

Referenzdaten: Die Datenbasis für konsistente Prozesse im Einkauf

Referenzdaten sind standardisierte Stammdaten, die als einheitliche Bezugsbasis für Geschäftsprozesse, Systeme und Berichte dienen. Im technischen Einkauf umfassen sie insbesondere Materialstammdaten, Lieferantenstammdaten, Klassifizierungssysteme sowie Preis- und Mengenreferenzen. Ohne saubere Referenzdaten funktionieren weder automatisierte Bestellprozesse noch belastbare Auswertungen. Dieser Artikel erklärt Aufbau, Qualitätssicherung und praxisnahe Nutzung von Referenzdaten im Beschaffungscontrolling.

Das Wichtigste in Kürze

- Referenzdaten sind standardisierte, häufig genutzte Stammdaten wie Materialnummern, Lieferantenstammsätze und Klassifizierungen, die als gemeinsame Bezugsbasis verschiedener Systeme dienen.
- Sie unterscheiden sich von Transaktionsdaten, die einzelne Vorgänge wie Bestellungen oder Wareneingänge abbilden.
- Häufig genutzte Standards sind UNSPSC, eCl@ss und GLN für die weltweite, eindeutige Identifikation von Materialien und Geschäftspartnern.
- Typische Qualitätskriterien sind Vollständigkeit, Eindeutigkeit, Aktualität und Konsistenz über alle angeschlossenen Systeme hinweg.
- Schlechte Referenzdaten führen in der Praxis zu Doppelbestellungen, Fehlteilen, falschen Preisen und hohen manuellen Nacharbeiten.
- Ein Master-Data-Management mit klar definierten Verantwortlichkeiten, sogenannten Data Ownern, ist die Grundlage dauerhaft sauberer Datenbestände.

Definition: Referenzdaten

Referenzdaten sind Stammdaten mit hohem Wiederverwendungsgrad, die als Referenz für operative Geschäftsprozesse dienen. Sie beschreiben dauerhaft gültige Objekte wie Materialien, Dienstleistungen, Lieferanten oder Kostenstellen und ändern sich im Vergleich zu Bewegungsdaten nur selten. Im Gegensatz dazu entstehen Transaktionsdaten bei jedem einzelnen Geschäftsvorgang, etwa bei einer Bestellung oder einem Wareneingang.

Abgrenzung zu Stammdaten und Transaktionsdaten

Der Begriff Stammdaten ist in der Praxis weiter gefasst und umfasst alle objektbezogenen Grunddaten. Referenzdaten sind die Teilmenge, die als einheitlicher Bezugspunkt über System- und Unternehmensgrenzen hinweg genutzt wird, etwa eine Materialnummer, die in ERP, Katalogsystem und EDI-Kommunikation identisch auftaucht. Transaktionsdaten wie Bestellungen oder Rechnungen verweisen stets auf diese Referenzdaten und sind ohne sie nicht eindeutig interpretierbar.

Zentrale Typen im Einkauf

Für Einkäufer sind vor allem vier Datenarten relevant: Materialstammdaten mit Beschreibung, Maßeinheit und Klassifizierung, Lieferantenstammdaten mit Identifikatoren und Bonitätsdaten, Kontrakt- und Preisreferenzen sowie technische Referenzdaten wie Zeichnungsnummern und Normen. Ergänzend spielen Referenzdaten für Nachhaltigkeitsangaben eine wachsende Rolle, etwa zur Zuordnung von Product Carbon Footprint (PCF)-Werten auf Materialebene.

Aufbau und Standards: Wie Referenzdaten strukturiert werden

Damit Referenzdaten zwischen Systemen, Abteilungen und Geschäftspartnern interpretierbar bleiben, benötigen sie eine einheitliche Struktur. Diese umfasst eindeutige Identifikatoren, kontrollierte Vokabulare für Bezeichnungen sowie Klassifizierungssysteme zur Gruppierung. In der Praxis hat sich ein gestuftes Vorgehen bewährt: erst Identifikation, dann Klassifizierung, dann Anreicherung mit attribuierten Fachdaten.

Klassifizierungssysteme

International verbreitete Standards sind UNSPSC (United Nations Standard Products and Services Code) für Waren und Dienstleistungen sowie eCl@ss, das im deutschsprachigen Maschinen- und Anlagenbau wegen seiner technischen Merkmale häufig genutzt wird. Ein harmonisiertes Klassifizierungssystem ermöglicht Spend-Analysen über Werke und Standorte hinweg und ist Grundlage für aussagekräftiges Benchmarking im Einkauf.

