

## Betriebliches Katalogdaten-Management

Prof. Dr.-Ing. Frank-Dieter Dorloff / Dipl.-Wirt.-Inform. Jörg Leukel /  
Dipl.-Inform. Volker Schmitz, Essen

**Viele Unternehmen müssen ihren Kunden Produktkataloge in standardisierten, XML-basierten Formaten bereitstellen. Dabei erweisen sich die bestehenden Informationssysteme und Arbeitsabläufe sowie die mangelnde Datenverfügbarkeit und -qualität häufig als Hindernisse. Da Katalogdaten für das E-Procurement jedoch unerlässlich sind, bedarf es eines effizienten Katalogdaten-Managements, das organisatorische, informatorische und technologische Aspekte berücksichtigt. Das Gleiche gilt für die einkaufenden Unternehmen, die die Katalogdaten in ein eigenes E-Procurement-System importieren, und für elektronische Marktplätze.**

### 1. Katalogdaten im E-Procurement

Unter elektronischer Beschaffung (E-Procurement) wird die Durchführung von Beschaffungstransaktionen zwischen Unternehmen mit Hilfe von (in der Regel internetbasierten) Informationssystemen verstanden. Beschaffungstransaktionen lassen sich allgemein in die vier Phasen Informations-, Vereinbarungs-, Abwicklungs- und After-Sale-Phase einteilen. Als ein wesentliches Medium der Informationsphase dienen Produktkataloge, also Leistungsbeschreibungen der angebotenen Güter. Analog dazu sind im E-Procurement elektronische Produktkataloge notwendig.

Mit Katalogdaten sind alle Daten gemeint, die für die Erstellung von elektronischen Produktkatalogen benötigt werden. Katalogdaten werden von den Lieferanten in unterschiedlichster Art und Weise bereitgestellt. Die Unterschiede beziehen sich auf **Syntax, Semantik und Komplexität der Daten**. Im Gegensatz zum im Weiteren nicht betrach-

**Katalogdaten auch in ERP-Systemen**

teten Business-to-Consumer-Bereich (B2C) zeichnet sich der Katalogeinsatz im Business-to-Business-Bereich (B2B) dadurch aus, dass die Katalogdaten des katalogerstellenden Unternehmens in ein Informationssystem des katalogempfangenden Unternehmens übernommen werden. Folglich sind Vereinbarungen über die zu verwendenden Formate und die zu übertragenden Inhalte notwendig (vgl. Baron/Shaw/Bailey 2000).

Eine Einteilung der Katalogdaten in **Datenbereiche** kann sich daran orientieren, wie die Katalogdokumente der eingesetzten Austauschformate aufgebaut und welche Daten enthalten sind (vgl. Dorloff/Leukel/Schmitz 2001). Da es sich bei Produkten um Objekte handelt, die in betriebswirtschaftlichen Informationssystemen (ERP-Systeme) abgebildet und verwaltet werden, kann alternativ eine Orientierung an den dort anzutreffenden Strukturierungskonzepten nützlich sein.

Beide Ansätze können derart zusammengeführt werden, dass eine Einteilung in sowohl dokumentenorientierte als auch in produktbezogene Datenbereiche vorgenommen wird:

- **Grunddaten** oder **Katalog-Metadaten** sind jene Daten, die selbst noch keine Informationen über Produkte enthalten, sondern den Katalog spezifizieren und für den Dokumentenaustausch notwendig oder hilfreich sind. Hierbei handelt es sich um Angaben zu Katalogersteller (= Lieferant) und -empfänger (= Kunde), Gültigkeitszeitraum und Festlegungen, die für den gesamten Katalog gelten, z.B. Sprache der Texte und Währung der Preise.
- **Strukturierungsdaten** dienen der Einteilung des Produktspektrums in Teilsortimente. Zu unterscheiden sind lieferantenspezifische Kataloggruppensysteme und übergreifende Klassifikationssysteme. Beide Systemtypen definieren eine mehrstufige Gruppenhierarchie, so dass insgesamt eine umgekehrte Baumstruktur entsteht. In diese Struktur können die einzelnen Produkte eingeordnet werden. Bei Klassifikationssystemen wird jedes Produkt genau einer Gruppe auf unterster Ebene, also einem Blatt der Baumstruktur, zugeordnet. Diese Restriktionen gelten nicht für Kataloggruppensysteme. Klassifikationssysteme sind ein Instrument der standardisierten Produktbeschreibung, wenn sie Merkmalsleisten enthalten, die je Produktklasse vorgeben, welche spezifischen produktbeschreibenden Merkmale im Katalog anzugeben sind (vgl. Otto/Beckmann 2001).
- **Produktdaten** im eigentlichen Sinn sind alle technisch relevanten Daten und Dokumente, die innerhalb des Produktentstehungsprozesses erzeugt und verwendet werden, um ein Produkt über seinen gesamten Lebenszyklus abbilden zu können. Mit dieser Sichtweise geht die Trennung in technische und betriebswirtschaftliche Daten über Produkte einher. Aus Sicht des Katalogdaten-Managements ist die Trennung jedoch nicht aufrechtzuerhalten, da für Kataloge beide Daten von Bedeutung sind. Daher wird der Begriff Produktdaten weiter gefasst: Produktdaten beschreiben die in Katalogen enthaltenen Produkte. Produkte werden durch verschiedenste Datenelemente beschrieben. Im Minimum sind dies ein Identifikator, die Bezeichnung und der Preis. Hinzu kommen weitere Angaben, z.B. Beschreibungstexte, Merkmalsausprägungen, Bestell- und Logistikinformationen, Zuordnung zu Kataloggruppen und Klassifikationsdaten.
- **Produktbeziehungsdaten** beschreiben semantische Beziehungen zwischen Produkten. Beispiele für Beziehungstypen sind Ersatzteil, Nachfolger, Zubehör und Komponente. Auf Datenebene bedeutet dies, dass ein Produkt ein anderes Produkt referenziert und der Beziehungstyp mit angegeben wird. Erweiterte Beziehungskonzepte ermöglichen die Abbildung von Varianten und Konfigurationen.
- **Multimediatdaten** ergänzen die bislang zeichenorientierten Daten um erklärende Dateien oder Dokumente. Typische Daten sind Produktabbildungen oder Benutzerhandbücher. Es kann sinnvoll sein, solche Daten nicht nur einzelnen Produkten zuzuordnen, sondern auch zu Kataloggruppen oder Klassifikationsgruppen, um diese zu illustrieren. Das Herausstellen der Multimediatdaten als eigener Datenbereich ist notwendig, da einerseits Multimediatdaten eine hohe Bedeutung für Kataloge besitzen und andererseits das Management von Multimediatdaten spezielle Anforderungen stellt.

