

Das Common Warehouse Metamodel als Referenzmodell für Metadaten im Data Warehouse und dessen Erweiterung im SAP® Business Information Warehouse

Michael Hahne

cundus AG
Freiherr-vom-Stein-Straße 13a
55559 Bretzenheim
michael.hahne@cundus.de

Abstract: Heterogene Data Warehouse-Landschaften sind durch eine Vielzahl verschiedener Softwarekomponenten gekennzeichnet, deren Integration zu einer funktionierenden Business Intelligence-Lösung eine besondere Herausforderung darstellt. Die Metadaten der beteiligten Komponenten stellen dabei einen viel versprechenden Ansatz der effektiven und effizienten Verknüpfung dar, die aber durch die proprietären Metadatenmodelle erschwert wird. Das von der Object Management Group verabschiedete Common Warehouse Metamodel (CWM) hat sich inzwischen als industrieweiter Standard zur Modellierung von Metadaten in Data Warehouse-Systemen sowie zu deren Austausch etabliert. Die Architektur- und Plattformunabhängigkeit dieses Metamodells bedingt zwar eine fehlende Berücksichtigung der spezifischen Objekte dedizierter Werkzeuge, das CWM stellt hierfür aber flexible Möglichkeiten der Erweiterung bereit. Das SAP Business Information Warehouse ist mit eines der führenden Data Warehouse-Systeme und setzt für den Austausch von Metadaten eine Erweiterung des CWM ein. In diesem spezifischen Metamodell findet sich neben der Abbildung des spezifischen Datenmodells des BW insbesondere auch die Berücksichtigung von Berechtigungen und des SAP Rollenkonzeptes.

1 Einleitung

Die Data Warehouse-Architektur unterscheidet für die spezifischen Aufgabenbereiche unterschiedlichste Softwarekomponenten, die den jeweiligen Data Warehouse-Prozess abdecken. Die hierdurch entstehenden heterogenen Data Warehouse-Landschaften stellen bezüglich der Integration zu einer funktionierenden Business Intelligence-Lösung eine besondere Herausforderung dar. Neben den Zielen eines effizienten Aufbaus und Betriebs eines Data Warehouse sowie der Unterstützung der Benutzer in ihren analytischen Anforderungen bietet das Management von Metadaten darüber hinaus auch einen viel versprechenden Ansatz zur Integration der Komponenten heterogener Data Warehouse-Lösungen.

In Form des Common Warehouse Metamodels steht ein von der Object Management Group verabschiedeter Standard für die Modellierung und den Austausch von Metadaten im Data Warehousing zur Verfügung, der darüber hinaus flexibel erweiterbar ist und so auch die spezifischen Anforderungen proprietärer Werkzeuge und Systemkomponenten berücksichtigen kann.

Unter dem Begriff *mySAP Business Intelligence* sind Anwendungskomponenten und Dienstleistungen für die Entscheidungsunterstützung in Unternehmen zusammengefasst, deren Hauptkomponente das SAP Business Information Warehouse (BW) als Basistechnologie zur Datenhaltung ist.¹ Für den Austausch von Metadaten des Business Information Warehouse steht eine Erweiterung des Common Warehouse Metamodels zur Verfügung, das neben den spezifischen Objekten des zugrunde liegenden Datenmodells auch Aspekte der Berechtigungen und des Rollenkonzeptes berücksichtigt.

Der Abschnitt 2 befasst sich mit der begrifflichen Definition von Metadaten und Metamodellen sowie deren Relevanz für Data Warehouses. In Abschnitt 3 erfolgt die Darstellung des Common Warehouse Metamodels, die Beschreibung der Erweiterungsmöglichkeiten und auch der noch fehlenden Bestandteile des Metamodels. Die Darstellung der Architektur und des Modells des SAP Business Information Warehouse ist Gegenstand von Abschnitt 4 und die Implementierung des erweiterten Common Warehouse Metamodels im Business Information Warehouse schließt sich in Abschnitt 5 an. Der Artikel schließt mit einer Zusammenfassung in Abschnitt 6.

2 Metadaten im Data Warehouse

Bei dem Aufbau von Informationssystemen entsteht eine Vielzahl von Informationen, die diese Systeme beschreiben. Hierzu gehören beispielsweise umfangreiche Datenmodelle mit Datenobjekten und Beziehungen. Für operative wie auch analyseorientierte Informationssysteme auf Basis des Data Warehouse Konzeptes und der OLAP-Technologie ist die Speicherung und Verwaltung von Metadaten eine zentrale Komponente.

Gerade für den effizienten Betrieb analyseorientierter Informationssysteme spielen Metadaten in vielfältiger Hinsicht eine herausragende Bedeutung. Sie sind einerseits für den gesamten ETL-Prozess (*Extraktion, Transformation, Laden*) von grundlegender Bedeutung aber auch andererseits wesentlich für die erfolgreiche Nutzung.

Die Definition von Metadaten steht zunächst in Abschnitt 2.1 im Vordergrund. Auf deren spezielle Bedeutung für analyseorientierte Informationssysteme geht Abschnitt 2.3 ein. Anschließend erfolgt eine Klassifikation von Metadaten.

¹ SAP und Business Information Warehouse sind ebenso wie SAP R/3 (eingetragene) Warenzeichen der SAP Aktiengesellschaft Systeme, Anwendungen, Produkte in der Datenverarbeitung, Neurtrotstraße 16, D-69190 Walldorf. Die SAP AG ist nicht Herausgeber des vorliegenden Artikels und auch nicht presserechtlich dafür verantwortlich.

2.1 Begriff der Metadaten im Kontext der Modellierung

Eine geläufige prägnante Beschreibung von Metadaten lautet, dass es "Daten über Daten" sind. Diese einprägsame Festlegung ist jedoch nicht hinreichend genau. Der Begriff Metadaten ist nach Devlin etwas präziser gefasst und es ist darunter „Data that describes the meaning and structure of business data, as well as how it is created, accessed and used.“[De97, S. 52] zu verstehen. Einer geläufigen Definition zufolge ist unter Metadaten jede Art von Information zu verstehen, die für den Entwurf, die Konstruktion und die Benutzung eines Informationssystems benötigt werden [BG01].

Dem konstruktivistischen Ansatz der Modellierung folgend sind Modelle das Ergebnis eines Modellierungsvorganges, bei dem die reale Diskurswelt, die so genannte Miniwelt, durch Abstraktion abgebildet wird [KE96, S. 21]. In der Literatur werden viele Modellbegriffe und Klassifikationsmöglichkeiten unterschieden. Eine mögliche Unterscheidung von Modellen basiert auf deren Funktion, wonach Beschreibungs-, Erklärungs- und Entscheidungsmodelle zu differenzieren sind [Gr69, Ko61]. In dieser Arbeit werden Beschreibungsmodelle betrachtet.

