

Leitfaden HR-Datenmodellierung

Wie HR-Daten für Analytics strukturiert werden — inkl. Historisierung und Privacy-by-Design
HR Business Transparency Consulting

1. Warum HR-Datenmodelle brechen

Ein Modell, das „Mitarbeiter gehört zu Kostenstelle“ als einfache Fremdschlüsselbeziehung abbildet, funktioniert exzellent — bis zur nächsten Reorganisation. Wird eine Kostenstelle umbenannt, zusammengelegt oder einer anderen übergeordneten Einheit zugeordnet, verändert sich diese eine Beziehung rückwirkend für alle historischen Daten, die auf sie verweisen. Ein Bericht mit Stand „Mai 2026“, der im August 2026 erneut aufgerufen wird, zeigt plötzlich eine andere Organisationsstruktur als im Mai tatsächlich galt — nicht, weil sich die Mai-Zahlen geändert hätten, sondern weil das Modell keine Vergangenheit kennt, nur einen aktuellen Zustand.

Der zweite häufige Bruchpunkt ist Datenschutz, der nachträglich in ein fertiges Modell hineingepresst wird — eine Row-Level-Security-Regel hier, eine Filterung dort —, statt von Anfang an Teil der Modellierung zu sein. Beides lässt sich vermeiden, wenn zwei Prinzipien von Beginn an eingeplant werden: Historisierung über Slowly Changing Dimensions und Privacy-by-Design auf Modellebene, nicht auf Report-Ebene.

Ein dritter, oft übersehener Bruchpunkt liegt noch vor der eigentlichen Modellierung: Ein Modell kann nur so gut sein wie das Verständnis der Rohdaten, die in es einfließen. Wer die Herkunft, Struktur und Aktualisierungslogik seiner Quellsysteme nicht kennt, modelliert an der Realität vorbei. Das folgende Kapitel schafft diese Grundlage, bevor Abschnitt 3 zeigt, wie daraus Fakten- und Dimensionstabellen werden.

2. HR-Datenquellen im Überblick

Bevor eine Faktentabelle oder Dimension entworfen werden kann, muss klar sein, aus welchen Quellsystemen die zugrundeliegenden Daten stammen, wie strukturiert sie sind und wie häufig sie sich ändern. HR-Datenquellen lassen sich in vier Kategorien einteilen, die sich in ihrer Standardisierung, Aktualisierungsfrequenz und Governance-Anforderung deutlich unterscheiden.

2.1 HR-Systemdaten (HRIS/ATS/LMS)

Kerndaten aus dem Human Resources Information System (z. B. Oracle, SAP, Workday) sowie angebundenen Systemen wie Bewerbermanagement (ATS) und Lernmanagement (LMS). Diese Quellen sind in der Regel am stärksten standardisiert und bilden das Rückgrat der meisten Faktentabellen und Dimensionen aus Abschnitt 3.

- **Recruiting-Daten** — Bewerberzahlen, Lebensläufe, Recruiting-Funnel und -Quellen.
- **Demografische Daten** — Mitarbeiter-ID, Geschlecht, Geburtsdatum, Wohnort, Position, Abteilung, Kostenstelle, Austrittsdatum; dienen häufig als Kontrollvariablen und liefern zugleich die Kandidaten für eine SCD-Typ-2-Historisierung (siehe Abschnitt 4).
- **Abwesenheitsdaten** — Krankheitstage, Elternzeit, Verspätungen.
- **Performance-Management-Daten** — Beurteilungen und Leistungsbewertungen.
- **Lernmanagement-Daten** — Kursangebote und Lernfortschritt (nicht alle Lerndaten liegen im LMS, z. B. externe Weiterbildungskosten im Finanzsystem).
- **Stellenarchitektur (Job Architecture)** — Grundlage der Vergütung; Rollen in Gehaltsbänder/Grade gruppiert. Hilft z. B. bei Fragen zu Gehaltskonsistenz, Aufstiegschancen oder Fluktuation je Grade.
- **Vergütung & Benefits** — Gehaltspakete und Zusatzleistungen, meist im HR- oder Payroll-System.
- **Nachfolgeplanung** — Daten zu Führungskräfteentwicklung und Nachfolgekandidaten.

- **Talententwicklung** — Programme zur internen Talentförderung (Leadership-Tracks, High-Potential-Listen, Mentoring-Teilnahme).
- **Austrittsgespräche (Exit Interviews)** — Gründe für Kündigungen; wertvoll zur Reduzierung von Fluktuation.