**Spezieller Datentyp**

**Frage 1:** Welche Zusammenhänge bestehen zwischen den fünf Datenbereichen?

**2. Charakteristika von Katalogdaten**

Katalogdaten weisen eine Reihe von Eigenschaften auf, die zu insgesamt hohen Anforderungen an das Datenmanagement und die der unterstützenden Informationssysteme

**Unterschiede zu innerbetrieblichen Daten**

führen. Diese Charakteristika beziehen sich vor allem auf die inner- und zwischenbetriebliche Verwendung der Katalogdaten.

Im Gegensatz zu den meisten anderen, in betrieblichen Informationssystemen verwalteten Daten werden Katalogdaten ihrer Bestimmung nach zwischen Unternehmen ausgetauscht. Anders als Daten z.B. der Buchhaltung, des Controllings oder des Vertriebs verbleiben Katalogdaten damit nicht innerhalb der Unternehmensgrenzen, sondern werden den Kunden bereitgestellt und von diesen genutzt. Die Nutzung erfolgt sowohl für operative Aufgaben, also für elektronische Bestellprozesse, als auch für strategische Aufgaben wie den katalogbasierten Lieferanten- und Produktvergleich. Hinzu kommt, dass der Katalogdatenaustausch nicht auf die Beziehung Lieferant-Kunde beschränkt ist. In vielen Branchen werden Katalogdaten entlang der gesamten Versorgungskette ausgetauscht, z.B. Hersteller – Großhandel – Industrie.

Für den Lieferanten besitzen Katalogdaten eine wesentliche Bedeutung. Sie beschreiben sein Sortiment und sind ein Instrument zur Differenzierung gegenüber den Wettbewerbern. Insofern können qualitativ hochwertige Katalogdaten als ein wertvolles **wirtschaftliches Gut** angesehen werden, welches gebündeltes Know-how über Produkte enthält. Dies zeigt sich am Großhandel, dessen Aufgabe unter anderem ist, seinen Kunden ein vollständiges Sortiment bereitzustellen, indem die Sortimente und damit die Kataloge der Lieferanten gebündelt werden. Der Handel aggregiert die erhaltenen Katalogdaten nicht nur, sondern „veredelt“ diese, indem fehlende Inhalte ergänzt und Daten vereinheitlicht oder normalisiert werden. Zugleich stellen Katalogdaten ein rechtlich relevantes Angebot dar. Unzulängliche oder fehlerhafte Katalogdaten können zu wirtschaftlichen Nachteilen führen.

**Dynamische, kundenspezifische Daten**

Katalogdaten haben überwiegend Stammdatencharakter (z.B. technische Produkteigenschaften), jedoch können sich bestimmte Daten (z.B. Preise) im Zeitablauf ändern. Wenn von „den“ Katalogdaten gesprochen wird, dann muss berücksichtigt werden, dass es sich nicht um die Daten nur eines Kataloges handelt. Vielmehr bilden Katalogdaten in ihrer Gesamtheit die Datenmenge, aus der Kataloge erstellt werden können. Die Bildung von kundenspezifischen Katalogen verweist auf ein wichtiges Charakteristikum: Kataloge besitzen eine Gültigkeit, die durch eine Reihe von Parametern bestimmt werden kann. Neben dem Kunden sind dies der Gültigkeitszeitraum, die Währung der Preise und die Sprache der sprachabhängigen Daten. Gemäß diesem **Multi-Abhängigkeitskonzept** kann jeder erstellte Katalog durch ein Tupel der vier Parameter, z.B. (Katalogkauf AG; 01.01.2002; EUR; DE), beschrieben werden.

Dem Typ nach umfassen Katalogdaten **strukturierte, semi-strukturierte und unstrukturierte Daten**. Die meisten Daten können bereits durch Datenelemente mit Standard-Datentypen dargestellt werden, z.B. lassen sich Produktpreise durch ein Preismodell wiedergeben, welches Basispreise, Rabatte, Mengenstaffeln und Steuern enthält. Multi-mediatdaten gehören zu den semi-strukturierten Daten, da zwar sowohl die Metainformationen als auch die Dateiobjekte selbst verwaltet werden, die Semantik des Objektes jedoch nicht abbildbar ist.

Zum Beispiel kann die Aussage einer Produktabbildung, dass das gezeigte Produkt aus zwei Komponenten besteht, nicht aus dem Dateiobjekt selbst abgeleitet werden. Produktbeschreibungstexte sind ein Fall von unstrukturierten Daten. Sie liegen als Zeichenketten vor, die Semantik wird jedoch nicht weiter modelliert. Oft enthalten Beschreibungstexte wichtige Aussagen zu den Produkten, die somit nicht direkt über spezielle Datenelemente zugreifbar sind.

**Verteilung**

Dem Inhalt nach handelt es sich bei Katalogdaten sowohl um technische als auch betriebswirtschaftliche Daten. Dies hat zur Folge, dass die Daten verteilt in unterschiedlichen betrieblichen Informationssystemen enthalten sind. Ebenso werden die relevanten Daten durch verschiedene organisatorische Einheiten im Unternehmen verwaltet. Die technischen Produktdaten sind primär den Bereichen Konstruktion und Fertigung zuzuordnen, während die betriebswirtschaftlichen Produktdaten überwiegend im Verantwortungsbereich des Vertriebs oder anderer kundenseitiger Einheiten liegen.

