

Datenbereinigung: Wie Einkäufer ihre Stammdaten auf Dauer in Schuss halten

Datenbereinigung bezeichnet die systematische Prüfung, Korrektur und Anreicherung von Datenbeständen, insbesondere von Material-, Lieferanten- und Einkaufsstammdaten. Im industriellen Einkauf ist sie Voraussetzung für verlässliche Prozesse, belastbare Auswertungen und automatisierte Workflows. Ohne saubere Daten funktionieren selbst moderne ERP- und E-Procurement-Systeme nur eingeschränkt. Der folgende Artikel erklärt Methoden, Kennzahlen und Praxismaßnahmen für eine nachhaltige Datenbereinigung im Einkauf.

Das Wichtigste in Kürze

- Datenbereinigung umfasst Identifizieren, Korrigieren und Deduplizieren fehlerhafter oder doppelter Datenbestände.
- Typische Problemfelder sind Materialstammsätze, Lieferantenstammdaten, Freitextbestellungen und Preis- sowie Konditionsdaten.
- Gut bereinigte Daten senken in der Praxis häufig Prozesskosten und reduzieren Fehlerquoten bei Bestellung, Wareneingang und Rechnungsprüfung.
- Zentrale Kennzahlen sind Füllgrad, Dublettenrate und Anteil standardisierter Materialklassifikationen.
- Regelbetrieb schlägt Großprojekte: Halbjährliche bis jährliche Bereinigungszyklen gelten in der Praxis als üblich.
- Klare Data Governance mit definierten Verantwortlichkeiten verhindert, dass Fehler nach der Bereinigung erneut entstehen.

Definition: Datenbereinigung

Datenbereinigung (Data Cleansing) ist der systematische Prozess, Datenbestände auf Fehler, Duplikate, Lücken und Inkonsistenzen zu prüfen und dauerhaft zu korrigieren. Im Kontext des Einkaufs richtet sie sich vor allem auf Stammdaten, die Bestell-, Beschaffungs- und Reportingprozesse steuern.

Abgrenzung zur Datenpflege

Datenbereinigung ist eine punktuelle oder zyklische Korrekturmaßnahme mit Projektcharakter, während Datenpflege die laufende Aktualisierung im Tagesgeschäft bezeichnet. Beide greifen ineinander: Eine einmal bereinigte Datenbasis muss durch kontinuierliche Pflege stabil gehalten werden. Ohne klare Regeln für Neuanlagen und Änderungen kehren Fehler typischerweise innerhalb weniger Monate zurück.

- Bereinigung: projektartige Korrektur von Bestandsdaten
- Pflege: fortlaufende Aktualisierung im Regelbetrieb
- Governance: Regeln, Rollen und Freigaben für Datenerfassung

Relevante Datenbestände im Einkauf

Im Mittelpunkt stehen Materialstammsätze mit Beschreibungen, Maßeinheiten und Klassifikationsschlüsseln sowie Lieferantenstammdaten mit Ansprechpartnern, Zahlungs- und Bewertungsinformationen. Dazu kommen Konditions- und Preisdaten, Ersatzteilzuordnungen sowie die Daten hinter jeder Freitextbestellung. Auch Dokumente wie Zeichnungen und Lastenhefte gehören dazu, wenn sie für Beschaffungsentscheidungen herangezogen werden.

Methoden und Vorgehen der Datenbereinigung

Ein wirksames Vorgehen folgt in der Praxis einem mehrstufigen Ablauf: Analyse, Standardisierung, Deduplizierung, Anreicherung und Validierung. Die Wahl der Methoden hängt vom Datenvolumen, der Systemlandschaft und der Fehlerkultur in der Erfassung ab.

Analyse und Dublettenprüfung

Zu Beginn wird ein Datenqualitätsbericht erstellt, der Füllgrade, Fehlerquoten und Inkonsistenzen pro Feld und Datenbereich ausweist. Dubletten werden über Fuzzy-Matching-Verfahren identifiziert, die auch Schreibvarianten, Abkürzungen und Übersetzungen erkennen. Bei Lieferantenstammdaten helfen zusätzlich Prüfkriterien wie Umsatzsteuer-Identifikationsnummer und Handelsregisterdaten, wirklich identische Lieferanten zu erkennen. Die Ergebnisse Priorisierung nach Geschäftskritikalität: Stammdaten für A-Teile und Kernlieferanten werden zuerst bereinigt.

Standardisierung und Klassifikation

Materialbeschreibungen werden nach einheitlichen Namenskonventionen aufgebaut, typischerweise im Muster Nomenklatur – Merkmal – Wert. Ein Klassifikationsschema wie eCl@ss oder UNSPSC erlaubt gruppenweise Auswertungen und unterstützt Compare-Sourcing. Die Zuordnung kann teilweise automatisiert über Mustererkennung erfolgen, braucht aber fachliche Freigabe. Wichtig ist die Definition von Pflichtfeldern, damit neue Datensätze von Anfang an vollständig erfasst werden.