2.2 Sonstige HR-Daten

Daten außerhalb des HRIS, oft weniger standardisiert, aber wichtig für den Kontext. Für die Modellierung bedeutet das: uneinheitliche Schlüssel, wechselnde Formate und häufig manuelle Konsolidierung, bevor diese Daten überhaupt an eine geteilte Dimension andocken können.

- **Reisedaten** — Kontext zu Rollen und Arbeitsmustern, meist im Finanz-/Reisebuchungssystem.
- **Mitarbeiter- & Engagement-Befragungen** — Oft durch Dritte erhoben, um Anonymität zu gewährleisten — eigene Datenquelle mit eigener Struktur.
- **Wellbeing/Gesundheit & Benefits** — Teilnahmedaten an Gesundheitsprogrammen sowie die Inanspruchnahme von sachbezogenen, brutto-/netto-optimierten Zusatzleistungen (z. B. Jobrad, Essensgutscheine, Kita-Zuschüsse, betriebliche Altersvorsorge oder Mobilitätsbudgets). Diese Daten liegen meist verteilt bei externen Benefit-Dienstleistern, Entgeltabrechnungs-Schnittstellen oder in Fachabteilungen und nicht direkt im HRIS.

2.3 Geschäftsdaten

Outcome-Daten aus dem operativen Geschäft, die HR-Kennzahlen erst mit geschäftlicher Wirkung verknüpfen.

- **CRM-Daten** — Kundenkontakte, NPS-Werte — als Outcome-Daten zur Messung der Wirkung von HR-Maßnahmen auf kundennahe Mitarbeitende.
- **Finanzdaten** — L&D-Ausgaben, Personalkosten, ROI-Berechnungen.
- **Produktionsdaten** — Betriebskennzahlen (Terminplanung, Servicequote, Durchlaufzeiten) als Outcome-Daten für Personalpolitik.
- **Vertriebsdaten** — z. B. Umsatz je Filiale zur Messung des Effekts von HR-Maßnahmen wie Trainingsprogrammen.

2.4 Externe Datenquellen

- **Abwesenheits-Benchmarks** (z.B. TK-Benchmarktool)
- **Arbeitsmarkt- und Gehaltsbenchmarks** — z. B. Glassdoor, Payscale, Bureau of Labor Statistics.
- **Wettbewerbs- und Branchenberichte** — z. B. McKinsey, SHRM, Deloitte.
- **Behörden- und Compliance-Datenbanken** — z. B. EEOC, OSHA, IRS.
- **Jobbörsen und Recruiting-Plattformen** — z. B. LinkedIn, Indeed, ZipRecruiter — liefern Signale zu Nachfrage und Marktstandards.

Für die Modellierung heißt diese Vielfalt: Nicht jede Quelle wird zu einer eigenen Faktentabelle. HR-Systemdaten liefern meist die geteilten Dimensionen (Datum, Kostenstelle, Mitarbeiter) und die zentralen Fakten. Sonstige HR-Daten, Geschäfts- und externe Daten werden dagegen häufig erst über gemeinsame Schlüssel — Datum, Kostenstelle, Mitarbeiter-ID — an dieses Kernmodell angebunden, nicht als gleichrangige eigene Sternschemata.

2.5 Erfolgsfaktoren für gute HR-Daten

Erfolgsfaktor	Worauf es ankommt
Daten zentralisieren	HRIS als Single Source of Truth; ATS, LMS, Payroll und Performance - Management-System (PMS) integrieren.
Compliance und Datensicherheit sicherstellen	Rollenbasierte Zugriffsrechte, Verschlüsselung, Backups — Grundlage für das Privacy-by-Design aus Abschnitt 6.
Datenkompetenz aufbauen	Im HR-Team Verständnis für Datenstruktur und -qualität verankern, nicht nur bei Analytics-Spezialisten.
Klare Datenverantwortung und Governance festlegen	Data Stewards und Zuständigkeiten je Funktion definieren.
Datenstrategien kontinuierlich verbessern	Redundante Quellen eliminieren, statt sie parallel weiterzuführen.
Datenqualität sicherstellen	Vollständig, fehlerfrei, aktuell — die Voraussetzung für jede Historisierung nach SCD Typ 2.

3. Grundprinzip: Sternschema mit geteilten Dimensionen

Die Grundstruktur folgt demselben Muster wie im Architektur-Dokument zum HR-KPI-Cockpit dieser Reihe: Faktentabellen (Personalkosten, Fehlzeiten, Fluktuation, Recruiting) hängen an geteilten Dimensionen (Datum, Kostenstelle, Mitarbeiter). Was in diesem Leitfaden neu hinzukommt, ist die Frage, was passiert, wenn sich eine dieser Dimensionen selbst verändert — und genau das passiert bei einer Reorganisation zwangsläufig.