Katalogdaten werden nicht nur für den elektronischen Datenaustausch verwendet, sondern für weitere **Einsatzzwecke**, die mit den verschiedenen Katalogtypen zusammenhängen. Zu diesen Typen zählen traditionelle Kataloge (papiergebunden), CD-ROM-Kataloge und in Eigenregie betriebene Online-Kataloge (Internet, Sell-Side-Systeme). Damit einher gehen vielfach weitere organisatorische und informationssystem-bezogene Festlegungen im Unternehmen.

**Frage 2:** Welche Unterschiede bestehen der Art nach zwischen Controllingdaten und Katalogdaten?

**Bestandteil des Informationsmanagements**

**3. Ziele des Katalogdaten-Managements**

Im Konzept des betrieblichen Informationsmanagements ist das Datenmanagement die zentrale Komponente. Das Datenmanagement stellt den Informationssystemen die benötigten Daten zur Verfügung. Ausgehend von diesem Verhältnis von Informations- und Datenmanagement kann das Katalogdaten-Management näher beschrieben werden. Unter Bezugnahme auf die allgemeinen Ziele und Aufgaben des Datenmanagements können die spezifischen Ziele und Aufgaben des Katalogdaten-Managements definiert werden.

Das Hauptziel des Katalogdaten-Managements liegt in der Befriedigung des betrieblichen Bedarfs an jenen Informationen, die in Zusammenhang mit elektronischen Produktkatalogen relevant sind. Charakteristisch für diese Zielformulierung ist die Ausrichtung der Daten am jeweiligen betrieblichen Informationsbedarf. Die Daten müssen derart vorliegen, dass aus ihnen die benötigten Informationen abgeleitet werden können. Bezogen auf die Katalogdaten lassen sich zunächst sechs Teilziele ableiten:

- **Richtigkeit:** Die Daten müssen die von ihnen abgebildeten Sachverhalte inhaltlich richtig wiedergeben. Da Katalogdaten von Kunden genutzt werden und Grundlage für Beschaffungsentscheidungen und nachfolgende Geschäftstransaktionen sind, ist die Richtigkeit von besonderer Bedeutung.
- **Konsistenz:** In sich müssen die Daten widerspruchsfrei und fehlerfrei sein. Wird z.B. in Katalogen von einem Produkt auf ein zugehöriges Ersatzteil verwiesen, so muss das referenzierte Produkt auch im Katalog enthalten sein.
- **Vollständigkeit:** Alle Daten, die für Kataloge relevant sind oder sein können, müssen von dem Katalogdaten-Management betrachtet werden, auch wenn die Datenquelle organisatorisch zu einer anderen Einheit gehört oder in andere Zuständigkeiten fällt.
- **Aktualität:** Die Daten müssen die Sachverhalte zeitlich korrekt wiedergeben. Zum Beispiel müssen Änderungen der Produktspezifikation rechtzeitig an die Katalognutzer propagiert werden.
- **Aufgabenbezogenheit:** Die Daten müssen den Anforderungen der Aufgaben entsprechen, die vom Unternehmen durchzuführen sind. Zu unterscheiden sind Anforderungen, die aus dem Erstellen, Importieren und Verwenden von Katalogdaten entstehen.
- **Zusammenhangbezogenheit:** Die Daten sind als vernetztes System anzusehen, welches Schnittstellen und Überschneidungen mit anderen Datenbereichen aufweist. Diese Zusammenhänge müssen bei der Gestaltung des Katalogdaten-Managements berücksichtigt werden.

*siehe "Time" relatives Modell*

**4. Bezugsrahmen für das betriebliche Katalogdaten-Management**

**Szenarien versus Systeme**

Geschäftsmodelle und Anwendungen im E-Business werden häufig in Abhängigkeit von dem kataloghaltenden System oder allgemein von dem Anwendungssystem in Buy-Side, Sell-Side und Marktplatz-System eingeteilt. Dieser Einteilung folgend kann das Katalogdaten-Management ebenfalls nach dem Systemtyp differenziert betrachtet werden. Entsprechend kann zum Beispiel von KDM in Buy-Side-Systemen gesprochen und die Unterschiede zum KDM in Marktplatz-Systemen herausgearbeitet werden.

Die systemorientierte Sichtweise vernachlässigt jedoch ein wesentliches Charakteristikum des Informationsmanagements. Das Informationsmanagement soll Gestaltungskonzepte in Abhängigkeit von den betrieblichen Anforderungen entwickeln. Im Begriff betriebliches Informationsmanagement kommt dies zum Ausdruck: Analog dazu nimmt das betriebliche Katalogdaten-Management ebenfalls eine **unternehmensorientierte Sichtweise** ein. Die weitere Betrachtung basiert daher auf der jeweiligen Stellung des Unternehmens in zwischenbetrieblichen B2B-Versorgungsketten und geht von drei Anwendungsszenarien aus:

- **Szenario Supplier:** Das Katalogdaten-Management bezieht sich auf die Vertriebsaktivitäten von (in der Regel produzierenden) Unternehmen.
- **Szenario Buyer:** Das Katalogdaten-Management bezieht sich auf die Einkaufsaktivitäten von in der Regel großen Unternehmen, die über hohe Einkaufsvolumina verfügen.

- **Szenario Intermediär:** Das Katalogdaten-Management bezieht sich auf die vermittelnden Aktivitäten von Intermediären, die Supplier und Buyer zusammenbringen.

**Frage 3:** Warum ist für das Katalogdaten-Management eines Lieferanten das Anwendungsszenario Supplier sinnvoller als ein Szenario Sell-Side-System?

