



Daten hinter Daten schaffen Intelligenz

Für erfolgreiche BI-Projekte ist vor allem die Nachvollziehbarkeit und Verlässlichkeit der Daten wichtig. Von essentieller Bedeutung sind dabei Meta- und Stammdaten, die als Daten hinter den Daten die Informationen definieren und strukturieren. Neue Lösungsansätze binden diese in ein umfassendes Integrationskonzept ein.

(erschienen in BI Spektrum 03/06, S. 23-25; Autor: Thomas Schumacher)

In der heutigen Zeit haben es IT-Manager, die Business Intelligence-Projekte betreuen, alles andere als leicht. Wachsende Datenvolumina sowie heterogene und verteilte Systeme machen den Experten ebenso schwer zu schaffen wie eine dynamische Unternehmenslandschaft, in der Restrukturierungen, Übernahmen und Zusammenschlüssen an der Tagesordnung sind. Vor diesem Hintergrund ist die optimale Nutzung der innerhalb einer Organisation vorhandenen Daten eine äußerst komplexe Aufgabe, sowohl aus technischer als auch aus organisatorischer Sicht. Für eine optimale Nutzung des „Rohstoffs“ Informationen setzen Unternehmen zunehmend auf Datenintegrationsplattformen, die den Austausch von Informationen zwischen den unterschiedlichsten Systemen automatisieren. Um Daten, die auf verschiedenen Systemen liegen, effektiv nutzen und plattformübergreifend austauschen zu können, etwa für Transaktionen oder Reports, müssen aber vor allem die Stammdaten (englisch Master Data) der betroffenen Systeme stets auf dem neuesten Stand zur Verfügung stehen.

Häufig werden die Begriffe Metadaten und Masterda-

ten/Stammdaten synonym gebraucht, dies ist jedoch nicht korrekt – eher sind es zwei Seiten einer Medaille. Bei Metadaten handelt es sich um technische beschreibende Informationen über die Daten. Die Zahl „10“ etwa hat für sich genommen keine Bedeutung, solange sie nicht mit weiteren Informationsteilen oder Metadaten (Datentyp, Feldbezeichnungen, etc.) verknüpft wird, wie z.B. „10 Euro ist der Preis einer bestimmten Aktie zu einem bestimmten Zeitpunkt“. Der Analyst Dr. Wolfgang Martin bezeichnet Metadaten als „Business-Vokabular des Unternehmens“. Masterdaten oder Stammdaten (die beiden Begriffe werden synonym verwendet) sind hingegen Informationen, die für unterschiedliche Blickwinkel auf das Geschäft dienen. Sie sind der Dreh- und Angelpunkt für jede Analyse und dienen unternehmensweit als Referenzwerk. Unternehmen, die mehrere Installationen einer ERP-Anwendung betreiben, haben beispielsweise ein Problem mit der Synchronisierung der Masterdaten, weil diese üblicherweise in jeder Instanz separat verwaltet werden müssen. Die Metadaten hingegen können über die In-

stanzen der Anwendung hinweg durchaus gleich sein.

Stammdaten dynamisch behandeln

Auch wenn die Bezeichnung Master Data Management – kurz MDM – derzeit stark in der Diskussion ist, handelt es sich im Grunde genommen um ein altes Problem, das aufgrund der technischen und wirtschaftlichen Entwicklung sowie von neuen Themen, vor allem Compliance ist hier zu nennen, in den Vordergrund gerückt ist. Den Unternehmen ist bewusst geworden, dass die schöne, neue Welt der überall verfügbaren und gemeinsam verwendeten Informationen auch Schwierigkeiten mit sich bringt und Risiken birgt. Stammdaten werden als „shared data“ zwischen verschiedenen Systemen geteilt genutzt, um Daten aus Transaktionen zu klassifizieren und zu definieren. Ein Beispiel: Heinz Müller (Vertriebsmitarbeiter) arbeitet in Bayern (Gebiet) und hat am 01. Oktober (Datum) 100 (Anzahl) Maschinenbauteile (Produkt) an die Firma Eisen Karl (Kunde) in Leipzig (Ort) für 150.000 Euro (Preis) verkauft. Zusammengefasst handelt es sich dabei um eine recht übersichtliche Transaktion, in der jedoch zahl-

