

Datenmigration: Risiko und Chance bei der ERP-Einführung

Matthias Knapp, Knapp Data Quality

Bei der Einführung und Umstellung von ERP-Systemen hängt der Projekterfolg maßgeblich von der Datenmigration ab. Die vorhandenen Datenbestände frühzeitig und gründlich zu analysieren und aufzubereiten, senkt aber nicht nur die Risiken, sondern eröffnet auch ganz neue Perspektiven. Der folgende Beitrag zeigt auf, wie Unternehmen mit guter Vorbereitung ein Quantensprung gelingt.

Der zentrale Wissensspeicher der meisten Unternehmen sind ihre ERP-Systeme. Sie bilden das Rückgrat aller betrieblichen Prozesse und Abläufe und das Fundament aller Auswertungen. Die wichtigsten Informationen sämtlicher Geschäftsobjekte und aller betrieblichen Aktivitäten werden im ERP-System verwaltet und gespeichert. Zum einen sind dies die Stammdaten, zum Beispiel Produkte und Leistungen, Geschäftspartner wie Kunden oder Lieferanten, Mitarbeiter, aber auch Werke, Anlagen, Arbeitsplätze und Lager. Zum anderen sind dies Bewegungsdaten, also beispielsweise Bestellungen und Aufträge mit ihren Bestell- bzw. Auftragspositionen, Fertigungsaufträge etc. Die meisten dieser Informationen werden regelmäßig in Auftragsabwicklungs-, Planungs- und Berichtsprozessen benötigt, kurz: für Entscheidungen genutzt. Daher ist die Nutzbarkeit von Daten ein wichtiges Kriterium für ihre Qualität.

Die Umstellung von ERP-Systemen bringt allerdings große Veränderungen mit sich. Dabei betreffen die Neuerungen bei weitem nicht nur den Funktionsumfang oder die User Experience, auch wenn dies in der Regel die entscheidenden Kriterien im Auswahlprozess sind. Die wesentlich tiefgreifenderen Veränderungen bleiben den Usern meist verborgen. Sie betreffen Datenmodelle und -strukturen weit unter der Oberfläche der Anwendungen. Und dennoch sind eben diese Daten entscheidend für die Leistungsfähigkeit, die Zuverlässigkeit und Akzeptanz der künftigen Anwendung.

Datenmigration bedeutet immer: Transformation der Daten

Eine große Herausforderung bei der Umstellung von ERP-Systemen besteht darin, dass die Daten in Quell- und Zielsystem unterschiedlich strukturiert sind. Bei der Umstellung müssen die Daten entsprechend transformiert werden. Ein simples Beispiel: Ein bisheriger Outlook-Kontakt soll als Kunde in ein ERP-System übertragen werden. Das Objekt „Kontakt“ umfasst aber Informationen zu zwei unterschiedlichen Datenobjekten: Angaben zur Firma (und ihrer Adresse) sowie Angaben zu einer Person (mit E-Mail-Adresse, Geburtsdatum etc.). Dementsprechend muss das Quellobjekt in zwei Zielobjekte aufgeteilt werden. Weitere Zuordnungen und Umwandlungen erfolgen auf Feld- und Werteebene: Das Feld „Website“ wird dem Zielfeld „URL“ zugeordnet, der Inhalt des Felds „Straße (geschäftlich)“ wird zerlegt und übertragen in „Adresse“ und „Hausnummer“. Die Logik der Übertragung von Quell- zu Zielfeldern bildet das sogenannte „Mapping“, den Kern der Datenmigration.

Besonders herausfordernd wird die Übertragung der Daten immer dann, wenn die Daten im Quellsystem schlecht oder uneinheitlich strukturiert sind, zum Beispiel, wenn in einem Feld unterschiedliche Informationen eingetragen sind. Eine eindeutige Zuordnung zu den Feldern im Zielsystem ist dann nicht möglich. Die Folge sind aufwendige, fehleranfällige Übertragungen von Hand oder nachträgliche Korrekturen im Zielsystem, oder alternativ die Programmierung komplexer Regeln. Ebenso kritisch sind fehlende oder fehlerhafte Daten. Schließlich kann nur übertragen werden, was auch vorhanden ist – Lücken und Inkonsistenzen richten dann im Zielsystem weiterhin den gleichen Schaden an wie zuvor.

Gerade bei „historisch gewachsenen“ Daten sollte der Aufwand für die Datenübernahme nicht unterschätzt werden. Ein frühzeitiges Assessment der Datenqualität kann hier helfen, den Umstellungsaufwand abzuschätzen und Risiken für das Projekt rechtzeitig zu erkennen.