Die einfachste, aber folgenreichste Modellierungsentscheidung: Verwendet die Faktentabelle die fachliche Kostenstellen-ID (z. B. „KST-042“) oder einen technischen Surrogatschlüssel als Fremdschlüssel? Nur der Surrogatschlüssel macht Historisierung überhaupt möglich — dazu mehr in Abschnitt 4.

4. Historisierung: Slowly Changing Dimensions

Eine „Slowly Changing Dimension“ (SCD) ist eine Dimensionstabelle, die Änderungen an ihren eigenen Attributen nicht überschreibt, sondern als neue, zeitlich gültige Version speichert. Für HR-Daten ist das keine akademische Feinheit, sondern die einzige Methode, mit der eine Kostenstellen-Umbenennung, ein Wechsel der Führungskraft oder eine Bereichszusammenlegung korrekt abgebildet werden kann.

Typ	Verhalten bei Änderung	Wann verwenden
SCD Typ 1	Alter Wert wird überschrieben, keine Historie	Korrektur echter Fehler (z. B. Tippfehler im Namen), nie für organisatorische Änderungen
SCD Typ 2	Neue Zeile mit neuem Surrogatschlüssel, Gültig-von/-bis und Aktuell-Flag	Standard für Kostenstellen, Positionen, Führungslinien — alles, was sich strukturell ändert
SCD Typ 3	Alter Wert bleibt in einer Zusatzspalte erhalten, nur ein Vorzustand	Selten in HR — nur wenn ausschließlich der unmittelbar vorherige Zustand relevant ist

In der Praxis ist SCD Typ 2 der Standard für HR-Dimensionen. Die Tabelle bekommt zusätzlich zur fachlichen ID einen technischen Surrogatschlüssel sowie drei Steuerspalten:

Spalte	Beispielwert (vor Reorg)	Beispielwert (nach Reorg)
Surrogatschlüssel (SK)	1001	1002
Kostenstellen-ID (fachlich, bleibt gleich)	KST-042	KST-042
Name	Vertrieb Nord	Vertrieb (zusammengelegt)
Übergeordnete Einheit	Vertrieb	Geschäftsleitung
Gültig von	01.01.2020	01.07.2026
Gültig bis	30.06.2026	31.12.9999
Aktuell	Nein	Ja

Der entscheidende Mechanismus: Faktentabellen referenzieren immer den Surrogatschlüssel, der zum Zeitpunkt des Ereignisses gültig war — nicht die fachliche ID und nicht den heutigen Stand. Eine Personalkosten-Zeile aus Mai 2026 verweist auf SK 1001 (Vertrieb Nord) und bleibt das für immer, unabhängig davon, wie oft sich KST-042 danach noch verändert.

5. Der Reorganisations-Fall im Bild

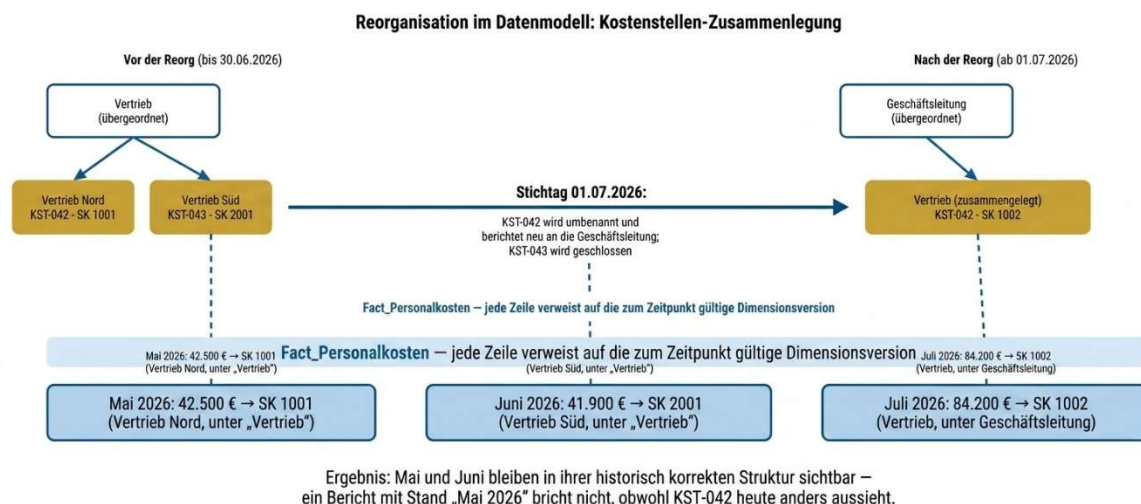


Abbildung 1: Zwei Kostenstellen werden zusammengelegt — historische Fakten bleiben trotzdem korrekt zugeordnet.

