

Reference data: the data foundation for consistent processes in procurement

Reference data is standardised master data that serves as a uniform basis of reference for business processes, systems and reports. In technical procurement, it comprises in particular material master data, vendor master data, classification systems as well as price and quantity references. Without clean reference data, neither automated ordering processes nor reliable reporting will function. This article explains the structure, quality assurance and practical use of reference data in procurement controlling.

Key facts

- Reference data is standardised, frequently used master data such as material numbers, vendor master records and classifications, serving as a common basis of reference across different systems.
- It differs from transaction data, which records individual transactions such as purchase orders or goods receipts.
- Frequently used standards are UNSPSC, eCl@ss and GLN for the worldwide, unambiguous identification of materials and business partners.
- Typical quality criteria are completeness, uniqueness, currency and consistency across all connected systems.
- Poor reference data leads in practice to duplicate orders, missing parts, incorrect prices and extensive manual rework.
- Master data management with clearly defined responsibilities, known as data owners, is the foundation of permanently clean data stocks.

Definition: reference data

Reference data is master data with a high degree of reusability that serves as a reference for operational business processes. It describes permanently valid objects such as materials, services, suppliers or cost centres and, compared with movement data, changes only rarely. By contrast, transaction data is generated with each individual business transaction, such as a purchase order or a goods receipt.

Distinction from master data and transaction data

In practice, the term master data is broader and covers all object-related base data. Reference data is the subset used as a uniform point of reference across system and company boundaries, for example a material number that appears identically in the ERP, catalogue system and EDI communication. Transaction data such as purchase orders or invoices always refers to this reference data and cannot be interpreted unambiguously without it.

Key types in procurement

For buyers, four types of data are particularly relevant: material master data with description, unit of measure and classification; vendor master data with identifiers and creditworthiness data; contract and price references; and technical reference data such as drawing numbers and standards. In addition, reference data for sustainability information is playing a growing role, for example for assigning Product Carbon Footprint (PCF) values at material level.

Structure and standards: how reference data is organised

So that reference data remains interpretable across systems, departments and business partners, it needs a uniform structure. This comprises unique identifiers, controlled vocabularies for designations, and classification systems for grouping. In practice, a staged approach has proven effective: first identification, then classification, then enrichment with attributed specialist data.

Classification systems

Internationally widespread standards are UNSPSC (United Nations Standard Products and Services Code) for goods and services, and eCl@ss, which is frequently used in German-speaking mechanical and plant engineering because of its technical characteristics. A harmonised classification system enables spend analyses across plants and locations and forms the basis for meaningful benchmarking in procurement.

Identifiers and master record logic

Every referenced object needs a unique key: material numbers in the company's own system, vendor numbers, GLN (Global Location Number) for locations in EDI exchange and GTIN for standard commercial items. What matters is the so-called golden record logic: there is exactly one leading record per object from which derived systems are fed. Duplicate and alias records are systematically identified and merged.

Data model and mandatory fields

A workable data model defines mandatory fields, value ranges and validation rules per object type. For material master data, this typically includes a description following a naming convention, unit of measure, commodity group, purchasing and sales organisation views, and technical characteristics. For vendor master records, bank details, payment terms, certifications and blocking indicators are relevant. The more comprehensively mandatory field completion is checked at creation, the lower the later correction effort.

Methods and approach: building and maintaining reference data

Building clean reference data follows a recurring cycle of cleansing, harmonisation and continuous maintenance. In practice, many companies start with a deduplication project on the material master, followed by the introduction of a central creation point with defined roles. What is important here is to organise data quality not as a one-off project but as an ongoing process.

Data cleansing and deduplication

Typically, the starting point is a stock analysis: how many materials are there, how many are duplicates, how many have been unused for years? In practice, stock analyses frequently show that a significant proportion of material master records are inactive or have been created in duplicate. Cleansing tools

identify duplicates using fuzzy search, normalisation of designations and comparison of technical characteristics. The approval of merged records should be carried out by subject-matter data owners.

Governance and data ownership

Reference data maintenance requires clear responsibilities: who creates, who reviews, who approves, who archives? A proven approach is a central master data team with defined service-level agreements for processing times, complemented by subject-matter data owners in the procurement and technology departments. Change processes must be documented, particularly for price-relevant or compliance-relevant fields. A procurement guide can set out binding naming conventions and creation processes [here](#).

Interfaces and EDI

Reference data only delivers its full benefit when it is exchanged consistently between systems. Via EDI (Electronic Data Interchange), master data, purchase orders and invoices are reconciled automatically; for this purpose, reference data such as vendor numbers and item identifiers must be maintained identically on both sides. In practice, it has proven effective to transfer changes to reference data in a versioned manner and to report discrepancies automatically, rather than maintaining them manually after the event.

Key KPIs for data quality

Data quality can be measured. In practice, a number of metrics have become established that are evaluated regularly – for example monthly. They make progress visible and help to justify resources for data maintenance.

Completeness and uniqueness

The completeness rate indicates what percentage of master records have all mandatory fields populated. The duplicate rate measures the proportion of objects created in duplicate. Both figures should be considered together, because a high level of completeness is of little use if the same item is maintained under five numbers.

Currency and consistency

The currency metric measures how many records have been reviewed or confirmed within a defined period. Consistency checks compare field contents between source systems, for example whether the vendor name is identical in the ERP and the e-catalogue. Discrepancies are typically assessed in a risk matrix by severity and frequency in order to derive correction priorities.