Kein alter Wein in neuen Schläuchen

Die Datenmigration hat das Ziel, alle relevanten Informationen in das Zielsystem zu übertragen. Relevant sind in diesem Sinne diejenigen Daten, die für die Funktion des Zielsystems notwendig sind. Gleichzeitig dürfen aber auch keine wichtigen Informationen aus dem Quellsystem verloren gehen. Hier schlägt die Stunde der Fachbereiche. Sie allein können und müssen entscheiden, welche Informationen für die betrieblichen Prozesse unverzichtbar sind. Gleichzeitig sind sie diejenigen, die veraltete oder nicht länger benötigte Daten zur Archivierung kennzeichnen und so dazu beitragen, dass nur aktive und gut gepflegte Daten migriert werden.

Und eine weitere Aufgabe kommt den Fachbereichen zu: Entwurf und Entwicklung moderner Konzepte, mit denen überholte Vorgehensweisen und Methoden abgelöst werden können. Typische Beispiele sind die Ablösung einer zuvor sprechenden Artikelnummer, die Einführung einer Klassifizierung zur strukturierten Produktbeschreibung, die Überarbeitung der bisherigen Warengruppensystematik oder die Nutzung von Statuskennzeichen, mit denen der Lifecycle von Produkt- oder Geschäftspartnerdaten dokumentiert werden kann. Wichtig ist es hier, die Umstellungsaufwände und -risiken ebenso wie mögliche interne Widerstände gegen die Veränderung realistisch einzuschätzen. Die Ergebnisse des Assessments können hier nützlich sein, um die eigene Datenlage zu visualisieren und angemessen zu berücksichtigen. Darüber hinaus empfiehlt sich eine Abstimmung mit der IT und dem Anbieter bzw. Implementierungspartner des Zielsystems, um dessen technische Möglichkeiten und Grenzen frühzeitig zu prüfen.

Migrationsablauf

In nahezu allen Fällen erfolgt die Datenmigration entlang einer sogenannten ETL-Strecke. Dabei stehen die Buchstaben ETL für die Begriffe „Extract“, „Transform“ und „Load“: mit einem Datenexport werden die Daten aus den Quellsystemen extrahiert und bereitgestellt, im Transformationsschritt aufbereitet und umgewandelt und schließlich per „Load“ in das Zielsystem eingespielt.

Um einen stabilen und klar definierten Ausgangszustand für die Datenmigration zu haben, wird in der Regel kurz vor dem Produktivstart des Zielsystems das Quellsystem „eingefroren“ („*system freeze*“). Sämtliche betrieblichen Prozesse werden angehalten, um alle Aktivi-

täten im Quellsystem zu unterbinden. Es liegt natürlich im Interesse des Anwender-Unternehmens, diese unvermeidliche Betriebsunterbrechung so kurz wie möglich zu halten.

Vor diesem Hintergrund weisen die drei genannten Schritte besondere Herausforderungen auf:

- **Extract:** Sämtliche Daten, d. h. eine Vielzahl von Tabellen müssen – möglichst in einem standardisierten Verfahren – in einem vordefinierten Format bereitgestellt werden. Aus den o. g. Gründen darf der Export nicht lange dauern. Typische Hürden sind in der Praxis fehlende Zugriffsrechte auf Datenbanken, Abhängigkeiten vom Hersteller oder Administrator des Quellsystems, lange Ladezeiten oder ungeeignete Formen der Datenübertragung.
- **Transform:** Die Transformation der Daten ist der Kern des gesamten Migrationsprozesses. Sie betrifft bei vielen Datenobjekten eine sehr große Anzahl von Datensätzen und ist daher im Allgemeinen nicht manuell zu bewältigen. Besondere Herausforderungen ergeben sich aus den erforderlichen strukturellen und konzeptionellen Änderungen. Kritisch sind auch die fehlende Verfügbarkeit und Kenntnis geeigneter Werkzeuge seitens der beteiligten User.
- **Load:** Die Übertragung in das Zielsystem erfolgt in der Regel über standardisierte Importprogramme oder Skripte, meist durch den Systemanbieter oder Implementierungspartner. Fallweise werden jedoch individuelle Routinen benötigt, die zuvor unbedingt ausreichend getestet werden sollten.