Zur Einordnung des Begriffes der Metadaten dient die Darstellung verschiedener Abstraktionsebenen der Modellierung in Abbildung 1 [BG01], wobei ergänzend zwischen den Modellen als Modellsprache und den eigentlichen Ausprägungen differenziert wird. Die Darstellung folgt der englischsprachigen Literatur, demzufolge die Modellierungssprache als „model“ bezeichnet wird, die damit erstellten „Modelle“ jedoch als „schema“ [Le95].

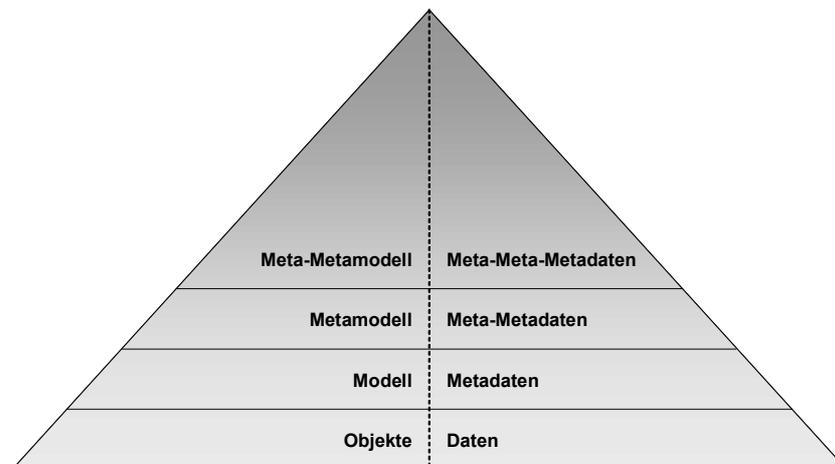


Abbildung 1: Abstraktionsebenen der Modellierung

Auf der untersten Abstraktionsebene sind die Daten und Objekte als Abbild der Miniwelt zu finden. Durch einen weiteren Abstraktionsschritt ergeben sich nun Modelle, die die Struktur der Objekte bzw. Daten beschreiben. Die Ausprägungen dieser Modelle sind die Metadaten, die sich demzufolge auf einer höheren Abstraktionsebene befinden. Von Metamodellen wird gesprochen, wenn Modelle und Modellbildung selbst zum Gegenstand der Modellierung werden. Eine weitere Stufe der Abstraktion ergibt sich beim Übergang zu Modellen, deren Fokus gerade die Beschreibung von Metamodellen ist. Dieser als Metaisierung bezeichnete Prozess kann im Prinzip weiter geführt werden [St98], jedoch reicht für die Belange in diesem Artikel die beschriebene vierstufige Metamodell-Architektur. Ziel der Modellbildung in diesem Kontext ist es, einen Gegenstandsbereich in einem Beschreibungsmodell abzubilden, zu deren Erstellung eine Sprache verwendet wird. Entsprechend dient auch eine Sprache zur Formulierung eines Metamodells, dass in diesem in diesem Sinn ein Modell eines Modells ist, „wobei es sich bei dem übergeordneten Modell um ein sprachliches Beschreibungsmittel handelt, das die Sprache, in der das untergeordnete Modell formuliert ist, abbildet.“ [St98, S. 3.]

2.3 Bedeutung von Metadaten im Data Warehouse

In der Data Warehouse Referenzarchitektur ist ein Metadatenbanksystem ein wesentlicher Bestandteil, in dem die Ablage der Informationen über die Systemkomponenten gespeichert und verwaltet werden. Neben den reinen administrativen Zwecken dient dieses insbesondere den Benutzern für ein schnelles und sicheres Auffinden benötigter Informationen im Data Warehouse [Ho99, S. 95]. Der Nutzer soll in die Lage versetzt werden, die im analyseorientierten Informationssystem gespeicherten Informationen weitgehend autonom für sich zu nutzen [Mü00, S. 133].

Für Metadaten im Data Warehouse sind im Wesentlichen die drei Nutzungsfelder der Administration des Systems, der Bedeutung gespeicherter Inhalte und der Navigationskomponente zu nennen.

Die mit der Entwicklung und Administration eines Data Warehouse befassten Mitarbeiter benötigen eher technisch orientierte Metadaten in Analogie zu den in operativen Systemen verwendeten Metadaten. Hierzu gehören Angaben über die Datenquellen, Regeln zur Verbesserung der Datenqualität, Regeln für Transformations- und Konsolidierungsschritte, Zuordnungs-Informationen zwischen den Datenquellen und den Data Warehouse-Modellen sowie die Metadaten der Datenmodelle des Data Warehouse selbst [Hu97, S. 229].

Damit die Benutzer eines Data Warehouses die enthaltenen Daten verstehen und bewerten können, benötigen sie ein breites Spektrum an Metadaten, die eine Interpretation der Data Warehouse-Inhalte ermöglichen. In diesem Zusammenhang sind insbesondere die Definitionen der verwendeten Geschäftsbegriffe sowie die Verbindung dieses Fachvokabulars mit den Datenobjekten im Data Warehouse zu nennen. Weiterhin sind auch Beschreibungen von erzeugbaren Berichten, zuständige Ansprechpartner und Voraussetzungen für Zugangsberechtigungen auf bestimmte Datenbereiche wichtige Metadaten [Ga97, S. 66].

Für einen weitgehend autonomen Zugang zu den Daten im Data Warehouse ist eine endbenutzergerechte Navigationskomponente notwendig, die ebenfalls auf Metadaten basiert. Diese ermöglicht beispielsweise Funktionen für die freie Abfrageformulierung, für Navigation, die elektronische Berichtsverteilung sowie den Durchgriff auf Daten in operativen Vorkomplexen [Mü00, S. 135].

Bei dem Aufbau einer unternehmensweiten Systemlandschaft zur Informationsversorgung für analytische Prozesse spielen zwei gegenläufige Grundprinzipien eine entscheidende Rolle: Neben dem „best-of-breed“-Ansatz, bei dem für jeden Aufgabenbereich eine separate besonders geeignete Software eingesetzt wird, versuchen „end-to-end“-Lösungen auf Basis einer einzigen BI-Suite alle Aufgaben abzudecken. Durch den Einsatz dedizierter Werkzeuge entsteht oftmals eine heterogene komplexe Systemlandschaft, die zu einer einheitlichen Data Warehouse Lösung zusammenzufügen ist. Oftmals finden sich unterschiedliche Tools für den Data Warehouse Datenspeicher, die Implementierung von ETL-Prozessen sowie OLAP-, Data Mining- und Reporting-Werkzeuge [MTH02]. Gerade für die Verbesserung der Interoperabilität in solchen heterogenen Landschaften ist der Austausch von Metadaten zwischen den einzelnen Tools ein viel versprechender Ansatz [DR00].

