

Unternehmensweite Daten- und Informationsmodellierung

Diplomarbeit

im Fach Management

Studiengang Informationswirtschaft

der

Fachhochschule Stuttgart –

Hochschule der Medien

Carmen Stier

Erstprüfer: Prof. Dr.-Ing. Peter Lehmann

Zweitprüferin: Prof. Dr. Bettina Schwarzer

Bearbeitungszeitraum: 4. April 2004 bis 3. September 2004

Stuttgart, August 2004

Kurzfassung

Diese Arbeit behandelt das Thema „Unternehmensweite Daten- und Informationsmodellierung“. Nach einer kurzen Einführung in die Grundlagen der Modellierung und der Darstellung der Besonderheiten der Daten in einem Data Warehouse, werden das Multidimensionale Datenmodell und seine strukturellen Anomalien beschrieben. Darauf aufbauend werden Anforderungen zusammengestellt, die ein konzeptioneller Modellierungsansatz für die adäquate Darstellung multidimensionaler Daten erfüllen muss. Anschließend werden verschiedene Modellierungsmethoden vorgestellt und bewertet. Die Umsetzung des multidimensionalen Modells in einem Informationssystem wird anhand des Datenmodells des SAP Business Content beschrieben. Den Abschluss bildet die Auswertung einer Umfrage zu diesem Thema.

Schlüsselwörter: multidimensionale Datenmodellierung, Vorgehensmodell, SAP Business Content, Umfrage unternehmensweite Daten- und Informationsmodellierung

Abstract

The subject of this diploma thesis is enterprise data and information modeling. After a short introduction to the basics of modeling and the special characteristics of data in data warehouses, the multidimensional data model and its structural anomalies are described. Following that a catalogue of requirements a conceptual model should fulfill is compiled. Subsequently, some models are introduced and evaluated. The implementation of the multidimensional model in an information system is shown by the data model of the SAP Business Content. Finally the results of a survey on the topic enterprise data and information modeling are presented.

Keywords: multidimensional modeling, process model, SAP Business Content, survey enterprise data and information modeling.

Inhaltsverzeichnis

Kurzfassung	2
Abstract.....	2
Abbildungsverzeichnis	5
Tabellenverzeichnis.....	6
Abkürzungsverzeichnis	7
1 Einleitung	8
1.1 Motivation der Arbeit	8
1.2 Aufbau der Arbeit	8
2 Grundlagen.....	10
2.1 Grundsätze ordnungsmäßiger Modellierung.....	10
2.2 Das Drei-Ebenen-Konzept	12
2.2.1 Konzeptionelle Ebene	13
2.2.2 Logische Ebene.....	14
2.2.3 Physische Ebene	14
2.3 Daten im Data Warehouse	14
3 Das multidimensionale Datenmodell.....	16
3.1 Strukturelle Aspekte	16
3.2 Dynamische Aspekte	19
3.3 Dimensionsstrukturen und -anomalien	20
3.3.1 Betriebswirtschaftliche Dimensionstypen.....	20
3.3.2 Strukturelle Dimensionstypen	20
4 Modellierungsansätze	25
4.1 Anforderungen an Modellierungsansätze	25
4.2 Multidimensional Entity Relationship Model.....	27
4.3 Dimensional Fact Model.....	29
4.4 Multidimensional Data Model.....	31
4.5 Multidimensional Aggregation Cube.....	33
4.6 Application Design for Analytical Processing Technologies	36
4.7 Multidimensional Unified Modeling Language	40
4.8 Kubenstrukturmodell	42
4.9 Semantisches Data Warehouse-Modell	44

4.10	Vergleich der Modellierungsansätze	51
5	Vorgehensmodelle zur Entwicklung von Data Warehouse Schemata	52
5.1	Analyse der operativen Systeme	52
5.2	Analyse der Geschäftsprozesse	55
5.2.1	Bestimmung des Zielsystems	55
5.2.2	Analysieren des Geschäftsprozessmodells	56
5.2.3	Konzeptuelles Objektschema	56
5.2.4	Semantisches Data Warehouse Schema.....	57
5.2.5	Logisches Data Warehouse Schema	59
5.2.6	Physisches Data Warehouse Schema.....	60
5.3	Informationsbedarfsanalyse	60
5.3.1	Initialisierungsphase	61
5.3.2	Ist-Analyse.....	61
5.3.3	Soll-Analyse	62
5.3.4	Fachkonzeptentwurf.....	64
6	Business Content	65
6.1	Komponenten des Datenmodells	65
6.1.1	InfoObjekte.....	66
6.1.2	Hierarchien	69
6.1.3	InfoCubes	71
6.1.4	ODS-Objekte	73
6.1.5	InfoSets	73
6.1.6	MultiProvider.....	74
6.2	Grafische Modellierungsmethode.....	74
6.2.1	Darstellung von Dimensionen	74
6.2.2	Hierarchien	75
6.2.3	Darstellung von InfoCubes und MultiCubes.....	76
7	Auswertung der Umfrage.....	77
8	Fazit	82
	Literaturverzeichnis.....	83
	Copyrightverweis	87
	Erklärung	88