reiche Elemente von Stammdaten enthalten sind: Vertriebsmitarbeiter, Gebiet, Datum, Anzahl, Produkt, Kunde, Ort und Preis. Das Problem: Metadaten und Stammdaten sind nicht statisch, sie unterliegen Veränderungen, die sich nicht nur durch Unternehmensübernahmen, sondern bereits aus dem alltäglichen Geschäft ergeben.

Angenommen, eine Firma nimmt eine neue Produktlinie in ihr Angebot auf, ein Artikel läuft aus oder ein neuer Kunde wird angelegt: die Änderung in den Stammdaten sollte bewirken, dass die neuen Informationen unverzüglich allen Systemen zur Verfügung stehen, um die Integrität der Daten zu gewährleisten. Solche Änderungen sind nicht die Ausnahme, sondern Alltag in großen Unternehmen. Banken haben nicht selten mehrere Tausend Änderungen an Stammdaten pro Monat zu verzeichnen. Erschwerend kommt hinzu, dass nach Angaben der Marktforscher der Tower Group etwa die Hälfte aller Unternehmen Stammdaten in elf oder mehr Quellsystemen pflegen. Im Gegensatz dazu handelt es sich bei Transaktionsdaten um die Aufzeichnung eines Geschäftsfalles, der nach Abschluss zu den Akten genommen werden kann. Stammdaten werden sowohl für Transaktionssysteme als auch für Reporting und Analysen benötigt, um Informationen über das Business bereitzustellen. Für das Berichtswesen oder mehrdimensionale Auswertungsmodelle beziehen sich die Stammdaten auf die Dimensionen oder Hierarchien, um die herum Kennzahlen organisiert oder ihnen zugeordnet sind – weniger auf die Fakten und Messungen selbst. Umsätze, Kosten oder Gewinne sind dabei Fakten,

während Zeitperioden, Orte, Kunden oder Lieferanten Dimensionen sind. Stammdaten sind so die Voraussetzung für einen konsistenten Blick auf das Unternehmen durch BI-, BPM- oder Reportingsysteme.

Steigerung der Datenqualität

Unternehmen, die annehmen, ein Data Warehouse würde die Stammdaten-Problematik lösen, da hier alle Daten zentral verwahrt werden, befinden sich auf einem Irrweg: in einem Data Warehouse werden Daten von mehreren Quellsystemen in einem Ziel zusammengefasst, um einen historischen Blick auf die gesamten Datenbestände zu ermöglichen. Dabei ist jedoch die Erstellung von Berichten ein Problem, wenn Änderungen an den Stammdaten nicht als Versionen behandelt werden. Wenn etwa Produktgruppen neu definiert werden, müssen auch bestehende Datensätze an die neuen Kategorien angepasst werden, da andernfalls spätere Auswertungen zu falschen Ergebnissen führen. Hinzu kommt, dass es sich beim klassischen Data Warehouse um eine Einbahnstraßen-Integration handelt: Daten aus verschiedenen Quellen werden in einem Ziel konsolidiert.

Ein MDM-Server muss jedoch in der Lage sein, Änderungen an den Stammdaten bidirektional zu integrieren, also auch mit den beteiligten operativen Systemen zu synchronisieren. Auf diese Weise stellt MDM auch eine hohe Datenqualität sicher, wenngleich das Konzept über einfache Data Quality-Funktionen weit hinausgeht. Die zentrale Verwaltung bringt Stammdaten aus mehreren Quellen an einer Stelle zusammen. Dabei können auch Data Quality-Prozeduren aufgerufen werden, um beispielsweise Doppelein-

träge herauszufiltern und Fehler zu berichtigen. Aber die Verwaltung und Pflege der Stammdaten geht erheblich weiter: Fragen wie „sollte der Produktkatalog anders strukturiert werden?“ oder „sollten die Vertriebsgebiete neu angeordnet werden?“ greifen tief in die Steuerung des Unternehmens ein.