Das Metadatenmanagement ist eine zentrale Funktion zur Steigerung von Effektivität und Effizienz einer Data Warehouse-Lösung [Ma00]. Die hierzu notwendigen Abstimmprozesse bedürfen neben den Werkzeugen zur Metadaten-Verwaltung eines organisatorischen Rahmens für das Metadatenmanagement zur Sicherstellung von Konsistenz, Aktualität und Qualität der Metadaten [Au02].

2.4 Klassifikation von Metadaten

Die Klassifikation von Metadaten in einem Data Warehouse kann nach unterschiedlichen Kriterien erfolgen. Eine sehr grobe Differenzierung unterscheidet operationale und DSS-Metadaten [Po96, S. 170f.]. Die operationalen Metadaten beinhalten Informationen über die Quellsysteme, Originaldatenquellen, Datenstrukturen, Feldbezeichnungen sowie Informationen über den Transformationsprozess und die angesprochenen Zieldatenquellen. Die DSS-Metadaten verbinden die physischen Data Warehouse-Daten mit den Modellen der Geschäftsprozesse und den Endbenutzerwerkzeugen und umfassen Angaben über die Speicherorte und -strukturen im Data Warehouse, über die Objekte des Business Modells sowie zu den Navigationsmöglichkeiten.

Die fünf Hauptfunktionen eines Data Warehouse führen zu einer anderen Sichtweise, in der die Metadatenverwaltung (*Meta-Data*) das Fundament für die anderen vier Hauptfunktionen bildet [Ho99, S. 99f.]. Diese umfassen die Identifikation von Datenquellen (*Source*), den Datentransfer- und Transformationsprozess (*Load*), die Speicherfunktion (*Storage*) mit den Aspekten der Speicherform, der Datenverteilung und Archivierung sowie die Abfrage- und Auswertefunktion (*Query*), in deren Vordergrund die Betrachtung von OLAP-Werkzeugen, Simulationsmethoden und Data Mining-Techniken steht.

Diese Klassifikation ist ähnlich der nach der Verwendung im Data Warehouse Prozess mit den Phasen *Design*, *Populate*, *Administer* und *Analyze*. Zusammen mit der Differenzierung von *Technical User* und *Business User* einerseits sowie der Unterscheidung in der Datensicht (*Operational*, *Data Warehouse* und *Data Mart*) ergibt sich eine dreidimensionale Klassifikation bezüglich Benutzer, Prozesse und Daten [DR00].

3 Das Common Warehouse Metamodel

Das Common Warehouse Metamodel (CWM) definiert ein Modell für Metadaten im Data Warehousing, das seit 1998 unter aktiver Beteiligung namhafter Softwarehersteller wie u. a. IBM, NCR, Hyperion Solutions und Oracle entwickelt wurde und von der Object Management Group (OMG) erstmals im Jahr 2000 als Standard verabschiedet wurde [OM01a]. Es ist damit ein Referenzmodell für Data Warehouse-Metadaten, das zunehmende Verbreitung findet.

3.1 Grundkonzept

Das Referenzmodell ist in die Model Driven Architecture (MDA) der OMG eingebettet [OM01c] und basiert auf über 200 Klassen und über 150 Assoziationen in einem UML-Klassenmodell. Der Schwerpunkt liegt dabei auf technischen Data Warehouse-Metadaten und dem Austausch von Metadaten zwischen verschiedenen Softwarekomponenten einer Data Warehouse Architektur. Für den Austausch von Metadaten auf Basis des CWM stehen aufgrund der Konformität zu anderen OMG-Standards die beiden Möglichkeiten über XMI (XML Metadata Interchange) für den datei- bzw. datenstrombasierten Austausch und des funktionsbasierten Zugriffs über CORBA-Schnittstellen zur Verfügung [MTH02, S. 148; Me03, S. 95ff.].

Wesentliche Leitlinien bei der Entwicklung des CWM waren die Vollständigkeit, Verständlichkeit, Unabhängigkeit, Kompatibilität, Interoperabilität, Erweiterbarkeit und Verfügbarkeit [Me03, S. 91]. Um die Plattformunabhängigkeit und Kompatibilität zu bestehenden Standards zu gewährleisten, bot sich die Verwendung der standardisierten OMG-Basistechnologien, die in der OMG Metamodell-Architektur (siehe Abbildung 2) zusammengefasst sind, hierbei an [Me03, S. 94f.].

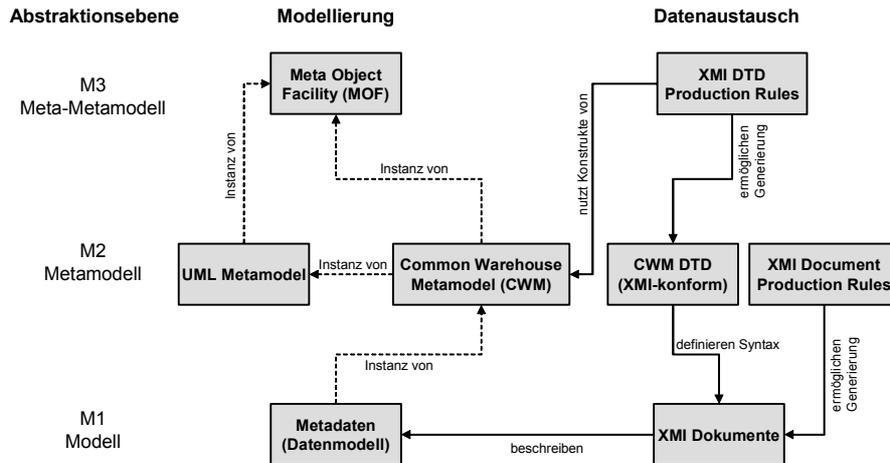


Abbildung 2: Metamodell-Architektur der OMG

Die Einbettung des Common Warehouse Metamodells in die Metamodell-Architektur der Object Management Group und die Integration mit den bestehenden Standards bietet damit neben der Sprache für die Modellierung von Metadaten auch Mechanismen zum Austausch von Metadaten und Metadatenmodellen.

3.2 Komponenten des CWM

Die Übersichtlichkeit der zahlreichen Klassen innerhalb des CWM wurde durch die Einführung thematisch fokussierter Pakete erhöht. Zusätzlich sind die Pakete in fünf Schichten angeordnet, die jeweils die Modellbestandteile einer Entwicklungsstufe des Data Warehousing umfassen [Me03]. In der untersten Schicht *Object Model* erfolgt die Festlegung der grundlegenden Konstrukte, auf denen alle weiteren Schichten aufbauen. Die Metamodelle zur Spezifikation von Datentypen, Ausdrücken und Architekturmodellen sind in der zweiten Schicht *Foundation* zusammengefasst. Die dritte Schicht *Resource* enthält Konstrukte zur Definition der Quell- und Zielstrukturen des Data Warehouse. In der *Analysis*-Schicht befinden sich die Pakete für Transformations- und (mehrdimensionale) Auswerteprozesse. Die Management-Schicht definiert Metamodelle für die Steuerung und Überwachung. Die Pakete in den einzelnen Schichten sind in Abbildung 3 zusammengefasst [Po02].