Vertrieb Nord (KST-042) und Vertrieb Süd (KST-043) werden zum 1. Juli 2026 zu einer neuen Kostenstelle „Vertrieb“ zusammengelegt, die künftig direkt an die Geschäftsleitung berichtet. Die fachliche ID KST-042 bleibt erhalten, bekommt aber einen neuen Surrogatschlüssel (SK 1002) und eine neue Gültig-von/-bis-Spanne. Die Mai- und Juni-Fakten bleiben unverändert an SK 1001 und SK 2001 hängen — ein Bericht mit historischem Stichtag zeigt weiterhin exakt die Organisationsstruktur, die zum jeweiligen Zeitpunkt galt.

Ohne SCD Typ 2 gäbe es nur einen Datensatz für KST-042, dessen „Übergeordnete Einheit“ beim Reorg-Update von „Vertrieb“ auf „Geschäftsleitung“ überschrieben würde — und plötzlich hätten auch die Mai- und Juni-Zahlen rückwirkend die neue, damals noch gar nicht existierende Berichtslinie.

6. Privacy-by-Design

Datenschutz, der erst im Report nachgerüstet wird, ist ein Report-Feature, kein Modell-Merkmal — und wird bei jedem neuen Bericht erneut vergessen. Die folgenden vier Prinzipien gehören ins Datenmodell selbst, nicht in jedes einzelne Dashboard erneut hinein.

Prinzip	Umsetzung im Modell
Datensparsamkeit	Nur Attribute aufnehmen, die für den definierten Analysezweck tatsächlich gebraucht werden — kein „das könnte später mal nützlich sein“
Pseudonymisierung	MitarbeiterID als Surrogatschlüssel ohne erkennbaren Bezug zum Namen; Klardaten in einer separat berechtigten Tabelle, nicht im Analytics-Modell
Minimum-Group-Size-Guard	Als DAX-Measure oder SQL-View-Logik direkt im Modell verankert — siehe Abschnitt 6 des HR-KPI-Cockpit-Architekturdokuments dieser Reihe
Löschkonzept	Aufbewahrungsfristen als Attribut der Dimension selbst modelliert (z. B. „Löschen ab“-Spalte), nicht als manueller Jahresprozess außerhalb des Modells

Ein Modell, das Klardaten (Name, Geburtsdatum, Kontaktdaten) und Analysedaten (Kostenstelle, Vertragsstunden, Fehltag) in getrennten Tabellen mit getrennten Zugriffsrechten hält, macht „wer darf was sehen“ zu einer Frage der Tabellenberechtigung statt einer Frage, die in jedem einzelnen Report neu beantwortet werden muss.

7. Checkliste: ein Modell, das die nächste Reorg übersteht

- ☐ Alle relevanten Datenquellen aus Abschnitt 2 identifiziert und ihrer jeweiligen Kategorie zugeordnet
- ☐ Alle Dimensionen, die sich strukturell ändern können (Kostenstelle, Position, Führungslinie), als SCD Typ 2 angelegt
- ☐ Faktentabellen verweisen auf Surrogatschlüssel, nie auf die fachliche ID direkt
- ☐ Gültig-von/-bis und Aktuell-Flag für jede historisierte Dimension vorhanden und gepflegt
- ☐ Ein dokumentierter Prozess für den Reorg-Fall: wer legt neue Surrogatschlüssel an, wann?
- ☐ Klardaten und Analysedaten in getrennten Tabellen mit getrennten Zugriffsrechten
- ☐ Minimum-Group-Size-Guard als Modell-Logik hinterlegt, nicht nur in einzelnen Reports
- ☐ Löschrufen als Modellattribut definiert, nicht als separater manueller Jahresprozess
- ☐ Ein historischer Testfall (z. B. ein vergangener Reorg) durchgespielt, bevor das Modell produktiv geht

Der letzte Punkt ist der wichtigste: ein Modell, das nie gegen einen echten Reorg-Fall getestet wurde, wird beim ersten echten Fall überraschen — im schlechtesten Moment, mit der Geschäftsführung im Publikum.