Entlang des skizzierten Migrationsprozesses tritt in der Praxis oft eine Zuständigkeits- und Kompetenzlücke zutage. Für den Load der Daten, also den Import in das Zielsystem, greift der Anbieter bzw. Implementierungspartner in der Regel auf standardisierte Templates und bewährte Prozeduren zurück. Er gibt das Format der bereitzustellenden Daten vor, das Anwenderunternehmen hingegen verantwortet die Bereitstellung der Inhalte. Die notwendige Aufbereitung und Transformation werden damit zur Aufgabe der User, da nur sie die Inhalte fachlich beurteilen und bearbeiten können – oft aber, ohne sie ausreichend auf diese Aufgabe vorzubereiten oder methodisch zu unterstützen („*migration gap*“, siehe Bild 1).

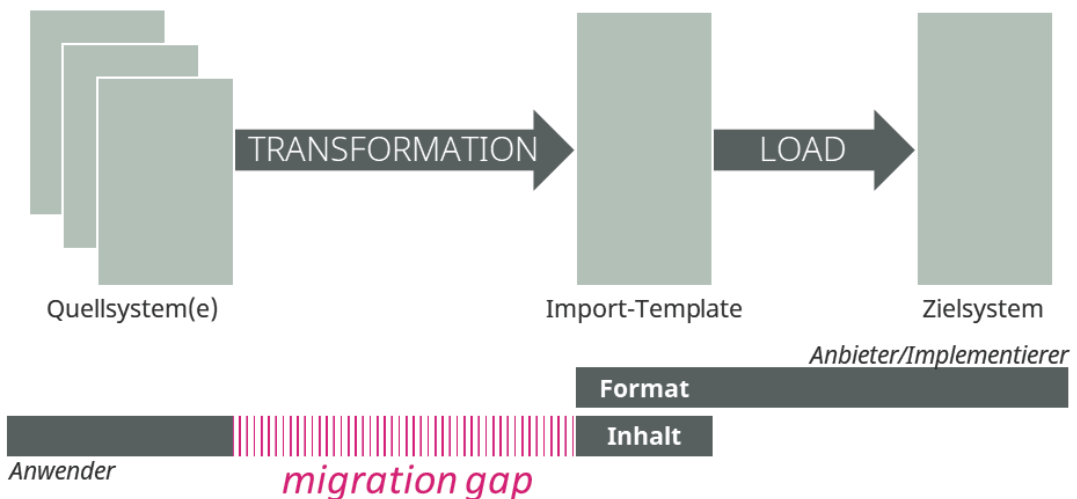


Bild 1: Verantwortungsbereiche im Migrationsprojekt. Die zur Vorbereitung der Datenimporte erforderlichen Schritte sind auf der Anwenderseite häufig weder geläufig noch methodisch unterstützt.

Die Durchführung einer Datenmigration ist mit einem Umzug vergleichbar: Für sämtliche Gegenstände im Haushalt muss entschieden werden, ob sie mitgenommen werden, ob sie gesäubert oder repariert werden müssen und für welches Zimmer sie bestimmt sind. Dabei ist es Sache der Bewohner, die Kartons zu packen und zu beschriften. Das Umzugsunternehmen stellt womöglich leere Kartons zur Verfügung, nimmt ansonsten aber nur die fertig gepackten Kartons entgegen und transportiert sie an den Bestimmungsort.

Vorgehen für eine qualitätsgesicherte Migration

Neben den drei technischen Phasen des ETL-Prozesses beruht die gründliche Vorbereitung einer Migration vor allem auf vier inhaltlichen Phasen: *Scouting* (Sichtung der Datenquellen), *Selecting* (Auswahl der relevanten Datensätze), *Cleansing* (Datenaufbereitung) sowie *Mapping* (Aufbau und Umsetzung der Transformationsregeln).

Scouting

In einem ersten Schritt sollten zunächst die für die Migration in Frage kommenden Datenquellen gesichtet werden. Auch wenn feststeht, welche Systeme abgelöst werden, müssen häufig nicht nur deren Daten für die Migration herangezogen werden. Neben einem (oder mehreren) Quellsystemen existieren in der Praxis oft weitere Datenquellen: vor- oder nebengelagerte Systeme (z. B. PLM, CRM), deren Daten womöglich aktueller oder umfangreicher sind, oder die sogenannte „Schatten-IT“, die auf lokalen Rechnern und Laufwerken eine undurchsichtige Parallelwelt von Daten bereithält. Gerade bei der Einführung von integrierten Anwendungen wie ERP-Systemen sollte darauf geachtet werden, möglichst viele Redundanzen zu beseitigen und vorhandene Datenquellen zusammenzuführen. Bewährt hat sich eine Befragung der User zu den eingesetzten Anwendungen (häufig auch: Dateien) sowie eine Sichtung der Datenquellen selbst. Im Idealfall lassen sich deren Metadaten mit geeigneten Analysetools im selben Schritt in Form eines Datenverzeichnisses dokumentieren.