Management	Warehouse Process			Warehouse Operation		
Analysis	Transformation		OLAP	Data Mining	Information Visualization	Business Nomenclature
Resource	Object Model	Relational	Record	Multidimensional		XML
Foundation	Business Information	Data Types	Expression	Keys and Indexes	Type Mapping	Software Deployment
Object Model	Core		Behavioral	Relationships		Instance

Abbildung 3: Pakete zur Strukturierung der Metamodelle im CWM

Innerhalb der *Object Model*-Schicht erfolgt die Strukturierung in die Pakete *Core*, *Behavioral*, *Relationships* und *Instance*. Das *Core*-Metamodell ist von keinem anderen Paket abhängig und umfasst die grundlegenden Metamodell Klassen und Assoziationen. Die verhaltensbezogenen Aspekte von Modellelementen wie Operationen, Ereignisse und Schnittstellen sind im Metamodell *Behavioral* zusammengefasst. Im Paket *Relationships* sind die verschiedensten Arten von Beziehungstypen festgelegt und das Paket *Instance* bildet die Basis für die Erstellung von Dateninstanzen [Me03].

In der *Foundation*-Schicht sind die Metamodelle *Business Information* für die Ergänzung der Modelle um beschreibende Informationen und Ansprechpartner, *Data Types* für elementare und komplexe Datentypen, *Type Mapping* für die Zuordnung einfacher Datentypen zueinander, *Keys and Indexes* zur Abbildung von Indizes und Schlüsseln sowie *Expression* und *Software Deployment* zusammengefasst [OM01a].

In der *Resource*-Schicht befinden sich die Metamodelle zur Beschreibung von Datenstrukturen wie etwa für relationale Datenbanken im Paket *Relational* und multidimensionale Datenstrukturen im Paket *Multidimensional*, das von spezifischen Systemen abstrahiert und allgemeine Elemente wie Dimensionen und Dimensionsobjekte spezifiziert. Das Paket *Record* dient der Berücksichtigung von strukturierten Datentypen und XML-Dateien sind über das *XML*-Paket berücksichtigt [OM01a].

In der mit den Metamodellen zur Analyse befassten Schicht *Analysis* befinden sich die Pakete *Transformation*, *OLAP*, *Data Mining*, *Information Visualization* und *Business Nomenclature*. Das *Transformation*-Paket enthält die Metamodellkomponenten für Zuordnungen (Mappings) und Transformationen, wobei diese Begriffe in der Terminologie des Common Warehouse Metamodels streng differenziert werden [Me03, S. 101]. Die Metamodellelemente zur Beschreibung von multidimensionalen Analysestrukturen sind in dem *OLAP*-Paket zusammengefasst. Diese Strukturen sind unabhängig von der physischen Speicherstruktur, an die es über die Mapping-Konstrukte des *Transformation*-Paketes angebunden ist, und es können neue Dimensionen und Hierarchien zur Analyse festgelegt werden. Die Pakete *Data Mining* und *Information Visualization* befassen sich mit Metamodellbestandteilen für Data Mining und Visualisierungsformen. Die Berücksichtigung von Fachvokabular und deskriptiven Informationen schlägt sich in dem Paket *Business Nomenclature* nieder.

Die *Management*-Schicht beinhaltet Metamodelle, die den technischen Betrieb des Data Warehouse unterstützen. Im Paket *Warehouse Process* befinden sich Konstrukte für Ereignisse, Aufgaben, Prozesse und zum Scheduling. Die Dokumentation von Prozessausführungen und Änderungsanfragen sind Gegenstand des Paketes *Warehouse Operation* [OM01a].

3.3 Erweiterungsmöglichkeiten

Das Common Warehouse Metamodel stellt von seinem Anspruch her einen architektur- und plattformunabhängigen Rahmen für Data Warehouse Metadaten zur Verfügung. In den zahlreichen Paketen sind nahezu alle allgemein relevanten Aspekte abgedeckt, die in diesem Sinne als eine gemeinsame Basis für allgemeine Metamodelle verschiedenster Data Warehouse Systeme fungiert. Das CWM erhebt darüber hinaus aber auch den Anspruch, für die Spezifika einzelner Softwarewerkzeuge den Rahmen für Metamodelle zur Verfügung zu stellen und bietet daher Mechanismen zur flexiblen Erweiterung des Metamodells an.

Zur Erweiterung des Common Warehouse Metamodels stehen zwei grundsätzliche Erweiterungstechniken zur Verfügung. Einerseits stellt das UML Objekt Modell die Erweiterungsmöglichkeiten über Tagged Values und über Stereotypen zur Verfügung und andererseits steht mit der Festlegung zusätzlicher Metamodellpakete eine nicht normative Möglichkeit der Erweiterung zur Verfügung [OM01b].

Bei der Erweiterung um Tagged Values können Modellelemente um weitere Attributwerte der Form Attributname (tag) = Attributwert (value) ergänzt werden. Diese Erweiterung ist einerseits im Allgemeinen unabhängig von Versionswechseln der CWM-Spezifikation, ist andererseits aber auf Werte von Typ String beschränkt und kann über die Standard-Schnittstellen nicht direkt abgerufen werden. Über Stereotypen können u. a. mehrere Tagged Values inhaltlich gruppiert werden, bieten aber ebenfalls keine Möglichkeiten der Definition zusätzlicher Assoziationen [Me03, S. 102f.].

Sehr flexibel einsetzbar ist die Möglichkeit der Definition von zusätzlichen Metamodellelementen wie Klassen, Assoziationen und Paketen, um das Common Warehouse Model zu erweitern. Diese Methode sollte möglichst eng an die Konzepte der Meta Object Facility angelehnt sein und auf die Vererbungsmöglichkeiten zurückgreifen, damit sich die Erweiterung nicht zu sehr vom Standard des CWM entfernt. Mit den Extension Packages bietet die OMG selbst einige Beispiele für nicht normative Erweiterungen des CWM, die beispielsweise für spezifische Systemprodukte wie Hyperion Essbase und Oracle Express Metamodelle vorschlagen [OM01b].

3.4 Erweiterungsbedarf des CWM

Eine detaillierte Betrachtung des Common Warehouse Metamodels zeigt, dass das Metamodel ein sehr starken Akzent bei den technischen Metadaten setzt. In der Literatur wurden daher einige Vorschläge für sinnvolle Erweiterungen aufgeführt [Me03, S. 107ff.].