ERP und MDM aus einer Hand

Auf dem Feld des Master Data Managements, kurz MDM, konkurrieren derzeit Anbieter mit unterschiedlichen Ansätzen. Auf der einen Seite finden sich die großen ERP- und Datenbankanbieter wie SAP, IBM oder Oracle: So hat SAP mit dem NetWeaver Master Data Management (SAP NetWeaver MDM) eine Lösung im Programm, mit der sich heterogene Systeme einbinden lassen, die sich an verschiedenen Standorten befinden können. Auf diese Weise wird eine hohe Datenkonsistenz erreicht. Die Lösung konsolidiert die Stammdaten aus diversen Systemen und überträgt sie an einen Datenspeicher, wo sie zu Analyse-zwecken abgerufen werden können. Funktionen zur Datenbereinigung und Ergänzung der bestehenden Inhalte komplettieren SAP Netweaver MDM. Globale Attribute können zentral verwaltet werden, um sicherzustellen, dass alle Systeme mit denselben Stammdaten arbeiten. Ein aktives Statusmanagement macht die Synchronisation und Verteilung der Stammdaten in jedem Schritt steuerbar, transparent und nachvollziehbar. Bereits bei der Anlage der Daten überprüft die Lösung existierende Stammdaten und schafft so die Basis für die unternehmensweite Einhaltung von Qualitätsstandards. Die an einer zentralen Stelle

angelegten Stammdaten stehen bei Bedarf allen Client-Systemen zur Verfügung. Produktinformationen können mit weiteren Inhalten angereichert werden und dann für die Veröffentlichung in Webseiten oder Druckerzeugnissen genutzt werden.

Viele Unternehmen haben aber Bedenken, ihre Infrastruktur zu stark auf einen Anbieter zuzuschneiden und sich so in eine Abhängigkeit vom Lieferanten zu begeben. Hinzu kommt, dass die „Platzhirsche“ MDM meist vor allem auf ihre eigenen Plattformen beziehen. IT-Verantwortliche beschleicht ein ungutes Gefühl, wenn sie beispielsweise Siebel-Daten mit einer Lösung von SAP integrieren oder Daten aus SAP mit einer IBM-Lösung aufbereiten sollen. Einzelne ERP-Systeme können in aller Regel keinen einheitlichen Blick auf alle Daten im Unternehmen ermöglichen, geschweige denn alle vorhandenen ERP-, BI oder Business Performance Management (BPM)-Lösungen anbinden – vor allem, da die beiden letzteren heute meist auf Abteilungsebene als Insellösungen agieren. Als Alternative bietet sich hier die Kombination verschiedener Lösungen neutraler Hersteller an. Datenintegration und Master Data Management ergänzen sich dabei in idealer Weise, wie anhand des Beispiels der jüngst bekannt gegebenen Kooperation von Informatica und Hyperion deutlich wird. Zu der Lösungsarchitektur steuert Informatica seine Datenintegrationsplattform PowerCenter bei, während Hyperion sein Hyperion System 9 Master Data Management einbringt.