Selecting

Im zweiten Schritt sollte anhand geeigneter Kriterien bestimmt werden, welche Daten überhaupt migrationsrelevant sind, um den Aufwand und die Komplexität für die nachfolgenden Phasen so gering wie möglich zu halten. Je nach Datenobjekt kommen unterschiedliche Kriterien in Frage. Das sicherlich wichtigste Kriterium: Wann wurde das Datenobjekt zum letzten Mal benutzt, also in einer Transaktion – beispielsweise einem Auftrag – verwendet? Objekte, die lange nicht genutzt wurden, sollten nur in einem Archiv gespeichert, nicht aber migriert werden; dabei muss der Betrachtungszeitraum unternehmensspezifisch festgelegt werden. Darüber hinaus ist es neben etwaigen Artikelbeständen für Produktionsunternehmen besonders wichtig zu prüfen, welche Komponenten in aktiven Stücklisten verwendet werden, im besten Fall durch eine vollständige Iteration durch alle Stücklisten und deren Positionen. Soweit keine Verwendung in den Stücklisten aktiver Komponenten und Produkte besteht, „erben“ auch untergeordnete Komponenten und Teile die Inaktivität. Im Rahmen dieser Berechnung entsteht als Nebenprodukt eine Obsoleszenzliste, also eine Liste von Teilen mit Bestand, die jedoch schon lange keine Bewegung mehr erfahren haben.

Ebenfalls im zweiten Schritt der Selektion gilt es, vorhandene Dubletten zu identifizieren. Es ist offensichtlich, dass von irrtümlich mehrfach angelegten Datensätzen nur ein Datensatz in das Zielsystem übertragen werden soll, um künftig eindeutige und verlässliche Auswertungen und Planungen erstellen zu können. Die Herausforderung der Dublettensuche besteht allerdings darin, dass Dubletten in der Regel nicht in allen Merkmalen übereinstimmen. Anhand geeigneter Kriterien müssen mit leistungsfähigen Algorithmen diejenigen Datensätze ermittelt werden, die eine ausreichend hohe Ähnlichkeit aufweisen. Dabei muss jeder Datensatz mit allen übrigen Datensätzen verglichen werden. Um möglichst viele Dubletten zu entdecken, aber falsch-positive Treffer zu vermeiden, sollte die Beschreibung der Objekte zuvor standardisiert werden, zum Beispiel im Rahmen einer Adressanreicherung oder -normalisierung.

In der Praxis hat es sich bewährt, die migrationsrelevanten Datensätze im Quellsystem zu kennzeichnen. Einerseits bietet die sichtbare Kennzeichnung für User die Möglichkeit, die Relevanzentscheidung zu validieren. Andererseits erleichtert sie es, die Datensätze sehr einfach zu filtern und so irrelevante Datensätze in den nachfolgenden Schritten zu ignorieren.

Cleansing

Im dritten Schritt, dem Cleansing, geht es darum, die Datendefekte in den als relevant eingestuften Daten zu beseitigen. Nur durch die Vervollständigung fehlender Daten und die Überprüfung und Korrektur inkonsistenter Daten kann verhindert werden, dass beim späteren Mapping unvollständige oder fehlerhafte Daten für den Upload in das Zielsystem erzeugt werden. Daher ist eine gründliche Analyse der vorhandenen Daten dringend zu empfehlen: einerseits mit dem Fokus auf die bestehenden Daten, andererseits im Vergleich mit den Spezifikationen des Zielsystems.

In diese Phase der Bereinigung und Aufbereitung fällt auch die Behandlung der zuvor identifizierten Dubletten. Damit tatsächlich je Dublettengruppe nur ein einziger Datensatz als „Golden Record“ migriert werden kann, sollten alle verfügbaren Informationen in diesem Datensatz vereinigt und konsolidiert werden. Zudem müssen bis zum Migrationszeitpunkt alle übrigen Dublettenkandidaten inaktiv sein.

Auch die Umstellung auf neue Konzepte wird in dieser Phase vorbereitet; so werden beispielsweise Artikelmerkmale aus der sprechenden Nummer abgeleitet und in ein Klassensystem übertragen. Falls die derart aufbereiteten Daten nicht in das Quellsystem zurückgespielt werden können, müssen sie übergangsweise bis zum Go-live in einer Staging Area vorgehalten werden.