Abbildungsverzeichnis

Abbildung 2-1: Drei-Ebenen-Konzept mit zugeordneten Modellierungsansätzen.	12
Abbildung 3-1: Ein typischer OLAP-Würfel.....	18
Abbildung 3-2: Nicht-hierarchische Dimension.	20
Abbildung 3-3: Einfache Hierarchie in extensionaler und intensionaler Sicht.	21
Abbildung 3-4: Hierarchie mit unterschiedlichen Pfadlängen in extensionaler und intensionaler Sicht.....	22
Abbildung 3-5: Heterarchie in extensionaler und intensionaler Sicht.	23
Abbildung 3-6: Parallele Hierarchie in extensionaler und intensionaler Sicht.....	24
Abbildung 4-1: Grafische Notationselemente des MERMs.	28
Abbildung 4-2: MERM.	29
Abbildung 4-3: Grafische Notationselemente des DFM.	31
Abbildung 4-4: Dimensional Fact Model.	31
Abbildung 4-5: Grafische Notationselemente des MD.	32
Abbildung 4-6: MD-Modell.....	33
Abbildung 4-7: Grafische Notationselemente des MAC.....	33
Abbildung 4-8: MAC Modell.	35
Abbildung 4-9: Grafische Notationselemente von ADAPT.....	38
Abbildung 4-10: ADAPT Modell.	39
Abbildung 4-11 Grafische Notationselemente der m UML.	41
Abbildung 4-12: m UML Modell.....	42
Abbildung 4-13: Grafische Notationselemente des Kubenstrukturmodells.	43
Abbildung 4-14: Das Kubenstrukturmodell.	44
Abbildung 4-15: Grafische Notationselemente der intensionalen Sicht.	46
Abbildung 4-16: Grafische Notationselemente der extensionalen Sicht.....	46
Abbildung 4-17: Grafische Notationselemente der Basiskennzahlensicht.....	47
Abbildung 4-18: Grafische Notationselemente des Kennzahlensystems.	48
Abbildung 4-19: Integrierte Gesamtsicht des SDWM.	50
Abbildung 5-1: Vorgehensweisen bei der Entwicklung von Data Warehouse Strukturen.	52
Abbildung 5-2: Entwicklung des Data Warehouse Schemas aus den operativen Modellen.....	53
Abbildung 5-3: Entwicklung des Data Warehouse Schemas aus den Geschäftsprozessen.	55
Abbildung 5-4: Entwicklung des Data Warehouse Schemas aus der Informationsbedarfsanalyse.....	61
Abbildung 6-1: Übersicht über InfoProvider.....	66

Abbildung 6-2: Unbalancierte Hierarchie mit bebuchbaren und nichtbebuchbaren Knoten.....	70
Abbildung 6-3: Erweitertes Sternenschema des InfoCubes.	72
Abbildung 6-4: Logisches Modell der Dimension im SAP BW.....	75
Abbildung 6-5: Logisches Modell von verschiedenen Hierarchien im BW.....	75
Abbildung 6-6: Logisches Modell des Basis Cube's mit Dimension.	76
Abbildung 6-7: Logisches Modell des Multi Cubes.	76
Abbildung 7-1: Nutzung der Datenmodelle.....	78
Abbildung 7-2: Anforderungen an ein Modell.	79

Tabellenverzeichnis

Tabelle 4-1: Überblick über die Modellierungsansätze und ihre Ursprünge.....	25
Tabelle 4-2: Vergleich der Modellierungsansätze.	51
Tabelle 5-1: Abbildungsregeln zwischen SDWM und Star/ Galaxy Schema.	59
Tabelle 5-2: Abbildungsregeln zwischen Star/ Galaxy Schema und Relationenmodell.	59

Abkürzungsverzeichnis

ADAPT	A pplication D esign for A nalytical P rocessing T echnologies
BW	B usiness W arehouse
bzw.	b eziehungsweise
CEO	C hief E xecutive O fficer
DBMS	D aten b ank m anagementsystem
DFM	D imensional F act M odel
DV-Systeme	D aten v erarbeitungssysteme
DWH	D ata W arehouse
ERM	E ntity R elationship Model
etc.	e t c etera
IAS	I nteraktionsschema
KOS	K onzeptuelles O bjektschema
KOT	K onzeptueller O bjekttyp
MD	M ultidimensional D ata Model
MERM	M ultidimensional E ntity R elationship M odel
MML	M ultidimensional M odeling L anguage
_m UML	M ultidimensional U nified M odeling L anguage
o. A.	o hne A utor
o. D.	o hne D atum
ODS	O perational D ata S tore
OLAP	O nline- A nalytical P rocessing
OMT	O bject M odeling T echnique
SDWM	S emantisches D ata W arehouse M odell
SERM	S trukturiertes E ntity R elationship M odell
SID	S tammdaten I dentifikationsnummer
SOM	S emantisches O bjekt m odell
u.a.	u nter a nderem
UML	U nified M odeling L anguage
VES	V organgs- E reignis- S chema
z.B.	z um B eispiel

1 Einleitung

*„Die Zahl derer, die durch zu viele Informationen nicht mehr informiert sind, wächst.“
Rudolf Augstein¹*

1.1 Motivation der Arbeit

Der Bedarf an unternehmensweit integrierten und konsistenten Informationen für Entscheidungsträger nimmt zu. Laut dem Bericht „Unternehmen vernachlässigen ihre taktischen Informationen“ der Computer Zeitung wird dieser Bedarf durch die bisher vorhandenen Informationssysteme in Unternehmen nicht gedeckt, da 14 % der CEO'S kein Vertrauen in die eigenen Zahlen haben und 67 % sie lediglich von der Tendenz her für richtig halten.² Dabei existieren mit dem Prinzip der Corporate Information Factory und der technologischen Lösung z.B. des SAP Business Information Warehouse, realistische Ansätze, um eine unternehmensweit integrierte Informationsversorgung zu gewährleisten.³ Die Leistungsfähigkeit und der Erfolg eines solchen Informationssystems hängen jedoch von der Mächtigkeit des zu Grunde liegenden Datenmodells ab.⁴ Ziel dieser Arbeit ist es, aufzuzeigen, welche Lösungen zur konzeptionellen Modellierung der Datenbasis eines Data Warehouse existieren. Darüber hinaus sollte durch eine Umfrage ermittelt werden, ob und wie im Rahmen von Data Warehouse Projekten konzeptionelle Datenmodelle entwickelt werden.