Anreicherung und Validierung

Nach der Korrektur werden Datensätze mit fehlenden Informationen ergänzt, etwa Herstellerschlüsseln, Zulassungsstatus oder Lieferantenbewertungen. Die Validierung erfolgt stichprobenartig durch die Fachbereiche – idealerweise über einen definierten Approval Workflow Einkauf, in dem geprüfte Datensätze freigegeben werden. Erst dann gilt ein Datensatz als bereinigt und wird für Automatisierungen wie Katalogbestellungen oder Abrufbestellungen freigegeben.

Wichtige KPIs zur Datenqualität

Datenqualität muss messbar sein, sonst verlieren Bereinigungsprojekte nach dem Go-live an Unterstützung. Bewährte Kennzahlen lassen sich direkt aus dem ERP-System ableiten und in das regelmäßige Einkaufscontrolling integrieren.

Vollständigkeits- und Korrektheitskennzahlen

Der Füllgrad gibt an, wie viel Prozent der Pflichtfelder pro Datensatz belegt sind. Die Korrektheitsquote misst, wie viele Datensätze eine fachliche Prüfung ohne Beanstandung passieren. Beide Werte lassen sich pro Materialgruppe oder Lieferantenbereich auswerten und in einem periodischen Datenqualitätsbericht zusammenfassen. So wird sichtbar, wo Nachschärfe nötig ist.

Struktur- und Prozesskennzahlen

Die Dublettenrate zeigt den Anteil doppelter Material- oder Lieferantensätze. Der Anteil klassifizierter Materialien und der Anteil katalogfähiger Positionen verraten, wie weit die Standardisierung fortgeschritten ist. Praktisch relevant ist auch der Anteil von Freitextbestellungen am Bestellvolumen: Ein hoher Wert deutet oft auf unvollständige Stammdaten hin und behindert Statistik und Einkaufscontrolling.

Risiken und Kontrollen bei der Datenbereinigung

Datenbereinigung birgt eigene Risiken: Falsch korrigierte Datensätze können schlimmere Folgen haben als die ursprünglichen Fehler. Kontrollen, Freigabestufen und Dokumentation sind daher unverzichtbar.

Fachliche und rechtliche Risiken

Das Zusammenführen von Dubletten kann historische Bestell- und Vertragsbeziehungen verlieren, wenn Änderungshistorien nicht mitwandern. Bei Lieferantenstammdaten sind bankbezogene Daten besonders sensibel: Fehler oder Manipulationen hier können zu Fehlzahlungen führen, weshalb Änderungen über einen gesicherten Approval Workflow Einkauf mit Vier-Augen-Prinzip laufen sollten. Rechtliche Vorgaben zum Datenschutz und zur Aufbewahrung sind im Einzelfall zu prüfen.

Organisatorische Kontrollen

Bewährt hat sich eine klare Rollenverteilung: Wer darf Materialien anlegen, wer freigeben, wer löschen? Change-Prozesse bei Stammdaten sollten so strukturiert sein wie Änderungen an Bauteilen – inklusive dokumentierter Freigabe, ähnlich einer Change Order. Ein temporäres Einfrieren von Neuanlagen während der Bereinigung verhindert, dass parallel neue Dubletten entstehen. Nach dem Projekt übernimmt eine fest definierte Stammteam-Organisation die laufende Kontrolle.

Typische Fehlerquellen

Häufig entstehen Fehler bereits bei der Erfassung: Freitextbeschreibungen aus dringenden Bestellungen, lokale Bezeichnungen einzelner Werke oder ungeprüfte Importe aus Altsystemen. Auch Schnittstellen zu Lieferantenportalen und zum Datenaustausch wie der IP-Übertragung technischer Unterlagen können Inkonsistenzen verursachen, wenn Mapping-Regeln fehlen. Diese Ursachen anzugehen ist nachhaltiger, als nur Symptome zu bereinigen.

Aktuelle Entwicklungen: Automatisierung und KI-gestützte Bereinigung

Datenbereinigung wird zunehmend durch Software unterstützt: Matching-Algorithmen, Sprachmodelle und Klassifikationstools übernehmen routineartige Aufgaben. Die menschliche Fachfreigabe bleibt aber zentral, gerade bei technischen Stammdaten.

Automatisierte Klassifikation und Matching

Moderne Tools schlagen für unklassifizierte Materialien automatisch Klassifikationsvorschläge vor und erkennen Dubletten auch über Standortgrenzen hinweg. In der Praxis erreichen solche Vorschläge hohe Trefferquoten bei einfachen Materialien, brauchen aber bei kritischen Teilen weiterhin fachliche Prüfung. Der Nutzen liegt weniger in der Einsparung von Erfassungszeit als in der Möglichkeit, große Altbestände überhaupt in angemessener Zeit zu bereinigen.