**Komplexitätsreduktion durch Sichten**

Mit den Anwendungsszenarien wird eine Differenzierung nach dem betrachteten Unternehmen und seiner Stellung in der Supply-Chain vorgenommen. Eng damit verbunden sind die organisatorischen Rahmenbedingungen des Katalogdaten-Managements. Im Sinne einer sichtenorientierten Gestaltung sind sie damit der Organisationsicht zuzuordnen und dort zu verfeinern. In **Anlehnung an das ARIS-Konzept** zur Beschreibung von Organisations- und Informationssystem soll daher nun ein Bezugsrahmen für das Katalogdaten-Management definiert werden, der für die weitere Betrachtung genutzt werden kann (s. Abb. 1).

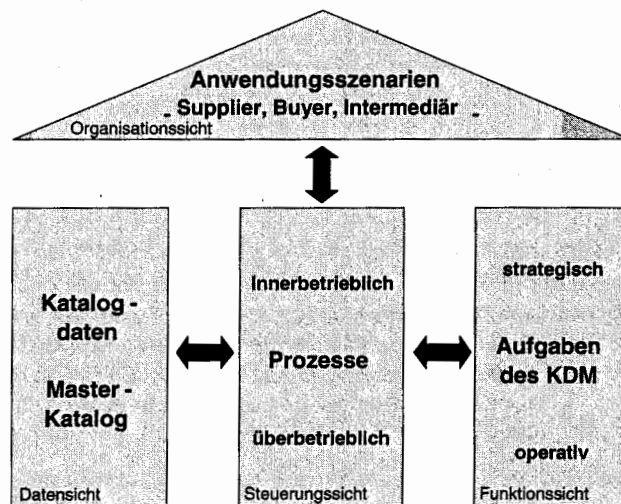


Abb. 1: Bezugsrahmen für das Katalogdaten-Management

Die **Datensicht** beschreibt Katalogdaten sowohl auf syntaktischer als auch auf semantischer Ebene. Während die syntaktische Ebene durch Datenmodelle formalisiert werden kann, stellt der Master-Katalog das inhaltliche Soll-Konzept für die Katalogdaten dar.

Die **Funktionssicht** enthält die Aufgaben, die im Katalogdaten-Management zu erfüllen sind. Dazu zählen strategische Aufgaben, die sich mit der Gestaltung befassen, und operative Aufgaben, die sich auf die Realisierung beziehen.

Die **Steuerungssicht** integriert die Sichten durch die Festlegung der durchzuführenden Prozesse, die sowohl innerbetrieblich (Datenmanagementprozesse) als auch zwischenbetrieblich (Austauschprozesse) zu sehen sind.

### 5. Strategische Aufgaben im Katalogdaten-Management

Die strategischen Aufgaben im Katalogdaten-Management befassen sich mit der Gestaltung des Katalogdaten-Managements. Die Hauptaufgabe besteht darin, eine **leistungsfähige Informationssysteminfrastruktur** aufzubauen, die das Erfüllen der oben beschriebenen Ziele sicherstellt.

In Anlehnung an den Bezugsrahmen können folgende Teilaufgaben formuliert werden:

1. Entwicklung eines Master-Katalogs
2. Entwicklung einer Katalog-Architektur
3. Entwicklung/Auswahl von KDM-Informationssystemen
4. Gestaltung der innerbetrieblichen KDM-Prozesse
5. Gestaltung der zwischenbetrieblichen Katalogdatenaustauschprozesse

**Management im Sinne von Führung**

**Welche Daten?**

Der **Master-Katalog** definiert unternehmensindividuell, welche Daten innerbetrieblich verwaltet werden. In den Szenarien Supplier und Intermediär kommt hinzu, welche Daten in welcher Form an Kunden weitergegeben werden. Die Entwicklung eines Master-Katalogs erfordert vor allem fachliches Wissen über und Erfahrungen mit Katalogdaten, Strukturierungssysteme, Katalogstandards und die derzeitigen und künftigen Anforderungen an die Daten.

Als Soll-Konzept enthält der Master-Katalog Vorgaben, z.B. für einen Lieferanten:

- „Unsere Katalogdaten decken zu 100% die Datenelemente des Katalogstandards BMEcat 1.2 ab; d.h., es werden unter anderem Produktreferenzen und -abbildungen verwaltet.“
- „Unsere Produkte werden standardmäßig nach dem Klassifikationssystem eCI@ss 4.0 klassifiziert und mit Merkmalen versehen.“
- „Unsere Produktpreise werden nur als Kundennettopreise verwaltet und weitergegeben. Die jeweils aktuellen Preise werden beim Katalogexport aus dem ERP-System übernommen.“
- „Unsere Produkte besitzen zum Teil kundenspezifische Artikelnummern und -bezeichnungen.“
- „Unsere Produkte werden zum Teil nach kundenspezifischen Klassifikations- und Kataloggruppensystemen eingeteilt.“

**Frage 4:** Welche Aussagen könnte der Master-Katalog zu semi-strukturierten und unstrukturierten Daten treffen?

Welche Informationssysteme?

Die **Katalog-Architektur** zeigt die beteiligten Informationssysteme und Datenflüsse auf, die insgesamt für die Verwaltung der unternehmenseigenen Katalogdaten notwendig sind. Eine wichtige Frage ist, wie viele Informationssysteme genutzt werden sollen und welches jeweils das datenführende System ist. Ein datenführendes System übernimmt für festgelegte Datenbereiche die Datenhaltung und -verwaltung und übergibt diese Daten gegebenenfalls an andere Systeme. Im Szenario Supplier wird häufig eine Soll-Architektur angestrebt, die aus lediglich zwei Systemen besteht:

- Das **ERP-System** ist das führende System für alle produktbezogenen Stammdaten, soweit sie abgedeckt werden können, und die kundenspezifischen Preise. Datenänderungen werden periodisch oder bei Bedarf an das Katalogsystem propagiert.
- Das **Katalog-System** ist das führende System für alle weiteren relevanten Daten, insbesondere für die Klassifizierung und Eingruppierung. Zusätzlich werden beliebige Multimediadaten verwaltet.