1.2 Aufbau der Arbeit

Nach einer kurzen Einleitung werden im zweiten Kapitel einige Grundsätze zum Thema Datenmodellierung allgemein erläutert und auf die Besonderheiten von Daten im Data Warehouse hingewiesen.

¹ Verlag für Deutsche Wirtschaft AG (2004), URL: <http://www.zitate.de> (letzter Zugriff: 20. August 2004)

² o. A. (2004a), S. 2.

³ Für eine genauere Beschreibung der Ansätze und der Aspekte des Enterprise Data Warehousing Vgl.: Preusschoff, S. (2003): Enterprise Data Warehousing. Gegenstand, Ansätze, Technologien.

⁴ Vgl.: Hahne, M. (2004), URL: http://www.cundus.de/downloads/MKWI2004_Hahne_paper.pdf (letzter Zugriff: 20. August 2004), S. 1.

Anschließend werden im dritten Kapitel das Multidimensionale Datenmodell mit seinen strukturellen Komponenten und dynamischen Aspekten vorgestellt. Im vierten Kapitel werden zunächst Anforderungen, die eine Modellierungsmethode für multidimensionale Daten erfüllen sollte, erhoben. Anschließend werden verschiedene Modellierungsansätze⁵ beschrieben und bewertet.

Die einfachste Möglichkeit, um ein Datenmodell für das Data Warehouse zu kreieren, wäre, es aus einem bereits bestehenden unternehmensweiten Datenmodell, in welchem alle analyse-relevanten Daten bereits gekennzeichnet sind, abzuleiten. Da dieses in der Realität nur sehr selten existiert, werden im fünften Kapitel drei Vorgehensmodelle für die Identifizierung von reporting-relevanten Daten vorgestellt.

Multidimensionale Datenmodelle müssen jedoch nicht nur modelliert, sondern auch umgesetzt werden. Wie die Strukturen und Komponenten in einem Informationssystem wiedergegeben werden können, wird im sechsten Kapitel an Hand des Datenmodells des SAP Business Content aufgezeigt. Des Weiteren wird ein logisches grafisches Modell vorgestellt, welches speziell für das BW konzipiert wurde.

Im siebten Kapitel erfolgt die Auswertung der Umfrage und das achte Kapitel schließt mit dem Fazit.

⁵ Anmerkung: Die noch in der Umfrage angesprochene Objekt Modelling Technique wurde bei der Vorstellung der Modellierungsansätze durch die _mUML ersetzt, da die Wahrscheinlichkeit, dass diese in der Praxis Verwendung findet, aufgrund der Bekanntheit der UML höher lag.

2 Grundlagen

2.1 Grundsätze ordnungsmäßiger Modellierung

Bei einem Modell handelt es sich um ein vereinfachtes Abbild eines Ausschnitts der Wirklichkeit. Trotz dieser Vereinfachung wird zwischen Modell und Wirklichkeit Strukturähnlichkeit, bzw. Gleichheit gefordert, so dass man von Merkmalen des Modells auf Merkmale der Wirklichkeit schließen kann.⁶

Ein Modell sollte folgenden sechs Grundsätzen der ordnungsmäßigen Modellierung entsprechen:

1. Grundsatz der Richtigkeit

Der Grundsatz der Richtigkeit beinhaltet zwei Aspekte: Zum einen die syntaktische Richtigkeit und zum anderen die semantische Richtigkeit. Erstere betrifft die Konsistenz und Vollständigkeit in Bezug auf das Metamodell. Dies bedeutet, dass die im Meta-Modell als unabdingbar definierten Konstrukte verwendet und die im Schema verwendeten Modellierungselemente durch das Metamodell beschrieben werden müssen. Die Forderung nach der semantischen Richtigkeit beinhaltet, dass das Modell die Realität struktur- und verhaltenstreu beschreibt. Des Weiteren muss das Modell widerspruchsfrei sein.⁷

2. Grundsatz der Relevanz

Auch der Grundsatz der Relevanz umfasst zwei Facetten. Zum einen müssen alle relevanten Aspekte der Realwelt im Modell enthalten sein. Man spricht dabei auch von externer Minimalität. Zum anderen muss das Modell daraufhin überprüft werden, ob alle enthaltenen Informationen mit der Zielsetzung des Modells übereinstimmen. Informationen, die für den jeweiligen Verwendungszweck des Modells irrelevant sind, sollten entfernt werden. Dies entspricht der Forderung nach interner Minimalität. Zum Grundsatz der Relevanz zählt auch die Frage nach der Angemes-

⁶ Vgl.: Heinrich, L. / Heinzl, A. / Roithmayr, F. (2004), S. 436.