In diesem Zusammenhang wird klar, dass die Methoden und Werkzeuge zur Prüfung und Aufbereitung der Daten sorgsam gewählt sein sollten. Mit Blick auf eine effiziente und fehlerfreie Bearbeitung ist eine Toolunterstützung unbedingt zu empfehlen, damit nicht durch manuelle Tätigkeiten neue Fehler entstehen. Wichtig ist auch: Während des Go-lives bleibt keine Zeit, noch manuelle Korrekturen an den Daten vorzunehmen!

Mapping

Im vierten Schritt, dem Mapping, entstehen schließlich die Transformationsregeln, mit denen die Quelldaten in die vom Zielsystem bzw. dem Import-Template vorgegebene Struktur um-

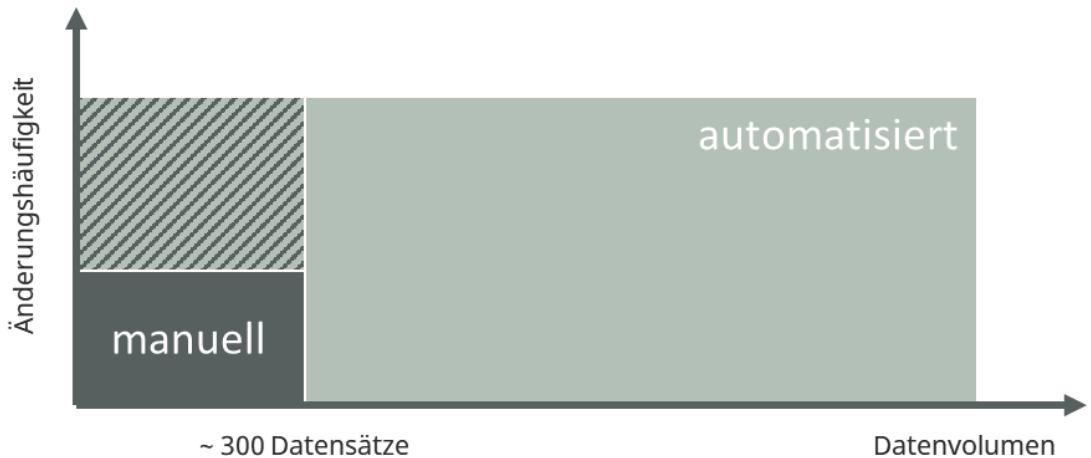


Bild 2: Während bei geringem Datenvolumen eine manuelle Übertragung per Copy/Paste denkbar ist, kommt bei höherem Datenvolumen und hoher Änderungshäufigkeit der Daten nur eine automatisierte, formel- oder skriptbasierte Implementierung des Mappings in Frage.

gewandelt werden. Dabei ist zu beachten, dass es sich in der Praxis nicht um ein einzelnes Mapping handelt, sondern um je ein Mapping pro importierter Tabelle – oft über hundert verschiedene Mappings. Jedes Mapping beschreibt dabei feldweise die Zuordnung von Quell- und Zielfeldern sowie die Art der Transformation: entweder als unveränderte Kopie oder als Konvertierung nach mehr oder weniger komplexen Regeln bzw. Zuordnungen auf Werteebene.

Zur Transformation der Quelldaten in die Zielstruktur ist eine manuelle Übertragung nur bei geringen Datenvolumen praktikabel. Auch Daten, die einer hohen Änderungsrate unterworfen sind, sollten eher regelbasiert umgewandelt werden, um Fehler zu vermeiden. Ein Datenvolumen bis zu etwa 300 Datensätzen kann erfahrungsgemäß als Obergrenze für eine manuelle Übertragung gelten (siehe Bild 2).

Praxistipp

Um von Anfang an ein transparentes Bild vom Umfang und der Komplexität der Datenmigration zu haben, empfiehlt es sich, bereits sehr früh die eigenen Daten auf den Prüfstand zu stellen. Im Rahmen eines Assessments können wichtige Erkenntnisse zu Risiken und Chancen der Datenmigration herausgearbeitet und notwendige Kompetenzen vermittelt werden. So werden bestehende Datendefekte rechtzeitig erkannt und beseitigt, neue Konzepte praxiskonform gestaltet und die Migration qualitätsgesicherter Daten gewährleistet.

Autor:

Matthias Knapp ist ein Pionier des Stammdatenmanagements. Mit dem Fokus auf Datenqualität begleitet er seit mehr als 15 Jahren zahlreiche Unternehmen weltweit bei Migrationsprojekten. Dabei verfügt das Team von KDQ über Erfahrung mit den meisten gängigen Anwendungen im Bereich ERP, PLM und CRM.

Kontakt:

KDQ Matthias Knapp
Auf der Eil 9
52078 Aachen
E-Mail: info@kdq.de
www.kdq.de