Data Warehouse-Systeme bieten sehr flexible Datenzugriffsmöglichkeiten und bedürfen daher einer besonderen Berücksichtigung von Aspekten der Datensicherheit und Berechtigungsverwaltung, wobei im Allgemeinen die verschiedenen Komponenten einer heterogenen Data Warehouse-Systemlandschaft eigene Konzepte zur Verwaltung von Benutzern, Rollen und Berechtigungen vorweisen. Eine Standardisierung von berechtigungsbezogenen Metadaten könnte hier den Abgleich von Berechtigungen sowie eine Erhöhung der Transparenz zur Vermeidung von Sicherheitslücken ermöglichen, sogar eine Single-Sign-On-Funktionalität könnte hierauf zurückgreifen. Das CWM bietet jedoch bisher noch keine Möglichkeiten zur Abbildung von Metadaten für Berechtigungen und Benutzer [DR00, S. 7].

Zur Sicherstellung einer dauerhaften Akzeptanz einer Data Warehouse-Lösung gehört ein Qualitätsmanagement, das den BI-Anwendungen eine konzeptionelle Qualität ermöglicht sowie funktionale und technische Qualität während Aufbau und Betrieb gewährleistet [Co97, S. 353]. Hierzu gehören auch Kennzahlen und Akzeptanzbereiche für Datenqualität und entsprechende Alerting-Mechanismen [He02, S. 138ff.]. Im Common Warehouse Metamodel finden sich allerdings bisher keine Modellbestandteile, die Metadaten dieser Form abzubilden vermögen.

Als Nutzer von Data Warehouse-Systemen sind neben den Knowledge-Workern, die mit flexiblen Werkzeugen verschiedenste Analysen durchführen, insbesondere auch die reinen Berichtsempfänger und Nutzer von Standard-Berichtslösungen zu nennen. Vielfältige Mechanismen der Berichtsgruppierung und Verteilung über unterschiedlichste Kanäle sowie Konstrukte der Berichtsdefinitionen sind im Common Warehouse Metamodel jedoch unzureichend berücksichtigt [Me03, S. 108] und bieten noch Potenzial zur verbesserten Integration unterschiedlicher Analysewerkzeuge durch entsprechende Modellbestandteile einer Erweiterung des CWM.

Die im CWM abgebildeten Modellbestandteile befassen sich zum Großteil mit den statischen Aspekten eines Data Warehouse, lediglich für die Abbildung von Transformationsprozessen stehen adäquate Metamodelle zur Verfügung. Gerade für die organisatorische Komponente eines Data Warehouse fehlen im CWM jedoch die Modellbestandteile. Für die Berücksichtigung der organisatorischen Data Warehouse-Prozesse bietet sich eine Erweiterung des CWM an [Au03, S. 132ff.].

4 SAP Business Information Warehouse

Analyseorientierte Informationssysteme zielen auf die Unterstützung der dispositiven und strategischen Prozesse in Unternehmen ab und haben eine zeitnahe Versorgung betrieblicher Entscheidungsträger mit relevanten Informationen zu Analysezwecken im Fokus. Das *Business Information Warehouse (BW)* als zentrale Data Warehouse-Lösung adressiert genau diesen Bereich und ist die Grundlage für vielfältige analytische Anwendungen.

Mit dem Business Information Warehouse ist die SAP AG in den Markt für Business Intelligence eingestiegen. Besonderes Leistungsmerkmal des BW ist der *Business Content* als eine Sammlung vorgefertigter betriebswirtschaftlicher Lösungen, die im BW neben den umfassenden Funktionalitäten zum Aufbau und Betrieb eines Data Warehouse zur Verfügung gestellt werden. Besonderes Augenmerk haben dabei Systemumgebungen, in denen SAP R/3 als vorgelagertes Quellsystem vorhanden ist.

Unter dem Begriff *mySAP Business Intelligence* sind Anwendungskomponenten und Dienstleistungen für die Entscheidungsunterstützung in Unternehmen zusammengefasst, deren Hauptkomponente das Business Information Warehouse als Basistechnologie zur Datenhaltung ist. Weitere Bestandteile sind u. a. der Business Explorer zur Datenauswertung und das Enterprise Portal als zentraler Zugangspunkt mit Single-Sign-On-Funktionalität zur Verteilung von Berichten und Informationen. Dies ist mittlerweile integraler Bestandteil von SAP NetWeaver. Das Strategic Enterprise Management, neuerdings Bestandteil der Lösung mySAP ERP (vormals mySAP Financials), beinhaltet Funktionen für Unternehmensplanung, Performance Management und Konsolidierung. Die Datenhaltungskomponente des SEM ist ebenfalls das BW.²

4.1 Architektur

Die Architektur des Business Information Warehouse ist angelehnt an die allgemeine Referenzarchitektur für ein Data Warehouse. Ausgehend von verschiedenen möglichen Quellsystemen erfolgt ein im Allgemeinen periodischer Datenladevorgang in die zentrale Datenhaltung des Business Information Warehouse (BW).

Im Mittelpunkt der Architektur steht der eigentliche Business Information Warehouse Server, der für die Ablage, Aufbereitung und Abfrage der Daten im BW verantwortlich ist. Die wesentlichen Objekte zur Datenablage sind dem Datenfluss folgend zunächst die Persistent Staging Area (PSA), der Operational Data Store (ODS) und die Info-Cubes. Während die PSA temporären Charakter hat, sind die Daten im ODS und in den Info-Cubes für die dauerhafte Ablage gedacht. In den relational strukturierten ODS-Objekten erfolgt die Speicherung von belegnaher Information, die Info-Cubes bilden die mehrdimensionalen Strukturen als Grundlage für OLAP Analysen ab.

Ausgehend von den Quellsystemen dient die Ebene der Persistent Staging Area der Speicherung der nicht aufbereiteten und nicht modifizierten Daten aus den vorgelagerten Quellsystemen. Eine erste Transformation ermöglicht die Datenweitergabe an die ODS-Objekte, die in der Darstellung der Konsolidierung verschiedener Datenquellen dienen, wobei auch ein direktes Befüllen der Info-Cubes aus der PSA unter Umgehung zwischengeschalteter ODS-Objekte möglich ist. In einer weiteren Stufe erfolgt die betriebswirtschaftliche Konsolidierung mit dem Aufbau von Info-Cubes [CGH05].

² Die Gruppierung und Begriffsbildung unterliegt häufigen Änderungen von Seiten der SAP AG.

Grundlage jeder Abfrage und Auswertung gespeicherter Daten im Business Information Warehouse sind die Queries, die bereits spezifische Sichten auf die vorhandenen Daten darstellen. Auf der Ebene der Queries erfolgt auch die Festlegung von berechneten bzw. abgeleiteten Kennzahlen, die nicht physisch gespeichert sind sondern erst zur Laufzeit kalkuliert werden. Jede Query bezieht sich stets auf genau einen Info-Provider, z. B. Info-Cubes oder ODS-Objekte.