Neutrale Anbieter bieten Alternativen

PowerCenter ermöglicht das MDM, indem es die Datenintegration und den Datentransport übernimmt, sowohl in Richtung Hyperion MDM als auch abgehend zu anderen Systemen. Die Exportdateien werden gleichzeitig für die schnelle und einfache Aktualisierung von Zuordnungen (Mappings), Lookup- oder Aggregationstabellen genutzt. Im Zentrum der Datenintegration stehen dabei immer Metadaten. Man unterscheidet grundsätzlich zwischen passiven und aktiven Metadaten. Passive Metadaten benötigen Interventionen durch Menschen, um eine direkte Verbindung zwischen der Beschreibung und dem Beschriebenen herzustellen, können also bestenfalls eine Momentaufnahme von Datenbeziehungen liefern. Ratsam ist daher die ausschließliche Verwendung von aktiven Metadaten, die zu jedem Zeitpunkt so aktuell wie möglich sind. Realisiert wird das Konzept der „Active Metadata“ zumeist durch einen Repository-Server in Verbindung mit einem Metadaten-Repository. Im Zusammenspiel dieser beiden Datenspeicher wird eine durchgängige Datenintegrität in jeder Phase eines Projekts sichergestellt. In diesem Repository wird quasi das Business-Vokabular festgelegt, mit dem die Backend-Systeme miteinander kommunizieren. Dazu müssen aber zuerst die Metadaten-Modelle aller beteiligten Systeme in die einheitliche Sprache transformiert werden. Durch die Zusammenfassung mehrerer Datenbestände zu einem zentralen ändert sich auch die Rolle der Datenintegration, die nun auch vorhandenen Anwendungen, denen der gewohnte Datenbestand abhanden gekommen ist, Services zur Verfügung stellen muss.

Der MDM-Server, in unserem Beispiel von Hyperion, agiert auf der nächsten Stufe als zentraler Server (IDC nennt diesen „Policy Hub“) für das Management der Änderungen an Masterdaten für zahlreiche Quellsysteme, darunter Transaktions- und Analyseanwendungen ebenso wie für Data Warehouses und Business-Performance-Management (BPM)-Lösungen. Er ermöglicht darüber hinaus den Abgleich und die Synchronisierung von Masterdaten zwischen verschiedenen Zielsystemen, sorgt für eine unternehmensweit konsistente Datenhaltung und schützt so die Datenintegrität, senkt die Kosten und Risiken, um Sicherheits- und Compliance-Anforderungen zu erfüllen und steigert die Produktivität. Das in der Praxis häufig noch manuell durchgeführte Stammdaten-Management entfällt damit. Zudem verschiebt der MDM-Server die Verantwortung, Veränderungen in den Stammdaten vorzunehmen, von der IT-Abteilung auf den Desktop der Fach- und Geschäftsanwender – die Änderungen werden dann automatisch unternehmensweit aktualisiert. Geschäftsverantwortliche haben mit akkuraten, zeitnahen Updates, die den Status Quo der Geschäftsentwicklung widerspiegeln, mehr Kontrolle über die Informationen, Anwender können damit BPM-Stammdaten anklicken und synchronisieren. Hierarchien, wie sie in Kontenrahmen, Finanz- und Vertriebsdaten oder in Niederlassungs- und Produktstrukturen bestehen, werden schnell und automatisch zwischen den Transaktions- und Analysesystemen synchronisiert.

Zwei Fliegen mit einer Klappe

Das Ergebnis ist eine hohe Exaktheit und Konsistenz der Daten und Datenstrukturen über sämtliche IT-Systeme hinweg. Damit ist die Basis geschaffen für den unternehmensweiten Einsatz von BPM- oder BI-Lösungen. Gleichzeitig ist damit der erste Schritt für die Einrichtung einer serviceorientierten Architektur getan – Unternehmen können also mit einer integrierten MDM-Lösung nicht nur ein akutes Problem lösen, sondern in einem Zug auch die Modernisierung der Infrastruktur voranbringen und somit im übertragenen Sinne zwei Fliegen mit einer Klappe schlagen. Mit Blick auf zukünftige Entwicklungen gibt es gute Gründe, in diesem Bereich eine Lösung zu wählen, die plattform- und herstellerunabhängig alle Daten im Unternehmen integriert und bereitstellt.