Die beschriebene Architektur ist gekennzeichnet durch Zentralisierung der Datenhaltung und eindeutige Abgrenzung der datenführenden Systeme. Ziel ist die redundanzfreie Datenpflege und die Ablösung von eventuell bestehenden, getrennten und nur lose gekoppelten Anwendungen für z.B. Bilder, Datenblätter, Produktmerkmale und Preise.

Welche Software-Produkte?

Häufig ist das Katalogsystem noch nicht vorhanden und muss daher ausgewählt bzw. entwickelt werden. Da es sich bei Katalogsystemen um einen relativ neuartigen Softwaretyp handelt, ist der **Markt sehr heterogen strukturiert** und wird noch nicht von wenigen Anbietern dominiert. Aufgrund der Berücksichtigung vieler unternehmensspezifischer Anforderungen sind wenige fertige „Plug & Play“-Lösungen verfügbar.

Unter **KDM-Prozessen** werden alle innerbetrieblichen Abläufe und Festlegungen für den Import, die Pflege und den Export von Katalogdaten verstanden.

Im Unterschied zum innerbetrieblichen Fokus der KDM-Prozesse geht es bei den **zwischenbetrieblichen Katalogdatenaustauschprozessen** um die Kommunikation und den Datenaustausch mit Lieferanten (Einkaufsseite) bzw. Kunden (Vertriebsseite).

#### 6. Operative Aufgaben im Katalogdaten-Management

Die operativen Aufgaben im Katalogdaten-Management umfassen die Tätigkeiten des laufenden Betriebs. Es handelt sich vor allem um Aufgaben, die direkt mit der Verwaltung der Katalogdaten zusammenhängen. Es können die drei Hauptaufgaben **Import, Bearbeitung und Export von Katalogdaten** unterschieden werden, die je nach Anwendungsszenario sehr unterschiedlich in Art und Umfang sind.

Formatunterschiede beherrschen

Sowohl beim Import als auch beim Export findet ein Datenmanagement-Konzept Verwendung, welches mit Data-Mapping bezeichnet wird. Das **Data-Mapping** übernimmt die Zuordnung der Datenelemente eines Quellformates zu den Datenelementen eines Zielformates. Damit ist es möglich, Daten eines Formates in ein anderes zu überführen. Dieses Konzept hat seinen Ursprung in der Verwaltung und Transformation von relationalen Datenmodellen und wird dort Schemaintegration genannt (vgl. Stonebraker/Hellerstein 2001). Im KDM wird das Data-Mapping genutzt für das Mapping der Importdaten auf die Datenstruktur des Katalogsystems (Import) und für das Mapping der Katalogsystemdaten auf die Exportformate (Export).

Data-Mapping definiert Anweisungen, die Daten zueinander in Beziehung bringen. Die Komplexität reicht von sehr einfachen direkten Zuordnungen bis zu fallbasierten Regeldefinitionen.

Es können zunächst drei Grundtypen identifiziert werden, die beschreiben, wie viele Datenelemente am Mapping beteiligt sind:

#### Einfachster Fall

- Das **1:1-Mapping** nimmt eine direkte Zuordnung zwischen je einem Datenelement des Quell- und Zielformates vor. Dieses Mapping wird typischerweise dann angewendet, wenn in den beiden Formaten die Datenelemente lediglich unterschiedlich bezeichnet sind, der Bedeutungsinhalt jedoch identisch ist. Beispiel: Mapping von ARTIKELNR nach SUPPLIER\_AID.
- Das **N:1-Mapping** ist dann anzuwenden, wenn im Quellformat die relevanten Daten auf mehrere Datenelemente verteilt sind und sie im Zielformat folglich zusammenzuführen sind. Im Katalogbereich kommt dies häufig vor, falls in (älteren) Informationssystemen Beschränkungen für die Länge von Zeichenketten bestehen. Beispiel: Mapping von NAME1+NAME2 nach DESCRIPTION\_SHORT.
- Das **1:N-Mapping** teilt unter Angabe der Teilungsposition die Daten eines Quelldatenelement auf mehrere Zieldatenelemente auf. Beispiel: Im Importformat enthält das Feld PREIS sowohl den Betrag als auch die Währung (z.B. „100,00DM“). Das Mapping teilt nun an der dritten Stelle von rechts das Feld in PRICE\_AMOUNT und PRICE\_CURRENCY, um Betrag und Währung getrennt voneinander zu halten.

#### Datenrepräsentation

Mit den drei Grundtypen ist es bereits möglich, eine Vielzahl von Transformationen durchzuführen. Jedoch ist eine notwendige Voraussetzung, dass die betrachteten Datenelemente keine Unterschiede in der Darstellung der Inhalte aufweisen. Der Umgang mit Darstellungsunterschieden erfordert **weitergehende Mapping-Konzepte**, die zu Datenmanipulationen führen. Die Manipulationen werden durch eine oder mehrere Regeln beschrieben und erweitern die Grundtypen wie folgt:

Im 1:1-Fall wird die **Darstellung der Inhalte** geändert. Die Änderung kann sich auf den Datentyp beziehen, wenn z.B. eine Ganzzahl in eine Fließkommazahl umgesetzt wird, oder auf die Kodierung des Inhaltes, z.B. Umsetzung von „DM“ im Quellformat gemäß ISO-Code nach „DEM“.

#### Formel als Regeltyp

Im datenändernden N:1-Fall erfolgt in der Regel eine **Reduktion oder Komprimierung der Inhalte**. Beispiel: Das Katalogsystem verwaltet nur Kundenendpreise (PRICE\_AMOUNT), im Quellformat wird jedoch der Listenpreis (PRICE) und der anzuwendende Rabatt (DISCOUNT, z.B. „10%“) dargestellt. Die Mapping-Regel lässt sich als Formel beschreiben:  $PRICE\_AMOUNT = PRICE * (1 - DISCOUNT)$ .