⁷ Vgl.: Rosemann, M. (1998); S. 6-7.

senheit des verwendeten Modells in Bezug auf seinen Verwendungszweck und die Modelladressaten.⁸

3. Grundsatz der Wirtschaftlichkeit

Dieser Grundsatz relativiert die anderen Ansprüche. So kann es aufgrund dieses Grundsatzes möglich sein, auf die strikte Einhaltung der anderen Grundsätze zu verzichten, wenn eine Einhaltung den Grundsatz der Wirtschaftlichkeit verletzen würde. Aufwandsreduzierend wären bspw. die Verwendung von Referenzmodellen und die Wiederverwendung von Modellbausteinen, sowie der Einsatz eines Modellierungstools. Bereits bei der Erstellung eines Modells ist darauf zu achten, seine Nutzungsdauer zu optimieren.⁹

4. Grundsatz der Klarheit

Ein Modell entspricht dem Grundsatz der Klarheit, wenn es strukturiert, verständlich bzw. zugänglich, übersichtlich und lesbar ist. Da diese Forderung adressatenindividuell ist, handelt es sich nicht um eine Forderung nach maximaler, sondern nach machbarer Klarheit. „Dabei darf der Nutzen, der aus der Aufdeckung eines Missverständnisses resultieren würde, nicht größer sein als die Aufwendungen, derer es bedarf, um die hierfür notwendige Klarheit herbeizuführen.“¹⁰

5. Grundsatz der Vergleichbarkeit

Auf Grund immer anschaulicherer Modellierungsmethoden und benutzerfreundlicherer Modellierungstools erfolgt die Arbeit an einem Modell vermehrt arbeitsteilig. Daher muss sichergestellt werden, dass die getrennt voneinander entwickelten Modelle konsolidiert und im Falle von Ist- und Sollmodellen vergleichbar sind. Die Erfüllung dieses Grundsatzes kann durch Modellierungsstandards unterstützt werden.¹¹

6. Grundsatz des systematischen Aufbaus

Wenn die Modellierung unter Berücksichtigung unterschiedlicher Sichten vorgenommen wurde, muss es möglich sein, diese einzelnen Sichten zu integrieren. Dazu ist ein sichtenübergreifendes Metamodell erforderlich.¹²

⁸ Vgl.: Rosemann, M. (1998), S. 7-8.

⁹ Rosemann, M. (1998), S. 8.

¹⁰ Rosemann, M. (1998), S. 9.

¹¹ Vgl.: Rosemann, M. (1998), S. 9-10.

¹² Vgl.: Rosemann, M. (1998), S. 10.

2.2 Das Drei-Ebenen-Konzept

Für die Entwicklung von klassischen Datenbanken hat sich der Entwurfsprozess über die drei Ebenen: konzeptionelle Ebene, logische Ebene und physische Ebene etabliert. Diese Entwurfsphasen sollen auch auf das Data Warehouse Umfeld übertragen werden. Dabei wird die konzeptionelle Ebene der fachlichen Modellierung zugerechnet, während die anderen Ebenen der softwaretechnischen Modellierung angehören.

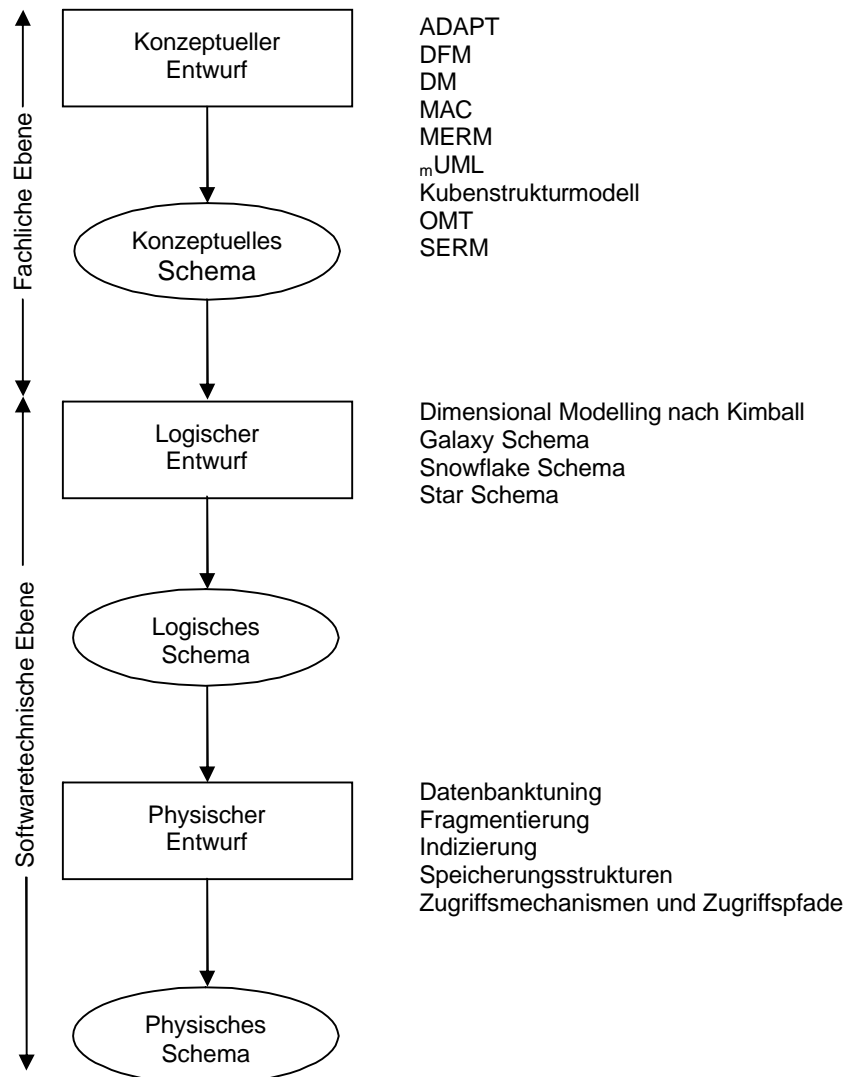


Abbildung 2-1: Drei-Ebenen-Konzept mit zugeordneten Modellierungsansätzen.¹³

¹³ Quelle: Böhnlein, M. (2001), S. 155.