Identifikatoren und Stammsatzlogik

Jedes referenzierte Objekt benötigt einen eindeutigen Schlüssel: Materialnummern im eigenen System, Lieferantenummern, GLN (Global Location Number) für Standorte im EDI-Austausch sowie GTIN für handelsübliche Artikel. Wichtig ist die sogenannte Golden-Record-Logik: Es gibt genau einen führenden Datensatz pro Objekt, aus dem abgeleitete Systeme gespeist werden. Doppel- und Alias-Sätze werden systematisch erkannt und zusammengeführt.

Datenmodell und Pflichtfelder

Ein praxistaugliches Datenmodell definiert Pflichtfelder, Wertebereiche und Validierungsregeln je Objekttyp. Für Materialstammdaten gehören dazu typischerweise Beschreibung nach Namenskonvention, Maßeinheit, Warengruppe, Einkaufs- und Verkaufsorganisationssichten sowie technische Merkmale. Für Lieferantenstammsätze zählen Bankverbindung, Zahlungsbedingungen, Zertifizierungen und Sperrklauseln. Je vollständiger die Pflichtfeldersetzung beim Anlegen geprüft wird, desto geringer ist der spätere Korrekturaufwand.

Methoden und Vorgehen: Referenzdaten aufbauen und pflegen

Der Aufbau sauberer Referenzdaten folgt einem wiederkehrenden Zyklus aus Bereinigung, Harmonisierung und kontinuierlicher Pflege. In der Praxis starten viele Unternehmen mit einem Deduplizierungsprojekt im Materialstamm, gefolgt von der Einführung einer zentralen Anlegestelle mit definierten Rollen. Wichtig ist dabei, Datenqualität nicht als einmaliges Projekt, sondern als laufenden Prozess zu organisieren.

Datenbereinigung und Deduplizierung

Am Anfang steht typischerweise eine Bestandsanalyse: Wie viele Materialien gibt es, wie viele davon

sind duplikativ, wie viele sind seit Jahren ungenutzt? In der Praxis zeigen Bestandsanalysen häufig, dass ein erheblicher Teil der Materialstammsätze inaktiv oder doppelt angelegt ist.

Bereinigungswerkzeuge erkennen Dubletten über Fuzzy-Suche, Normalisierung von Bezeichnungen und Vergleich technischer Merkmale. Die Freigabe zusammengeführter Sätze sollte durch fachliche Data Owner erfolgen.

Governance und Data Ownership

Referenzdatenpflege braucht klare Zuständigkeiten: Wer legt an, wer prüft, wer gibt frei, wer archiviert? Bewährt hat sich ein zentrales Master-Data-Team mit definierten Service-Level-Agreements für Bearbeitungszeiten, ergänzt durch fachliche Data Owner in den Einkaufs- und Technologiebereichen. Änderungsprozesse müssen dokumentiert sein, insbesondere bei preis- oder compliance-relevanten Feldern. Ein Einkaufsleitfaden kann hier verbindliche Namenskonventionen und Anlageprozesse beschreiben.

Schnittstellen und EDI

Referenzdaten entfalten ihren vollen Nutzen erst, wenn sie konsistent zwischen Systemen ausgetauscht werden. Über EDI (Electronic Data Interchange) werden Stammdaten, Bestellungen und Rechnungen automatisiert abgeglichen; Referenzdaten wie Lieferantennummern und Artikelidentifikatoren müssen dafür auf beiden Seiten identisch gepflegt sein. In der Praxis hat sich bewährt, Änderungen an Referenzdaten versioniert zu übertragen und Abweichungen automatisch zu melden, statt sie manuell nachzupflegen.

Wichtige KPIs zur Datenqualität

Datenqualität lässt sich messen. In der Praxis haben sich einige Kennzahlen etabliert, die regelmäßig – etwa monatlich – ausgewertet werden. Sie machen Fortschritte sichtbar und helfen, Ressourcen für die Datenpflege zu rechtfertigen.

Vollständigkeit und Eindeutigkeit

Die Vollständigkeitsquote gibt an, wie viel Prozent der Stammsätze alle Pflichtfelder befüllt haben. Die Dublettenquote misst den Anteil doppelt angelegter Objekte. Beide Werte sollten zusammen betrachtet werden, denn eine hohe Vollständigkeit nützt wenig, wenn derselbe Artikel unter fünf Nummern geführt wird.

Aktualität und Konsistenz

Die Aktualitätskennzahl misst, wie viele Datensätze innerhalb eines definierten Zeitraums geprüft oder bestätigt wurden. Konsistenzprüfungen vergleichen Feldinhalte zwischen Quellsystemen, etwa ob der Lieferantename in ERP und E-Katalog identisch ist. Diskrepanzen werden typischerweise in einer Risikomatrix nach Schwere und Häufigkeit bewertet, um Korrekturprioritäten abzuleiten.

