, duplicate rate, proportion of classified materials and proportion of catalogue-eligible items. In addition, the share of free-text orders provides indications of master data gaps. These KPIs can usually be extracted directly from the ERP system and evaluated periodically.

? Can data cleansing be automated?

Partly, yes: duplicate matching and classification suggestions can today be automated well, especially for simple materials. However, expert approvals, particularly for critical parts or sensitive supplier data, remain manual. Automation accelerates the cleansing of large volumes but does not replace expert review.

Find the right suppliers on Amplifa

Verified industrial suppliers in Germany, Austria and Switzerland. Describe your requirement and receive comparable offers.

amplifa-marketplace.ai