2.2.1 Konzeptionelle Ebene

Bei der konzeptionellen Modellierung entsteht ein „[...] fachlich orientiertes Modell des abzubildenden Sachverhalts.“¹⁴ Dieses ist vollständig unabhängig vom Ziel-DBMS, damit mögliche Einschränkungen oder spezielle Eigenschaften des Ziel-Systems nicht berücksichtigt werden müssen.¹⁵ Dadurch wird eine natürliche, benutzernahe Darstellung des Sachverhalts ermöglicht,¹⁶ und das konzeptionelle Modell kann als Diskussionsgrundlage für Mitarbeiter der Fach- und der IT-Abteilung verwendet werden. Außerdem können konzeptionelle Modelle als Vorgabe für das Data Dictionary, bzw. Repositories eingesetzt werden und den Endanwender als Navigationshilfe im Datenpool unterstützen. Durch ihre Unabhängigkeit vom verwendeten System behalten konzeptionelle Modelle auch bei einem Tool- oder Release-Wechsel ihre Gültigkeit.¹⁷ Somit bilden sie eine Basis für eine langfristig ausgerichtete Systemdokumentation.¹⁸ Außerdem können sie „[...] bei auftretenden Modifikationen an den Datenstrukturen wertvolle Orientierungshilfen bieten und Änderungen damit vereinfachen.“¹⁹ Darüber hinaus hat das konzeptionelle Modell noch einen Koordinationseffekt. Da Data Warehouses meist schrittweise entwickelt werden, werden im unternehmensweiten Datenmodell die verschiedenen Aspekte des Data Warehouse integriert, zusammenhängend und vollständig modelliert, so dass alle Entwickler mit einem zentralen Bauplan arbeiten und die Teilergebnisse aufeinander abgestimmt sind.²⁰ Abhängig davon, wie viele Teilbereiche des Unternehmens damit abgedeckt werden, spricht Inmon im Falle eines begrenzten Bereichs von einem „corporate data model“, ansonsten von einem „enterprise data model“.²¹

¹⁴ Böhnlein, M. (2001), S. 152.

¹⁵ Vgl.: Vossen, G. (2000), S. 75.

¹⁶ Vgl.: Böhnlein, M. (2001), S. 152.

¹⁷ Vgl.: Gabriel, R. / Gluchowski, P (1997), S. 20.

¹⁸ Vgl.: Böhnlein, M. (2001), S. 152-153.

¹⁹ Gabriel, R. / Gluchowski, P (1997), S. 20.

²⁰ Vgl.: Inmon, W.H. (1997), URL: <http://www.inmoncif.com/library/whiteprs/techtopic/tt03.pdf> (letzter Zugriff: 20. August 2004), S. 1.

²¹ Vgl.: Inmon, W.H. (2000a), URL: <http://www.inmoncif.com/library/whiteprs/earlywp/ttdwdmod.pdf> (letzter Zugriff: 20. August 2004), S. 5.

2.2.2 Logische Ebene

In der logischen Entwurfsphase wird das konzeptionelle Modell in ein logisches Modell überführt.²² Auch dieses berücksichtigt noch keine Einzelheiten der physischen Repräsentation der Daten, ist aber bereits auf das zu verwendende konkrete Datenmodell ausgerichtet, welches nur bestimmte Beschreibungsmittel zulässt. Das logische Modell ist die Grundlage für die spätere Datenbankimplementierung.²³

2.2.3 Physische Ebene

Das physische Modell ist nicht mehr unabhängig vom einzusetzenden System, sondern behandelt Implementierungsaspekte wie Performance- und Kapazitätsgesichtspunkte, welche vom eingesetzten DBMS abhängig sind. Im physischen Modell werden das interne Datenbankschema und damit zusammenhängende Systemparameter, wie Fragmentierung oder Indexierung, definiert.²⁴

2.3 Daten im Data Warehouse

Die Daten im Data Warehouse unterscheiden sich in einigen Punkten von den Daten in den operationalen Systemen. Sie sind themenorientiert, integriert, zeitbezogen und nichtflüchtig.²⁵

- **Themenorientierung**

Dies bedeutet, dass die Daten in Dimensionen zusammengefasst werden, so dass Entscheidungsträger die Unternehmensdaten aus unterschiedlichen Blickwinkeln betrachten können.²⁶

- **Integration**

„Im Data Warehouse müssen für alle Daten und Informationen eindeutige Bezeichnungen vorliegen.“²⁷ Denn die operationalen DV-Systeme können sich als inkonsistente Datenquellen erweisen. So kann es vorkommen, dass in den einzelnen Daten-

²² Vgl.: Vossen, G. (2000), S. 77.

²³ Vgl.: Böhnlein, M. (2001), S. 153.

²⁴ Vgl.: Böhnlein, M. (2001), S. 153.

²⁵ Vgl.: Behme, W. / Mucksch, H.(2000), S. 9-13.