Das datenändernde 1:N-Mapping ist dann anzuwenden, wenn im Zielformat die **Inhalte aufwändiger und strukturierter modelliert** sind. Dieser Fall kommt in der Praxis sehr oft vor. Die Quelldaten weisen eine recht simple Struktur auf und sind in ein System zu importieren, welches Katalogdaten möglichst allgemeingültig verwaltet und diese daher durch komplexere Modelle repräsentiert. Beispiel: Im Quellformat enthält das Postleitzahlfeld auch ein Landeskennzeichen (z.B. „D-45359“), das Zielformat trennt beide. Die Mapping-Regel lautet dann: Position 1 umsetzen gemäß ISO-Code nach PLZ; Position 2 ignorieren; ab Position 3 mappen auf COUNTRY.

In der Praxis stellt sich die Mapping-Problematik aufwändig und arbeitsintensiv dar, da neben den oben beschriebenen Fällen weitere Gesichtspunkte berücksichtigt werden müssen:

Die Regelformulierung in den datenändernden Fällen ist oft schwierig oder benötigt Konstrukte, die z.B. auf Zeichenketten **nicht nur einfache String-Funktionen** anwenden. Nur so ist es möglich, in dem zuvor dargestellten PLZ-COUNTRY-Mapping korrekt „NL-72011“ umzusetzen. Hier ist das Länderkürzel zwei statt ein Zeichen lang, die oben definierte Regel würde ein falsches Ergebnis liefern.

#### Metadaten sind notwendig

Das Mapping benötigt nicht nur Wissen über die Syntax der zu importierenden Daten, sondern ebenso über deren Bedeutung (vgl. Omelayenko/Fensel 2001). Dies ist dann problematisch, wenn die Quellformate wenig oder nicht dokumentiert sind und über die **Semantik der Datenelemente** genaue Angaben fehlen. Beispiel: Das Quellformat besitzt ein Datenelement PREIS3, das den Preis inklusive ermäßigter Umsatzsteuer enthält. Ohne dieses Wissen ist ein korrektes Mapping  $PRICE\_AMOUNT = PREIS3$  und  $TAX = „.07“$  nicht möglich.

Bisher wurde implizit unterstellt, dass die zu integrierenden Datenmodelle sich nur auf Datenelementebene unterscheiden, von ihrer Beziehungsstruktur jedoch gleichartig sind, die vorhandenen Beziehungen und Kardinalitäten zwischen den Datenbereichen also identisch sind. Dies ist jedoch meistens nicht der Fall. Oft stellen nicht standardisierte Quelldatenformate wie Excel-Tabellen oder komma-separierte Dateien 1:N-Beziehungen über eine Verbundoperation als **flache Struktur** dar. Beispiel: Im Quellformat können zu jedem Produkt eine beliebige Anzahl an Preisen durch jeweils angefügte Datenelemente enthalten sein. Das Mapping muss die aufgelöste Beziehung wieder entstehen lassen und jeden Datensatz zunächst in ein Produktdatenelement (ARTICLE) umsetzen und diesem in Abhängigkeit von der Anzahl der Preise je ein neues Preisdatenelement zuordnen (ARTICLE\_PRICE\_DETAIL). Die Umsetzung wird als schleifenartige Regel beschrieben.

Umgekehrt zu flachen Quelldatenformaten können die zu importierenden Daten auch auf mehrere Formate und Dateien verteilt sein. In diesen Fällen müssen unter Beachtung der Verbindungsdatenelemente zwischen den Formaten **übergreifende Regeln** definiert werden.

**Frage 5:** Nach welchen Gesichtspunkten kann das Data-Mapping-Konzept strukturiert werden?

**Realisierung des Data-Mapping**

Data-Mapping-Funktionen sind ein Bestandteil von Katalogsystemen. Beim Import und Export von standardisierten Datenformaten (z.B. BMEcat, xCBL) ist das Mapping bereits definiert und bleibt dem Benutzer verborgen. Sollen jedoch nicht-standardisierte Daten importiert oder exportiert werden, so ist das Mapping in der Regel einmal individuell festzulegen und wird dann im Katalogsystem in **Import- bzw. Exportkonfigurationen** gespeichert. Darüber hinaus enthalten die Konfigurationen weitere Informationen über die durchzuführenden Vorgänge, insbesondere legen die Exportkonfigurationen fest, welche Produkte bzw. Sortimente und welche Datenbereiche im zu erstellenden Katalogdokument enthalten sind.

**Weiteres Datenmanagementkonzept**

Die Bearbeitung von Katalogdaten erstreckt sich prinzipiell auf alle Datenfelder. Es sind Funktionen für die Verwaltung von Produkten, Preisen, Katalog- und Klassifikationsgruppen, Merkmalen, Multimediadaten, Artikelreferenzen und vieles mehr bereitzustellen. Neben den allgemeinen Datenverwaltungsfunktionen wie Anlegen, Ändern und Löschen treten weitere, katalogspezifische Funktionen, die mit **Content-Mapping** bezeichnet werden können.

Mit Content-Mapping werden Katalogdaten auf inhaltlicher Ebene zu anderen Daten in Beziehung gesetzt. Im Unterschied zum Data-Mapping erfolgt das Mapping, um die Daten semantisch zu strukturieren. Content-Mapping zeigt sich zum Beispiel durch folgende **Katalogsystemfunktionen**:

- Zuordnung von Produkten zu Kataloggruppen
- Zuordnung von Produkten zu Klassifikationsgruppen, um Produkte zu klassifizieren
- Zuordnung von Kataloggruppen zu Klassifikationsgruppen, um alle zugehörigen Produkte zu klassifizieren
- Zuordnung von Multimediadaten zu Produkten
- Zuordnung von Multimediadaten zu Kataloggruppen und Klassifikationsgruppen
- Zuordnung von Produkten oder Gruppen zu Kunden, um kundenspezifische Kataloge zu bilden
- Zuordnung von internen Merkmalen auf die Merkmale von Klassifikationssystemen, um alle Merkmale eines Produktes nur einmal zu verwalten
- Zuordnung von Gruppen eines Klassifikationssystem zu Gruppen eines anderen, um eine teil-automatisierte Klassifizierung zu erreichen

**Tool für große Kataloge**

Content-Mapping erleichtert den Umgang mit den Daten sehr vieler Produkte, indem z.B. nicht jedes Produkt einzeln einem Kundenkatalog zuzuordnen ist, sondern dies mit Rückgriff auf Strukturierungsdaten geschieht.