Integraler Bestandteil des Business Information Warehouse ist das Werkzeug Business Explorer Analyzer als Excel-basiertes OLAP-Werkzeug. Der Bereich des Webreportings wird durch die neue Technologie auf Basis des Web Application Servers und dem Web Application Designer adressiert.

Eine grobe Übersicht der Architektur und des Zusammenspiels der beteiligten Komponenten des Business Information Warehouse ergibt sich aus Abbildung 4 [Ha03].

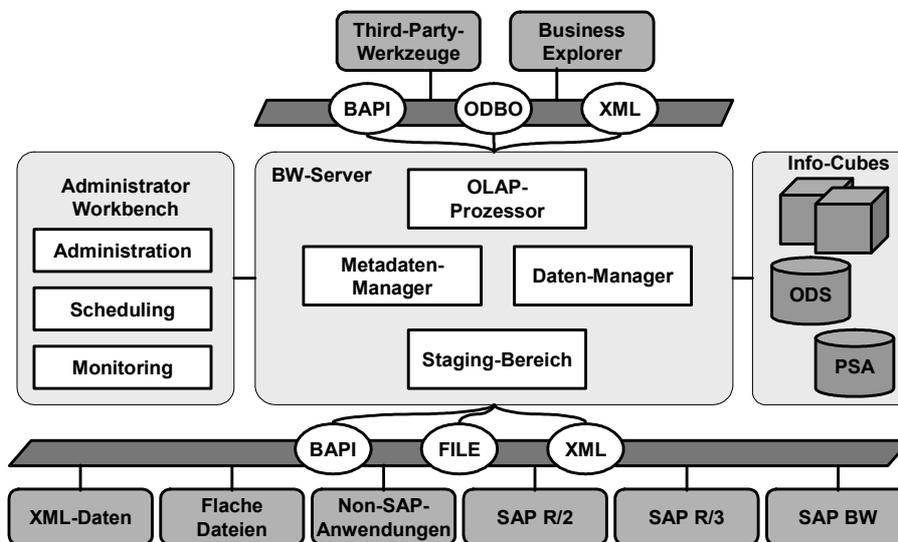


Abbildung 4: Architektur des SAP Business Information Warehouse

4.2 Datenmodell

Die Datenhaltung des Business Information Warehouse basiert auf der relationalen Datenbanktechnik und die Speicherung mehrdimensional strukturierter Informationen folgt dem Ansatz eines Star Schemas, das von der SAP AG um eigene Konstrukte erweitert wurde. Dieses so genannte erweiterte Star Schema im BW ist gekennzeichnet durch eine große Anzahl von Tabellen, über welche die Informationen verteilt abgelegt sind [Ha05].

Die elementaren Objekte im Business Information Warehouse sind die global definierten und über ihren technischen Namen eindeutig identifizierbaren Info-Objekte. Die im BW-Modell differenzierten betriebswirtschaftlichen Auswertungsobjekte sind zum einen die Merkmale, die den allgemeinen Dimensionselementen entsprechen, und die Kennzahlen, mit denen die Fakten beschrieben werden. Beide sind technisch gesehen Info-Objekte.

Der Info-Cube ist das zentrale Objekt zur Speicherung mehrdimensional strukturierter Daten im BW. Dass die Stammdaten übergreifend gültig sind und über alle Info-Cubes hinweg zur Verfügung stehen ist dabei ein wesentliches Grundprinzip. Die Faktentabelle bildet im erweiterten Star Schema den Mittelpunkt des Modells und ist von Dimensionen umgeben, die jeweils aus der Dimensionstabelle und den Stammdatentabellen bestehen. Dabei fungiert die Dimensionstabelle als Verbindungstabelle von der Faktentabelle zu den Stammdatentabellen, in denen sich die eigentlichen Dimensionswerte mit beschreibenden Texten sowie deren Attribute befinden (siehe Abbildung 5).

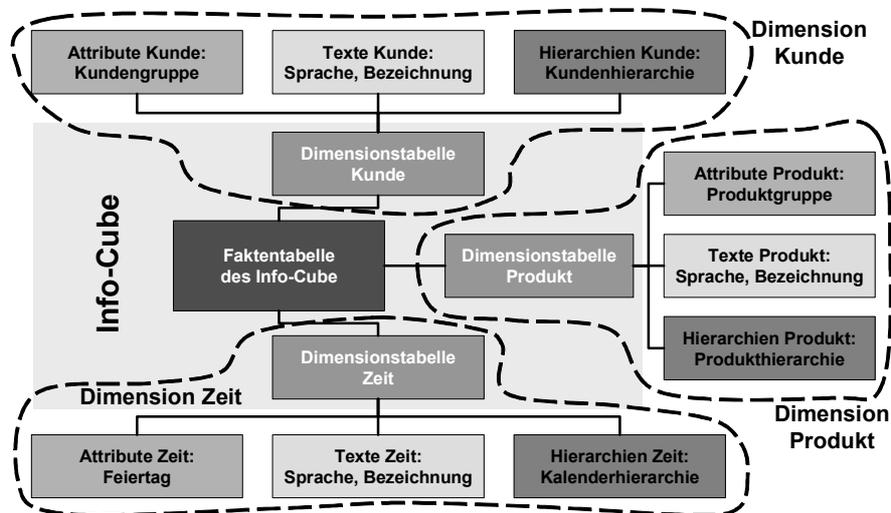


Abbildung 5: Erweitertes Star Schema des SAP Business Information Warehouse

Die Ablage der beschreibenden Texte von Merkmalen erfolgt im Business Information Warehouse in einer separaten Texttabelle, wobei die Texte in verschiedenen im System definierten Sprachen verfasst sein können. Die Berücksichtigung von Dimensionsattributen ist im erweiterten Star Schema über Stammdatentribute eines Merkmals gewährleistet. Diese werden in einer separaten Stammdatentabelle eines Merkmals gespeichert. Im BW-Sprachgebrauch wird allgemein von Attributen gesprochen, wobei zwischen reinen Anzeigeattributen und Navigationsattributen zu differenzieren ist. Letztere können auch ohne das zugrunde liegende stammdatenträgende Merkmal in Abfragen berücksichtigt werden.

Für die Modellierung hierarchischer Dimensionsstrukturen gibt es im Business Information Warehouse drei verschiedene Varianten. Neben den auf Stammdaten basierenden Varianten der Abbildung über Externe Hierarchien und über Navigationsattribute gibt es noch die Bewegungsdatensicht der Gestaltung über Merkmale.

Im erweiterten Star Schema des BW sind die physisch abgelegten Kennzahlen gerade die Wertspalten in der Faktentabelle, die somit keiner hierarchischen Strukturierung unterliegen. Die Berechnung dieser Kennzahlen erfolgt im Rahmen des ETL-Prozesses während der zumeist periodischen Aktualisierungsvorgänge des BW-Datenbestandes.