²⁶ Vgl.: Behme, W. / Mucksch, H.(2000), S. 10.

²⁷ Holthuis, J. (1999), S. 74.

beständen die Datenfelder identischer Daten unterschiedlich benannt (Synonyme) werden oder unterschiedliche Datenfelder gleich bezeichnet werden (Homonyme). Darüber hinaus müssen neben den Datenfeldbezeichnungen auch die unterschiedlichen Datenformate, z.B. Abkürzungen und Werteeinheiten, angepasst werden.²⁸

- **Zeitbezug**

Unter Zeitbezug werden mehrere Eigenschaften zusammengefasst. Die Daten in einem DWH werden zehn Jahre und länger gesammelt, um die Entwicklung des Unternehmens zu analysieren, damit Trends erkannt werden können. Außerdem werden zu den Schlüsselfeldern der einzelnen Datensätze während der Datenübernahme in das Data Warehouse Zeitmarken hinzugefügt, welche Zeitpunkte, abgeschlossene Zeiträume oder nicht abgeschlossene Zeiträume kennzeichnen. Des Weiteren kann der Nutzer den Zeitraum, über welchen er die Daten in seiner Analyse betrachten möchte, selbst bestimmen.²⁹

- **Nicht-Volatilität**

Volatilität beschreibt den Grad, mit dem sich Daten im Laufe ihrer normalen Nutzung ändern. Die Daten im DWH werden jedoch nach der fehlerfreien Übernahme in der Regel nicht mehr geändert oder aktualisiert. Änderungen sind nur dann erlaubt, wenn die Fehler im Rahmen der Datenübertragung aufgetreten sind oder wenn die Daten in den operationalen Systemen fehlerhaft waren und dort erst nach der Datenübertragung in das DWH korrigiert wurden. Dadurch, dass die Daten nicht volatil sind, wird sichergestellt, dass alle erstellten Analysen jederzeit nachvollzogen werden können.³⁰

²⁸ Vgl.: Behme, W. / Mucksch, H.(2000), S. 11-12.

²⁹ Vgl.: Behme, W. / Mucksch, H.(2000), S. 10-11.

³⁰ Vgl.: Behme, W. / Mucksch, H.(2000), S. 13.

3 Das multidimensionale Datenmodell

Die Daten in einem Data Warehouse werden nach für den Nutzer relevanten Entscheidungsobjekten angeordnet. Daher hat sich die multidimensionale Darstellung von Daten zur Modellierung der Datenbasis des DWH durchgesetzt,³¹ da die multidimensionale Datenstruktur der Sichtweise des Managements entspricht.³²

Die Grundlage des multidimensionalen Modells ist die Unterscheidung zwischen qualifizierenden und quantifizierenden Daten. Dabei werden die quantifizierenden Daten mit Hilfe der qualifizierenden Daten aus mehreren Blickwinkeln betrachtet.³³ Dies wird nachfolgend näher erklärt.

3.1 Strukturelle Aspekte

Das Datenmodell setzt sich aus mehreren Objekten zusammen, die seine Struktur bestimmen. Diese werden im Folgenden vorgestellt.

- **Kennzahlen**

Bei quantifizierenden Daten handelt es sich um betriebswirtschaftliche Kennzahlen, die auch als Variablen, Fakten, Measures und Measured Facts bezeichnet werden. In Übersetzungen werden Measures sowohl mit Kennzahlen als auch mit Maßeinheiten übersetzt, während man unter measured facts Fakten versteht.³⁴ Bei den Kennzahlen kann es sich um Basisgrößen oder abgeleitete Zahlen handeln.³⁵

- **Fakten**

Ein Fakt setzt sich zusammen aus einem Faktbezeichner und Faktwerten.³⁶ Er kann also mehrere Kennzahlen enthalten.

³¹ Vgl.: Wirtz, K. (2002), S. 59.

³² Vgl.: Schelp, J. (2000), S. 136.

³³ Vgl.: Böhnlein, M. / Ulbrich-vom-Ende, A. (2000):. URL: <http://www.seda.wiai.uni-bamberg.de/ceus/publikationen/downloads/BoUI2000.pdf> (letzter Zugriff: 20. August 2004), S. 4.

³⁴ Vgl.: Totok, A. (2000), S. 86.

³⁵ Vgl.: Böhnlein, M. / Ulbrich-vom-Ende, A. (2000):. URL: <http://www.seda.wiai.uni-bamberg.de/ceus/publikationen/downloads/BoUI2000.pdf> (letzter Zugriff: 20. August 2004), S. 4.

³⁶ Vgl.: Wirtz, K. (2002), S. 60.

- **Dimensionen**

Qualifizierende Daten werden Dimensionen genannt.³⁷ Sie beschreiben die Perspektiven, unter welchen die Kennzahlen betrachtet und analysiert werden können.³⁸ Dimensionen beantworten also die Fragen wo, wann und warum eine Kennzahl aufgetreten ist.³⁹ Typische Dimensionen sind bspw. Zeit oder Ort.