Risks and controls: what poor reference data costs

Faulty reference data has a direct impact on operational and financial results. Typical consequences are duplicate orders, mispostings, delayed supply processes and inadequate compliance. Controls should therefore be applied preventively at creation and detectively during ongoing operations.

Operational risks

Duplicate material numbers lead to unnecessary orders and inventory; unclear units of measure cause incorrect quantities. When reference data for supplier evaluation is faulty, supplier performance is assessed incorrectly – for example when parts per million (PPM) figures are distorted by incorrect

material assignment. In the quality context, incomplete master data also impedes the traceability of materials, which is mandatory particularly in regulated industries.

Financial and compliance risks

In the three-way match between purchase order, goods receipt and invoice, deviating reference data leads to discrepancies and manual clarifications. Outdated price references can result in overpayments; missing blocking indicators in the vendor master allow payments to blocked or fraudulent payees. Compliance-relevant data such as a Restricted Substances List (RSL) must be referenced without gaps at material level, otherwise supply chain breaches loom.

Control instruments

Proven controls include mandatory field validation at creation, four-eyes approval for sensitive fields, periodic duplicate and inactivity scans, and reconciliation reports between systems. Change logs and regular sampling through internal audit rounds complement the control system. It is important that control findings feed into an improvement process rather than languishing as isolated reports.

Current developments: automation and sustainability data

Two trends are currently having a noticeable impact on the handling of reference data: automated data maintenance through AI and growing sustainability and regulatory requirements. Both significantly raise the demands placed on data quality and structure.

AI-supported classification and deduplication

Language models and matching algorithms now classify free-text item descriptions largely automatically into standard systems such as UNSPSC or eCl@ss. In practice, this significantly shortens creation processes, but still requires human approval, particularly for technical specifications. AI-supported duplicate detection goes beyond fuzzy matching and recognises semantically similar designations even across language boundaries.

Sustainability and supply chain data

Reporting obligations relating to the supply chain and emissions figures require machine-readable reference data for suppliers, materials and countries of origin. Assigning PCF values at material level presupposes that material master data is consistently classified. Traceability of materials, for example for conflict minerals or recycled content, is likewise mapped via referenced batch and supplier data. Companies that harmonise their reference data foundation early gain a structural advantage here.

Typical metrics

Completeness rate: > 95%

Proportion of master records with all mandatory fields fully populated

Duplicate rate: < 2%

Proportion of duplicated objects in the total stock of records

Currency rate: > 90%

Proportion of master records reviewed or confirmed within the defined period

EDI automation level: > 80%

Proportion of master data transmissions that can be processed without manual correction

Practical example

A medium-sized mechanical engineering company with several plants discovers during a spend analysis that the same ball bearing is maintained under six different material numbers and descriptions. As a result, volume tiers are not exploited, stock is built up twice, and supplier enquiries contain contradictory information. The company launches a master data project with the following approach:

- Duplicate analysis of all active and inactive material master records, followed by consolidation by subject-matter data owners.
- Introduction of a central creation point with a naming convention, mandatory fields and eCl@ss classification as a mandatory entry.
- Establishment of a monthly data quality report with completeness, duplicate and currency metrics.
- Reconciliation of the cleansed material references with supplier catalogues and EDI interfaces, including correction of outdated price references before the next round of negotiations.

Conclusion

Reference data is the invisible foundation of every efficient procurement function. It determines whether spend analyses are reliable, whether automation via EDI works and whether compliance requirements can be traced without gaps. The effort required for cleansing and governance typically pays for itself quickly through reduced duplicate work, a stronger negotiating position and fewer missing parts. What is decisive is to organise data quality as a permanent process with clear responsibilities, not as a one-off project. Those who have their reference data foundation in order create the prerequisites for digital procurement, sustainability reporting and reliable procurement controlling.

Frequently asked questions

What are reference data in simple terms?

Reference data is standardised master data that serves as a uniform basis of reference for business processes and systems. This includes, for example, material numbers, vendor master records and classifications. It changes rarely and is referenced by many transactions such as purchase orders and invoices.

What is the difference between reference data and transaction data?

Reference data describes permanently valid objects such as materials or suppliers and changes rarely. Transaction data is generated with each individual business transaction, such as a purchase order or invoice, and refers to reference data. Without consistent reference data, transaction data cannot be evaluated unambiguously.

Which standards exist for reference data in procurement?

Common standards include UNSPSC and eCl@ss for the classification of goods and services, as well as GLN and GTIN for the identification of locations and items. In German-speaking mechanical engineering, eCl@ss is frequently used because of its technical characteristics. The choice of standard should be made consistently across the group.

How do you measure the quality of reference data?

Typical metrics are the completeness rate (proportion of master records with mandatory fields populated), the duplicate rate, the currency rate and consistency across systems. In practice, these

figures are evaluated regularly, for example monthly. They make progress measurable and justify resources for data maintenance.

What does poor data quality cost in procurement?

Consequences include duplicate orders, unnecessary inventory, manual clarifications during invoice reconciliation and incorrect spend analyses. Negotiation outcomes also suffer because quantities and volumes cannot be aggregated correctly. In practice, this results in significant ongoing costs that can be substantially reduced through data cleansing.

Who is responsible for maintaining reference data?

A proven model is one with a central master data team responsible for process and systematics, plus subject-matter data owners in the procurement and functional departments. These approve changes and carry content responsibility. Clear roles and defined processing times are decisive for lasting data quality.