Die Abbildung umfassender Kennzahlensysteme erfolgt im Business Information Warehouse nicht auf der Ebene der physischen Datenstrukturen, sondern erst auf der Ebene der Abfragen. Diese so genannten Queries stellen im BW eine mehrdimensionale Sichtweise auf den in Info-Providern abgelegten Datenbestand dar und ermöglichen die Implementierung komplexer Berechnungen und Operationen mit berechneten und eingeschränkten Kennzahlen.

5 Implementierung des CWM im SAP BW

Für den Austausch von Metadaten nutzt das SAP BW die Möglichkeiten über die XML Metadata Interchange (XMI) Specification. Das zugrunde liegende Metamodell ist angelehnt an das Common Warehouse Metamodel und stellt im Sinne einer Erweiterung des CWM eigene Pakete, Klassen und Assoziationen zur Verfügung.

5.1 Erweitertes CWM im BW

Die Metadatenverwaltung im Business Information Warehouse benutzt für die Präsentation der abgelegten Informationen auf der Basis der Spezifikation des Common Warehouse Metamodels neun Pakete, in denen einzelne Klassen jeweils thematisch gruppiert abgelegt sind. Diese Pakete sind in Abbildung 6 in die einzelnen Schichten des CWM eingegliedert dargestellt.

Management	whm		
Analysis	transformation	bex	authorization
Resource	core	relational	olap
Foundation	types		foundation

Abbildung 6: Pakete des erweiterten CWM im SAP BW

Die Spezifikation der spezifischen Datentypen *Boolean*, *Integer*, *Name*, *String* und *Time* erfolgt in dem Paket *types*. In dem Paket *association* sind neben der ebenfalls *Association* benannten Klasse gerade die einzelnen Assoziationen wie beispielsweise für die Verbindung der Objekte entlang des Datenflusses im Business Information Warehouse abgelegt. Aber auch die in diesen Assoziationen verwendeten Attribute für Objekt-Eigenschaften wie etwa der Eigner, der Status und beschreibende Texte sowie Business Content Informationen sind in diesem Paket festgelegt.

In dem *core*-Paket erfolgen die Klassendefinitionen für Info-Objekte, Währungsumrechnungsarten, Personen, ABAP-Routinen und diverse (Quell-)Systeme in den Klassen *InfoObject*, *CurrencyTranslation*, *Person*, *RoutineABAP* sowie *SystemDB*, *SystemBW*, *SystemMyself* und Klassen für weitere Systeme wie etwa Flat-Files und generische Extraktoren (*SystemDB*, *SystemExternal*, *SystemFile*, *SystemGeneric* und *SystemR3*).

Das Paket *olap* umfasst im Business Information Warehouse im Wesentlichen die Klassen für die Info-Provider und Aggregate (u. a. die Klassen *InfoCube*, *InfoProvider* und *MultiProvider* sowie *VirtualCube* und *Aggregate*) aber auch der Reporting Agent findet hier seinen Eingang (Klassen *ReportAgentPackage* und *ReportAgentSetting*). Die zu berücksichtigenden Klassen für relational strukturierte ODS-Objekte, Info-Sets und Indizes sind in dem *relational*-Paket mit den Klassen *ODS*, *InfoSet* und *Index* zusammengefasst.

Zu den Inhalten des *transformation*-Paketes gehören u. a. die Klassen *TransferStructure*, *CommunicationStructure*, *InfoSource* und *UpdateRule* die damit die gesamten Transformationsmöglichkeiten des BW entlang des systeminternen Datenflusses abbilden. Weitere hier spezifizierte Klassen sind *FileSource*, *InfoSourceMapping*, *InfoSourceMD* und *TransferStructureMD*.

Die Aspekte der Datenauswertung auf Basis des BEx Analyzer finden ihren Niederschlag im Paket *bex*, dass die Klassen *ExcelWorkbook*, *InfoSetQuery*, *Query*, *Query-CalculatedKeyFigure*, *QueryFormula*, *QuerySelection*, *QueryStructure* und *Query-Variable* für die entsprechenden BEx-Objekte zur Verfügung stellt.

In dem Paket *authorization* sind die Klassen *Authorization*, *Role* und *UserGroup* zur Ablage der Berechtigungsinformationen entsprechend des SAP Rollenkonzeptes zusammengefasst.

Die für den Betrieb des Business Information Warehouse notwendigen Objekte der Prozessketten, Info-Packages und –Gruppen sind mit den Klassen *EventChain*, *InfoPackage* und *InfoPackageGroup* in dem Paket *whm* zusammengefasst. Hierin finden sich darüber hinaus auch die Klassen für die Objektstrukturierung mit Hilfe von Katalogen, Bereichen und Anwendungskomponenten (*InfoObjectCatalog*, *InfoArea* und *Application*).

5.2 Metadaten austausch

Für den Austausch der Metadaten des Business Information Warehouse steht eine Implementierung auf Basis des XMI-Standards zur Verfügung. Die XMI-Dokumente können dabei einerseits direkt aus dem System als Datei verarbeitet und andererseits über einen http-Service genutzt werden.

Die Ablage aller Metadaten erfolgt im Business Information Warehouse im Metadata-Repository. Aus diesem können die Metadaten zu nahezu allen BW-Objekten auf Basis der Metamodell-Spezifikation als Erweiterung des CWM exportiert werden. Teilweise steht auch ein Import dieser Metadaten zur Verfügung, so dass auch die Anlage oder Modifikation von Objekten im System möglich ist. Dieser Mechanismus steht derzeit aber nur für wenige Objekte zur Verfügung. Für den Import werden die eingehenden Metadaten zunächst in einer Queue gesammelt und dann durch den XML-Parser geschleust und nach erfolgreicher Überprüfung an die entsprechende Business API übergeben, die das Metadata-Repository füllt.

6 Zusammenfassung

Das von der Object Management Group verabschiedete Common Warehouse Metamodel (CWM) hat sich inzwischen als industrieweiter Standard zur Modellierung von Metadaten in Data Warehouse-Systemen sowie zu deren Austausch etabliert. Es bietet losgelöst von proprietären Metadatenstrukturen eine solide Basis auf der Grundlage anerkannter Standards zur Modellierung umfangreicher Metadaten im Kontext des Data Warehousing. Zur Erweiterung stehen flexible Möglichkeiten zur Verfügung, so dass auch systemspezifische Anforderungen mit berücksichtigt werden können.

Wesentliche Schwachpunkte des Common Warehouse Metamodels sind die fehlenden bzw. unzureichenden Metamodelle in den Bereichen der Berechtigungen, der Semantik von Data Warehouse-Inhalten, dem Qualitätsmanagement, des Berichtswesens und der Unterstützung organisatorischer Prozesse.