Um die Verständlichkeit des Datenmodells zu erhöhen, empfehlen Gabriel und Gluchowski, die Anzahl der Dimensionen pro Fakt auf sechs bis acht Dimensionen zu beschränken. Außerdem sollten die Hierarchiestufen innerhalb einer Dimension den Richtwert von sieben Hierarchiestufen nicht übersteigen.⁴⁰

- **Hierarchieebenen**

Innerhalb einer Dimension können die einzelnen Elemente einer Dimension hierarchisch zu Ebenen zusammengefasst werden. Diese Ebenen können durch nicht-dimensionale Attribute näher beschrieben werden.⁴¹ Allerdings werden in der Literatur teilweise auch diese als dimensionale Attribute bezeichnet.⁴²

- **Verdichtungsalgorithmus**

Die Struktur der Dimension legt fest, welche einzelnen Dimensionsebenen zu übergeordneten Ebenen verdichtet werden. Man spricht dabei auch vom Verdichtungs- oder Konsolidierungspfad. Nach welchem Algorithmus dies geschieht, wird aber erst durch den Verdichtungsalgorithmus angegeben. Die wichtigsten Aggregationsoperatoren sind Summe, Durchschnitt, Minimum, Maximum, Anzahl der Werte, Median und Modalwert.⁴³

- **Würfel**

Bei der Datenstruktur des multidimensionalen Modells handelt es sich um einen so genannten Hypercube. Dieser wird zur Veranschaulichung meist dreidimensional abgebildet, da in einer Grafik nicht mehr Dimensionen dargestellt werden können.⁴⁴ Die Dimensionen stellen dabei die Achsen des Würfels dar, auf welchen die Ele-

³⁷ Vgl.: Böhnlein, M. / Ulbrich-vom-Ende, A. (2000):. URL: <http://www.seda.wiai.uni-bamberg.de/ceus/publikationen/downloads/BoU12000.pdf> (letzter Zugriff: 20. August 2004), S. 4.

³⁸ Vgl.: Determann, L. (2002), S. 78.

³⁹ Vgl.: Herden, O. (2001), S. 19.

⁴⁰ Vgl.: Gabriel, R. / Gluchowski, P. (1998), S. 496.

⁴¹ Vgl.: Determann, L. (2002), S. 96-97.

⁴² Vgl.: Böhnlein, M. (2001), S. 137.

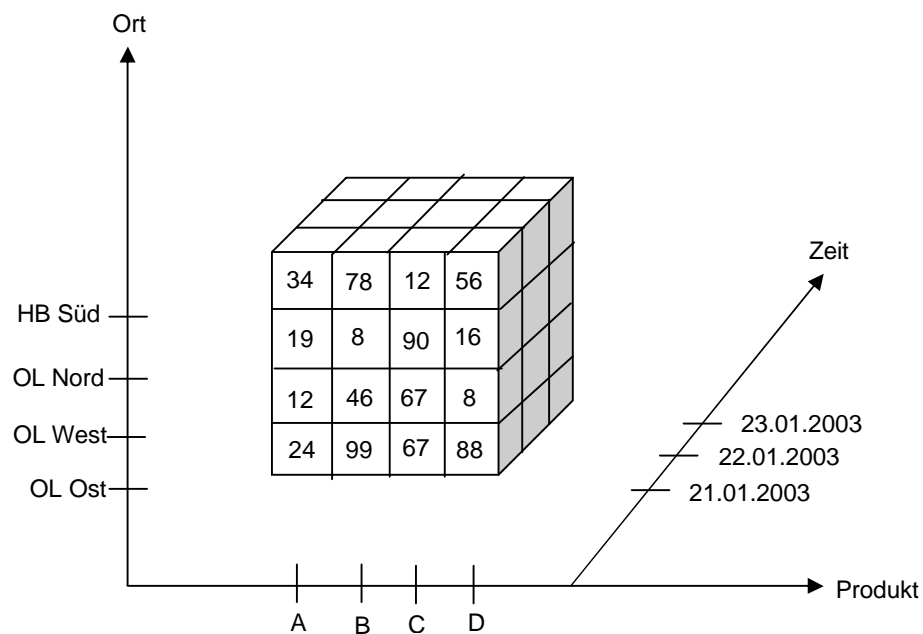
⁴³ Vgl. Wieken, J.-H. (1999), S. 136-137.

⁴⁴ Vgl.: Holthuis, J. (1999), S. 42.

mente einer Dimension abgetragen werden. Im Inneren des Datenwürfels befinden sich die Kennzahlen.⁴⁵

Ein Beispiel zum besseren Verständnis: Ein Unternehmer möchte die Verkaufszahlen von Produkten pro Tag und Filiale analysieren. Somit sind die Verkaufszahlen die betriebswirtschaftlichen Kennzahlen. Zeit, Ort und Produkt stellen die Dimensionen dar.

Abbildung 3-1: Ein typischer OLAP-Würfel.⁴⁶



⁴⁵ Vgl.: Böhnlein, M. / Ulbrich-vom-Ende, A. (2000):. URL: <http://www.seda.wiai.uni-bamberg.de/ceus/publikationen/downloads/BoUI2000.pdf> (letzter Zugriff: 20. August 2004), S. 4.

⁴⁶ Quelle: Herden, O. (2001), S. 20.

3.2 Dynamische Aspekte

Unter dem dynamischen Aspekt des Multidimensionalen Modells versteht man die Operationen, die man zur Navigation durch die Daten verwenden kann.⁴⁷

- **Drill Down**

Mittels des Drill Down wechselt man von einer höheren Verdichtungsstufe innerhalb einer Dimensionshierarchie zur nächstniederen, z.B. von Land zu Region.

- **Roll-Up**

Der Roll-Up ist die gegensätzliche Operation zum Drill Down. Hierbei wird von einer niedrigeren Verdichtungsstufe zu einer höheren gewechselt.