Datenqualität als Fundament für Automatisierung

Agile Einkauf-Ansätze, automatische Bestellanforderungen und KI-gestützte Lieferantenauswahl funktionieren nur mit verlässlichen Daten. Gleiches gilt für Auswertungen wie den Plan-Ist-Vergleich Bestand oder die Ableitung von Bestellmengen über die Economic Order Quantity (EOQ): Falsche Stammdaten verfälschen jede Optimierung. Datenbereinigung ist damit kein isoliertes IT-Projekt, sondern Grundlage der digitalen Einkaufsstrategie.

Typische Kennzahlen

Füllgrad Pflichtfelder: > 98 %

Anteil vollständig belegter Pflichtfelder pro Datensatz, häufig als Zielwert für kritische Materialgruppen

Dublettenrate: < 1 %

Anteil doppelter Datensätze; in unbereinigten Altbeständen typischerweise deutlich höher

Anteil klassifizierter Materialien: > 95 %

Anteil Materialien mit gültigem Klassifikationsschlüssel, z. B. eCl@ss

Anteil Freitextbestellungen: < 10 %

Hohes Niveau deutet auf unvollständige Stammdaten und mangelnde Katalogabdeckung hin

Praxisbeispiel

Ein mittelständischer Maschinenbauer mit mehreren Standorten stellte fest, dass dasselbe Kugellager unter vier verschiedenen Materialnummern geführt wurde – mit unterschiedlichen Beschreibungen und Preisen. Eine dublettenbereinigung über alle Werke senkte die Materialmasterzahl um ein Vielfaches und machte Bündelungspotenziale sichtbar. Begleitend führte das Unternehmen einen Datenqualitätsbericht ein, der quartalsweise Füllgrade und Dublettenraten auswies.

- Dublettenprüfung aller Materialstammsätze über Standortgrenzen hinweg
- Einführung verbindlicher Namenskonventionen und Pflichtfelder für Neuanlagen
- Freigabeprozess für Stammdatenänderungen mit Vier-Augen-Prinzip
- Quartalsweiser Datenqualitätsbericht als Steuerungsinstrument des Einkaufscontrolling

Fazit

Datenbereinigung ist im industriellen Einkauf kein einmaliges Projekt, sondern eine dauerhafte Aufgabe mit messbarem Nutzen. Saubere Material- und Lieferantenstammdaten ermöglichen Bündelung, automatisierte Prozesse und belastbare Kennzahlen. Entscheidend ist die Kombination aus methodischer Bereinigung, klarer Governance und kontinuierlicher Pflege im Regelbetrieb. Unternehmen, die ihre Datenbasis systematisch aufbereiten, schaffen die Grundlage für Sourcing-Strategien, Automatisierung und verlässliches Reporting. Der Aufwand amortisiert sich typischerweise über reduzierte Prozesskosten und bessere Entscheidungsgrundlagen.

Häufige Fragen

Was bedeutet Datenbereinigung im Einkauf?

Datenbereinigung bezeichnet die systematische Prüfung und Korrektur von Einkaufsdaten wie Material-, Lieferanten- und Konditionsstammsätzen. Dazu gehören das Entfernen von Dubletten, das Vervollständigen fehlender Felder und die Standardisierung von Beschreibungen. Ziel ist eine

Datenbasis, auf die Bestellprozesse, Auswertungen und Automatisierungen verlässlich aufbauen können.

Warum sind doppelte Materialstammsätze problematisch?

Dubletten fragmentieren Einkaufsvolumina: Der gleiche Artikel wird unter mehreren Nummern bei verschiedenen Lieferanten oder Preisen beschafft. Dadurch gehen Bündelungspotenziale verloren, Bestandsdaten werden verfälscht und Auswertungen liefern ein verzerrtes Bild. Auch Lagerbestände und Disposition leiden unter unklaren Materialidentitäten.

Wie oft sollte man Einkaufsstammdaten bereinigen?

In der Praxis haben sich halbjährliche bis jährliche Bereinigungszyklen für kritische Datenbereiche etabliert, ergänzt um kontinuierliche Pflege im Tagesgeschäft. Wichtiger als der Takt ist die Verhinderung neuer Fehler durch klare Erfassungsregeln und Freigabeprozesse. Ein laufender Datenqualitätsbericht zeigt, wo nachzusteuern ist.

Welche Kennzahlen messen Datenqualität im Einkauf?

Bewährt sind Füllgrad der Pflichtfelder, Dublettenrate, Anteil klassifizierter Materialien und Anteil katalogfähiger Positionen. Zusätzlich liefert der Anteil von Freitextbestellungen Hinweise auf Stammdatenlücken. Diese Kennzahlen lassen sich meist direkt aus dem ERP-System ziehen und periodisch auswerten.

Kann man Datenbereinigung automatisieren?

Teilweise ja: Dubletten-Matching und Klassifikationsvorschläge lassen sich heute gut automatisieren, besonders bei einfachen Materialien. Fachliche Freigaben, bei kritischen Teilen oder sensiblen Lieferantendaten, bleiben aber manuell. Automatisierung beschleunigt die Bereinigung großer Bestände, ersetzt aber nicht die fachliche Prüfung.