Die Implementierung des Metamodels im SAP Business Information Warehouse unterstützt Aspekte des Berichtswesens durch entsprechende Klassen und Assoziationen, eine Berücksichtigung der semantischen Ebene fehlt aber ebenso wie auch der organisatorischen Prozesse. Die Unterstützung auf der Ebene der Auswertung und Analyse ist ebenfalls nur sehr rudimentär, der gesamte Bereich der Web-Anwendungen etwa ist in dem Metamodel nicht mit berücksichtigt.

Die Ausführungen zeigen, dass das Common Warehouse Metamodel ein mächtiges Werkzeug zur Metadatenmodellierung und -verwaltung an die Hand gibt, dessen Akzeptanz aber wesentlich von der Implementierung und Nutzung der Werkzeug-Hersteller abhängt. Das Beispiel des SAP Business Information Warehouse belegt, dass hier sinnvolle Erweiterungen des Common Warehouse Metamodels möglich sind, die auch eine erste Diskussionsgrundlage für weitergehende Standardisierungen im Rahmen der Spezifikation sein können.

Literaturverzeichnis

- [Au02] Auth, G.: Prozessorientiertes Metadatenmanagement für Data Warehouse-Systeme. In (von Maur, E.; Winter, R. Hrsg.): Vom Data Warehouse zum Corporate Knowledge Center, Proceedings der Data Warehousing 2002, Physica-Verlag, Heidelberg, 2002.
- [BG01] Bauer, A.; Günzel, H.: Data Warehouse Systeme, dpunkt-Verlag, Heidelberg, 2001.
- [CGH05] Chameni, P.; Gluchowski, P.; Hahne, M.: Business Information Warehouse – Perspektiven betrieblicher Informationsversorgung und Entscheidungsunterstützung auf der Basis von SAP-Systemen, Springer-Verlag, Berlin Heidelberg, 2005.
- [Co97] Conrad, W.: Qualitätsmanagement in Data Warehouse-Projekten – Methoden und Verfahren für die Projektpraxis. In (Mucksch, H.; Behme, W. Hrsg.): Das Data Warehouse-Konzept, 2. Auflage, Gabler, Wiesbaden, 1997, S. 351-389.
- [De97] Devlin, B.: Data Warehouse: From architecture to implementation, Addison Wesley, New York et al., 1997.
- [DR00] Do, H. H., Rahm, E.: On Metadata Interoperability in Data Warehouses, Technical Report Nr. 01(2000), Institut für Informatik, Universität Leipzig, Leipzig, 2000.
- [Ga97] Gardner, S. R.: Data Warehouses and Metadata: The Importance of Metadata Management. In (Siu, B. et al. Hrsg.): Data Mining, Data Warehousing & Client/Server Databases, Proceedings of the 8th International Database Workshop, Industrial Volume, Singapore, 1997.
- [Gr69] Grochla, E.: Modelle als Instrument der Unternehmensführung. In: ZfbF 21 (1969), S. 382-397.
- [Ha03] Hahne, M.: Logische Datenmodellierung zur Abbildung mehrdimensionaler Datenstrukturen im SAP Business Information Warehouse. In (Weikum, Gerhard; Schönig, Harald; Rahm, Erhard Hrsg.): BTW 2003 Datenbanksysteme für Business, Technologie und Web, Tagungsband der 10. BTW-Konferenz, 26.-28. Februar 2003, Leipzig, Köllen Druck + Verlag, Bonn, 2003, S. 630-647.
- [Ha05] Hahne, M.: SAP Business Information Warehouse – Mehrdimensionale Datenmodellierung, Springer-Verlag, Berlin Heidelberg, 2005.
- [He02] Helfert, M.: Planung und Messung der Datenqualität in Data-Warehouse-Systemen, St. Gallen, 2002.
- [Ho99] Holthuis, J.: Der Aufbau von Data Warehouse-Systemen: Konzeption- Datenmodellierung- Vorgehen, 2. Aufl., Deutscher Universitäts-Verlag, Wiesbaden, 1999.
- [Hu97] Hufford, D.: Metadata Repositories: The Key to Unlocking Information in Data Warehouses. In (Barquin, R.; Edelstein, H. Hrsg.): Planning and Designing the Data Warehouse, Upper Saddle River, 1997.
- [KE96] Kemper, A.; Eickler, A.: Datenbanksysteme: Eine Einführung, Oldenbourg Verlag, München et al., 1996.
- [Ko61] Kosiol, E.: Modellanalyse als Grundlage unternehmerischer Entscheidungen. In: ZfbF 13 (1961), S. 318-334.
- [Le95] Lehner, F.: Modell und Modellierung. In (Lehner, F.; Hildebrand, K.; Maier, R., Hrsg.): Wirtschaftsinformatik: theoretische Grundlagen, Wien, 1995.
- [Ma00] Marco, D.: Building and Managing the Metadata Repository. A Full Lifecycle Guide, New York et al., 2000.
- [Me03] Melchert, F.: Das CWM als Standard für Metadaten im Data Warehousing. In (von Maur, E.; Winter, R. Hrsg.): Data Warehouse Management, Das St. Galler Konzept zur ganzheitlichen Gestaltung der Informationslogistik, Springer-Verlag, 2003.
- [MTH02] Meister, J.; Tapken, H.; Harren, A.: Metadaten austausch unter Einsatz von Korrespondenzmetamodellen. In (von Maur, E.; Winter, R. Hrsg.): Vom Data Warehouse zum Corporate Knowledge Center, Proceedings der Data Warehousing 2002, Physica-Verlag, Heidelberg, 2002.

- [Mü00] Müller, J.: Transformation operativer Daten zur Nutzung in Data Warehouses, Deutscher Universitäts-Verlag, Wiesbaden, 2000.
- [OM01a] Object Management Group (OMG): Common Warehouse Metamodel (CWM) Specification. Volume 1. Version 1.0, OMG, o.O., 2001.
- [OM01b] Object Management Group (OMG): Common Warehouse Metamodel (CWM) Specification. Volume 2 – Extensions, Version 1.0, OMG, o.O., 2001.
- [OM01c] Object Management Group (OMG) Architecture Board ORMSC: Model Driven Architecture (MDA), OMG, o.O., 2001.
- [OM03] Object Management Group (OMG): XML Metadata Interchange (XMI) Specification. Version 2.0, OMG, o.O., 2003.
- [Po96] Poe, V.: Building a Data Warehouse for Decision Support, Upper Saddle River, 1996.
- [Po02] Poole, J. et al.: Common Warehouse Metamodel : An Introduction to the Standard for Data Warehouse Integration, John Wiley & Son Inc., New York et al., 2002.
- [St98] Strahinger, S.: Ein sprachbasierter Metamodellbegriff und seine Verallgemeinerung durch das Konzept des Metaisierungsprinzips. Proceedings der Modellierung '98, Münster, 1998.