**7. Zwischenbetriebliches Katalogdaten-Management**

**Katalogstandards lassen Fragen offen**

Das zwischenbetriebliche Katalogdaten-Management erfordert trotz des Einsatzes standardisierter Katalogaustauschformate nach wie vor einen hohen Koordinations- und damit verbundenen Kommunikationsaufwand zwischen den katalogerzeugenden und katalogverarbeitenden Unternehmen. Gerade in der **Initialisierungsphase**, also dem Ein-



richten des ersten Austausches zwischen den beteiligten Partnern, müssen die Vorgaben der eingesetzten XML-Formate wesentlich konkretisiert werden.

**Frage 6:** Welchen Beitrag kann das Content-Mapping für die Klassifizierung von Produkten leisten?

Sämtliche Standards lassen bei ihrer Anwendung **Wahlmöglichkeiten** offen und beinhalten auch **Interpretationsspielräume**. Diese Unklarheiten sollten vor dem Beginn des Datenaustausches geklärt werden, um einen reibungslosen Import der Katalogdaten zu ermöglichen. Abstimmungsaufwand ergibt sich einerseits aus den unterschiedlichen Geschäftsprozessregeln und den dafür benötigten Daten und orientieren sich andererseits an den technischen Möglichkeiten und Rahmenbedingungen der verarbeitenden Informationssysteme. Der konkrete Klärungsbedarf umfasst beispielsweise folgende Bereiche:

- Unterstützte Katalogtransaktionen (Katalog, Katalog-Update, Preis-Update)
- „Verschärfung“ der Standards auf Feldebene (z.B. Einforderung von Bildern oder sonstigen Marketinginformationen)
- Erweiterung der Standards auf Feldebene (kundenspezifische Informationen, wie z.B. Artikelnummern des Kunden, Kostenstellen oder Schlagworte)
- Beschränkungen des Import-Systems bezüglich Feldlängen oder nicht interpretierbarer Datenelemente
- Vorgabe von Standardwerten (Sprache, Währungen, Zeiträume, Klassifikationssysteme)
- Konkretisierung des Standards (z.B. Dateiformate der Bilddaten)

**Prozesse  
sind nicht automatisiert**

Diese Abstimmung vor und während des Austausches der Katalogdaten erfolgt zur Zeit zwar unter Verwendung von standardisierten Formaten, jedoch sind die Austauschprozesse selbst noch stark durch manuelle oder halb-automatische Prozessschritte gekennzeichnet, die fehleranfällig und zeit- und kostenaufwändig sind. Aus diesem Grund ist zu überlegen, inwieweit hier eine Vollautomatisierung erreicht werden kann. Ein Ansatz liegt darin, ein **Modell für die Kommunikation zwischen katalogverarbeitenden Software-Systemen** bereitzustellen und zugehörige XML-Nachrichtentypen zu entwickeln. Diese Nachrichten könnten den Katalogaustauschprozess initiieren, steuern und überwachen. Die Nachricht RFC (Request for Catalog) könnte zum Beispiel von einem einkaufenden Unternehmen an einen potenziellen Lieferanten übermittelt werden und bereits alle oben erläuterten Festlegungen zum Katalogaustauschprozess beinhalten. Weitere Nachrichtentypen wären „Request for Catalog Update“ oder „Catalog Import Report“, welcher eine von katalogexportierenden System verarbeitungsfähige Liste der Importstatus (z.B. Artikelnummer und Fehlercode) übermittelt.

#### Literaturempfehlungen:

- Baron, J.P./Shaw, M.J./Bailey, A.D.: Web-based E-catalog Systems in B2B Procurement. In: Communications of the ACM (CACM), Vol. 43 (2000), No. 5, S. 93 - 100.
- Dorloff, F.-D./Leukel, J./Schmitz, V.: Standards für den Austausch elektronischer Produktkataloge. In: WISU, 30. Jg. (2001), S. 1528 - 1536.
- Omelayenko, B./Fensel, D.: A Two-Layered Integration Approach for Product Information in B2B E-commerce. In: Proceedings of the 2nd International Conference on Electronic Commerce and Web Technologies EC WEB-2001, München, S. 226 - 239.
- Otto, B./Beckmann, H.: Klassifizierung und Austausch von Produktdaten auf elektronischen Marktplätzen. In: Wirtschaftsinformatik, 43. Jg. (2001), S. 351 - 362.
- Stonebraker, M./Hellerstein, J.M.: Content Integration for E-Business. In: Proceedings ACM SIGMOD Conference 2001, Santa Barbara (California), 2001.

**Die Beantwortung der Fragen erfolgt im WISU-Repetitorium.**

## Wirtschaftsinformatik/Hauptstudium

Fragen und Antworten 1 - 6 zu „Betriebliches Katalogdaten-Management“ von Prof. Dr.-Ing. F.-D. Dorloff/Dipl.-Wirt.-Inform. J. Leukel/Dipl.-Inform. V. Schmitz. WISU 3/02, S. 369 - 377.

### Frage 1: Welche Zusammenhänge bestehen zwischen den fünf Datenbereichen?

Zwischen den fünf Datenbereichen besteht eine Vielzahl von Zusammenhängen. Zunächst spezifizieren die Metadaten den Katalog und geben damit zum Beispiel Währung, Sprache und Kunde (Produktdaten) vor. Zusätzlich wird darüber informiert, ob Kataloggruppen- bzw. Klassifikationssysteme (Strukturierungsdaten) und Multimedia-daten mit übertragen werden. Die Verbindung zwischen Strukturierungsdaten und Produktdaten erfolgt dadurch, dass zu den Produkten die relevanten Gruppen und gegebenenfalls Merkmalsausprägungen angegeben werden. Die Produktbeziehungsdaten stellen schon ihrem Namen nach Beziehungen zwischen Produkten (Produktdaten) her. Schließlich finden sich Multimediadaten in drei Datenbereichen. Auf Metadatenebene kann zum Beispiel das Logo des Lieferanten enthalten sein. Zu Katalog- oder Klassifikationsgruppen und den Produkten selbst können erklärende bzw. ergänzende Abbildungen und Dokumente angefügt werden.