Risiken und Kontrollen: Was schlechte Referenzdaten kosten

Fehlerhafte Referenzdaten wirken sich direkt auf operative und finanzielle Ergebnisse aus. Typische Folgen sind Doppelbestellungen, Fehlbuchungen, verzögerte Zulieferprozesse und mangelhafte Compliance. Kontrollen sollten daher präventiv beim Anlegen und detektiv im laufenden Betrieb ansetzen.

Operative Risiken

Doppelte Materialnummern führen zu unnötigen Bestellungen und Lagerbeständen; unklare Einheiten verursachen Fehlmengen. Wenn Referenzdaten zur Lieferantenbewertung fehlerhaft sind, werden Lieferantenleistungen falsch bewertet – etwa wenn Parts per Million (PPM)-Werte wegen falscher Materialzuordnung verzerren. Im Qualitätskontext behindern unvollständige Stammdaten auch die Nachverfolgbarkeit von Materialien, die gerade in regulierten Branchen verpflichtend ist.

Finanzielle und Compliance-Risiken

Im Three-Way-Match zwischen Bestellung, Wareneingang und Rechnung führen abweichende Referenzdaten zu Abweichungen und manuellen Klärungen. Veraltete Preisreferenzen können zu Überzahlungen führen; fehlende Sperrkennzeichen im Lieferantenstamm ermöglichen Zahlungen an gesperrte oder betrügerische Empfänger. Compliance-relevante Daten wie eine Restricted Substances List (RSL) müssen lückenlos auf Materialebene referenziert sein, sonst drohen Lieferkettenverstöße.

Kontrollinstrumente

Bewährte Kontrollen sind Pflichtfeld-Validierung beim Anlegen, vier-Augen-Freigabe bei sensiblen Feldern, periodische Dubletten- und Inaktivitätsscan sowie Abstimmungsreporte zwischen Systemen. Änderungsprotokolle und regelmäßige Stichproben durch interne Auditrunden ergänzen das Kontrollsystem. Wichtig ist, dass Kontrollbefunde in einen Verbesserungsprozess münden, statt als Einzelmeldungen zu versanden.

Aktuelle Entwicklungen: Automatisierung und Nachhaltigkeitsdaten

Zwei Trends verändern den Umgang mit Referenzdaten derzeit spürbar: die automatisierte Datenpflege durch KI und wachsende Nachhaltigkeits- und Regulatory-Anforderungen. Beide erhöhen den Anspruch an Datenqualität und Struktur deutlich.

KI-gestützte Klassifizierung und Deduplizierung

Sprachmodelle und Matchingsalgorithmen klassifizieren freitextliche Artikelbeschreibungen heute weitgehend automatisch in Standardsysteme wie UNSPSC oder eCl@ss. In der Praxis verkürzt das Anlageprozesse deutlich, erfordert aber weiterhin menschliche Freigabe, insbesondere bei technischen Angaben. KI-gestützte Dublettenerkennung arbeitet über Fuzzy-Matching hinaus und erkennt semantisch ähnliche Bezeichnungen auch über Sprachgrenzen hinweg.

Nachhaltigkeits- und Lieferkettendaten

Berichtspflichten zur Lieferkette und zu Emissionswerten verlangen maschinenlesbare Referenzdaten für Lieferanten, Materialien und Herkunftsländer. Die Zuordnung von PCF-Werten auf Materialebene setzt voraus, dass Materialstammdaten konsistent klassifiziert sind. Auch die Nachverfolgbarkeit von Materialien, etwa bei Konfliktmineralien oder Rezyklatanteilen, wird über referenzierte Chargen- und Lieferantendaten abgebildet. Unternehmen, die ihre Referenzdatenbasis früh harmonisieren, schaffen hier einen strukturellen Vorteil.

Typische Kennzahlen

Vollständigkeitsquote: > 95 %

Anteil der Stammsätze mit vollständig befüllten Pflichtfeldern

Dublettenquote: < 2 %

Anteil doppelt angelegter Objekte am Gesamtbestand

Aktualitätsquote: > 90 %

Anteil der Stammsätze, die im definierten Zeitraum geprüft oder bestätigt wurden

Automatisierungsgrad EDI: > 80 %

Anteil der ohne manuelle Korrektur verarbeitbaren Stammdatenübertragungen

Praxisbeispiel

Ein mittelständischer Maschinenbauer mit mehreren Werken erkennt bei einer Spend-Analyse, dass dasselbe Kugellager unter sechs verschiedenen Materialnummern und Bezeichnungen geführt wird. Als Folge werden Mengenstaffeln nicht ausgeschöpft, Bestände doppelt aufgebaut und Lieferantenanfragen enthalten widersprüchliche Angaben. Das Unternehmen startet ein Master-Data-Projekt mit folgendem Vorgehen:

- Dublettenanalyse aller aktiven und inaktiven Materialstammsätze mit anschließender Zusammenführung durch fachliche Data Owner.
- Einführung einer zentralen Anlegestelle mit Namenskonvention, Pflichtfeldern und eCI@ss-Klassifizierung als Pflichtangabe.
- Aufbau eines monatlichen Datenqualitätsreports mit Vollständigkeits-, Dubletten- und Aktualitätskennzahlen.
- Abgleich der bereinigten Materialreferenzen mit Lieferantenkatalogen und EDI-Schnittstellen, inklusive Korrektur veralteter Preisreferenzen vor der nächsten Verhandlungsrunde.

Fazit

Referenzdaten sind das unsichtbare Fundament jeder effizienten Beschaffung. Sie entscheiden darüber, ob Spend-Analysen belastbar sind, ob Automatisierung über EDI funktioniert und ob Compliance-Anforderungen lückenlos nachvollzogen werden können. Der Aufwand für Bereinigung und Governance amortisiert sich typischerweise rasch durch reduzierte Doppelarbeiten, bessere Verhandlungspositionen und weniger Fehlteile. Entscheidend ist, Datenqualität als dauerhaften Prozess mit klaren Verantwortlichkeiten zu organisieren, nicht als einmaliges Projekt. Wer seine Referenzdatenbasis in Ordnung hat, schafft die Voraussetzung für digitale Beschaffung, Nachhaltigkeitsberichterstattung und belastbares Einkaufscontrolling.

Häufige Fragen

Was sind Referenzdaten einfach erklärt?

Referenzdaten sind standardisierte Stammdaten, die als einheitliche Bezugsbasis für Geschäftsprozesse und Systeme dienen. Dazu zählen etwa Materialnummern, Lieferantenstammsätze und Klassifizierungen. Sie ändern sich selten und werden von vielen Vorgängen wie Bestellungen und Rechnungen referenziert.

Was ist der Unterschied zwischen Referenzdaten und Transaktionsdaten?

Referenzdaten beschreiben dauerhaft gültige Objekte wie Materialien oder Lieferanten und ändern sich selten. Transaktionsdaten entstehen bei jedem einzelnen Geschäftsvorgang, etwa einer Bestellung oder Rechnung, und verweisen auf Referenzdaten. Ohne konsistente Referenzdaten sind Transaktionsdaten

nicht eindeutig auswertbar.

Welche Standards gibt es für Referenzdaten im Einkauf?

Verbreitete Standards sind UNSPSC und eCl@ss für die Klassifizierung von Waren und Dienstleistungen sowie GLN und GTIN für die Identifikation von Standorten und Artikeln. Im deutschsprachigen Maschinenbau wird häufig eCl@ss wegen seiner technischen Merkmale genutzt. Die Wahl des Standards sollte konzernweit einheitlich erfolgen.

Wie misst man die Qualität von Referenzdaten?

Typische Kennzahlen sind Vollständigkeitsquote (Anteil der Stammsätze mit befüllten Pflichtfeldern), Dublettenquote, Aktualitätsquote und Konsistenz zwischen Systemen. Diese Werte werden in der Praxis regelmäßig, etwa monatlich, ausgewertet. Sie machen Fortschritte messbar und begründen Ressourcen für die Datenpflege.

Was kostet schlechte Datenqualität im Einkauf?

Folgen sind Doppelbestellungen, unnötige Lagerbestände, manuelle Klärungen im Rechnungsabgleich und falsche Spend-Analysen. Auch Verhandlungsergebnisse leiden, weil Mengen und Volumina nicht korrekt aggregiert werden können. In der Praxis entstehen dadurch nennenswerte laufende Kosten, die sich durch Datenbereinigung deutlich reduzieren lassen.

Wer ist für die Pflege von Referenzdaten verantwortlich?

Bewährt ist ein Modell mit zentralem Master-Data-Team für Prozess und Systematik sowie fachlichen Data Ownern in den Einkaufs- und Fachbereichen. Diese geben Änderungen freigaben und tragen inhaltliche Verantwortung. Klare Rollen und definierte Bearbeitungszeiten sind entscheidend für dauerhafte Datenqualität.