### Frage 2: Welche Unterschiede bestehen zwischen Controllingdaten und Katalogdaten?

Controllingdaten unterscheiden sich grundlegend von Katalogdaten. Zuerst handelt es sich bei Controllingdaten um historische Daten über bereits durchgeführte Transaktionen, während Katalogdaten zeitaktuelle Daten für zukünftige Geschäftstransaktionen sind. Auch werden Controllingdaten fast ausschließlich innerbetrieblich verwendet, demgegenüber werden Katalogdaten zwischenbetrieblich ausgetauscht. Außerdem zeichnen sich Controllingdaten dadurch aus, dass sie durch Aggregationen und Transformationen aus operativen Daten entstehen, d.h., sie verdichten und selektieren Daten. Dies ist bei Katalogdaten nicht der Fall. Schließlich ist der Zentralisierungsgrad bei Controllingdaten wesentlich höher. Durch ERP-Systeme und Data Warehousing werden alle unternehmensrelevanten Daten dem Controlling zugänglich gemacht. Katalogdaten sind dagegen häufig noch stark verteilt.

### Frage 3: Warum ist für das Katalogdaten-Management eines Lieferanten das Anwendungsszenario Supplier sinnvoller als ein Szenario Sell-Side-System?

Das Szenario Sell-Side-System für Lieferanten würde sich auf jene Datenverwaltungsaspekte beschränken, die mit dem Betrieb eines Sell-Side-Systems, also eines Shops, zusammenhängen. Damit wird jedoch vernachlässigt, dass Lieferanten auch Teil von anderen E-Business-Anwendungen sind, insbesondere stellen sie Kataloge für Buy-Side-Systeme und Marktplätze bereit.

Hinzu kommen Offline-Kataloge (Print- und CD-Kataloge). Im Sinne einer ganzheitlichen Betrachtung sollte das Katalogdaten-Management alle katalogdatenrelevanten Aktivitäten eines Unternehmens zum Gegenstand haben und dafür ein integriertes Informationssystem- und Managementkonzept entwickeln. Dies kommt im Anwendungsszenario Supplier zum Ausdruck.

### Frage 4: Welche Aussagen könnte der Master-Katalog zu semi-strukturierten und unstrukturierten Daten treffen?

Da der Master-Katalog das inhaltliche Soll-Konzept der Katalogdaten abbildet, könnten in ihm Anforderungen enthalten sein, die das Management von semi-strukturierten und unstrukturierten Daten erleichtern. Bezüglich der semi-strukturierten Daten, in der Regel Multimediadaten, ist zu fordern, dass zusätzlich Metadaten, d.h. Daten über die Daten zu verwalten sind, z.B. Bezeichnung, Version, Beschreibung, verantwortliche Stelle, Erstellungsdatum und Dateityp. Dies sind typische Informationen, die für das Dokumentenmanagement notwendig sind. Unstrukturierte Daten sind nach Möglichkeit zu vermeiden, insbesondere sollten Beschreibungstexte keine oder nur wenige Informationen enthalten, die bereits an anderer Stelle im Katalog vorhanden sind, um zu einer redundanzfreien Datenhaltung

zu gelangen. Zum Beispiel sollten Merkmalsausprägungen bei den Merkmalen abgelegt und nicht in der Produktbeschreibung werden.

### Frage 5: Nach welchen Gesichtspunkten kann das Data-Mapping-Konzept strukturiert werden?

Zunächst kann das Data-Mapping-Konzept nach der Anzahl der beteiligten Datenelemente auf Quell- und Zielseite strukturiert werden. Entsprechend lassen sich die drei Grundtypen 1:1, N:1- und 1:N-Mapping ableiten. Ein zweiter Gesichtspunkt ist die Datenänderung, d.h., ob das Mapping lediglich eine Zuordnung vornimmt, so dass die Inhalte der Elemente unverändert übernommen werden, oder ob Dateninhalte durch das Mapping geändert werden. Beim datenändernden Mapping, welches auf alle drei Grundtypen übertragen werden kann, sind weitergehende Regelformulierungen notwendig. Der dritte Strukturierungsaspekt beschäftigt sich mit der Frage, in welcher Struktur die Quelldaten vorliegen: Liegen 1:N-Beziehungen zwischen Datenelementen aufgelöst in einer flachen Struktur vor oder sind sie verteilt auf mehrere Quellformate bzw. -dateien?

### Frage 6: Welchen Beitrag kann das Content-Mapping für die Klassifizierung von Produkten leisten?

Das Content-Mapping kann zwei Beiträge für die Klassifizierung von Produkten leisten. Der erste Beitrag liegt darin, dass Gruppen aus Kataloggruppensystemen zu Klassifikationsgruppen zugeordnet werden. Die Folge ist, dass alle Produkte, die der Kataloggruppe angehören, damit auch einer Klassifikationsgruppe zugeordnet sind. Der zweite Beitrag entsteht durch das Mapping von Klassifikationsgruppen eines Systems zu Klassifikationsgruppen eines anderen Systems, z.B. von UNSPSC nach eCI@ss. Damit wäre es möglich, die Einführung eines neuen standardisierten Klassifikationssystems im Unternehmen erheblich zu erleichtern, da die Produkte nur nach einem System manuell zu klassifizieren wären. Allerdings ist es in beiden Fällen nicht zu erwarten, dass immer eine Gruppe auf Blattebene gefunden werden kann. Im schlimmsten Fall kann nur eine Gruppe auf sehr hoher Ebene ermittelt werden, so dass eine manuelle Klassifizierung durchzuführen ist. Auch sind Mappings zwischen Klassifikationssystemen heute noch nicht frei verfügbar.