- **Selection**

Mit der Selection wählt der Nutzer bestimmte Würfeldata aus. Daher hat die Selection eine Filterfunktion. Die Operatoren Slice und Dice sind Spezialfälle der Selection.

- **Slice**

Die Slice-Operation entspricht dem Herausschneiden einer Scheibe aus dem Hypercube. Die Anzahl der zu betrachtenden Dimensionselemente wird dadurch verringert.

- **Dice**

Durch die Operation Dice wird ein Teilwürfel aus dem Hypercube betrachtet.

- **Rotate**

Dabei wird der Hypercube um seine Achsen gedreht, so dass der Nutzer durch das Tauschen der Reihenfolge der Dimensionen unterschiedliche Sichten erhält.

- **Nesting**

Durch das Nesting ist es möglich, mehrere Dimensionen in einer zweidimensionalen Matrix darzustellen, indem die Hierarchiestufen der Dimension in den Spalten und Zeilen geschachtelt dargestellt wird.

- **Olap Join**

Damit können mehrere Würfel verbunden werden, allerdings nur, wenn sie über gemeinsame Dimensionen verfügen.

⁴⁷ Vgl.: Böhnlein, M. / Ulbrich-vom-Ende, A. (2000):. URL: <http://www.seda.wiai.uni-bamberg.de/ceus/publikationen/downloads/BoUI2000.pdf> (letzter Zugriff: 20. August 2004), S. 5-7.

3.3 Dimensionsstrukturen und -anomalien

In den Dimensionen spiegeln sich die Merkmale wider, welche für die Nutzer bei der Analyse der Kennzahlen relevant sind. Diese Dimensionen lassen sich u.a. in betriebswirtschaftliche und strukturelle Dimensionstypen klassifizieren.⁴⁸

3.3.1 Betriebswirtschaftliche Dimensionstypen

Die betriebswirtschaftlichen Dimensionen werden wiederum unterschieden in Standarddimensionen und individuelle, bereichsabhängige Dimensionen. Dabei werden zum Beispiel die Dimensionen Zeit, Wertetyp/ Szenario (Istzahlen, Sollwerte, Plangrößen) und Maßeinheiten den Standarddimensionen zugerechnet. Dagegen werden Dimensionen wie Kunde, Artikel und Region zu den individuellen bereichsabhängigen Dimensionen gezählt, da sie für einen bestimmten Aufgabenbereich, z.B. für den Vertrieb, gelten.⁴⁹

3.3.2 Strukturelle Dimensionstypen

- **Nicht-hierarchische Dimensionen**

Bei nicht-hierarchischen Dimensionen besteht zwischen den Positionen keine vertikale Verbindung; somit ist es auch nicht möglich, die Positionen zu verdichten. Eine solche Dimension ist bspw. die Dimension Wertart mit den Positionen Istzahlen, Sollzahlen und Plangrößen.⁵⁰

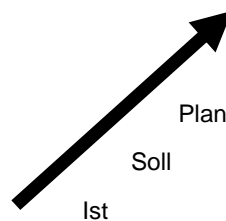


Abbildung 3-2: Nicht-hierarchische Dimension.⁵¹

⁴⁸ Vgl.: Holthuis, J. (1999), S 121-138.

⁴⁹ Vgl.: Holthuis, J. (1999), S 122-124.

⁵⁰ Vgl.: Holthuis, J. (1999), S. 124-125.

⁵¹ Quelle: Holthuis, J. (1999), S. 125

- **Hierarchische Dimensionen**

Wenn innerhalb einer Dimension vertikale Beziehungen zwischen ihren Positionen bestehen, so handelt es sich um eine Hierarchie mit unterschiedlichen Verdichtungsstufen.

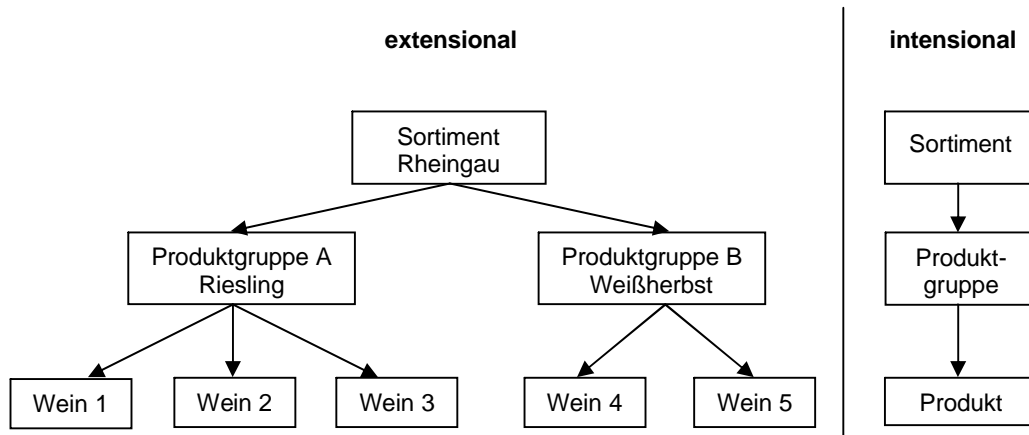


Abbildung 3-3: Einfache Hierarchie in extensionaler und intensionaler Sicht.⁵²

In einer Hierarchie wird zwischen Elementarpositionen und verdichteten Positionen unterschieden. Unter einer Elementarposition versteht man die Verdichtungsstufe mit dem niedrigsten Niveau, bzw. der niedrigsten Granularität, während es sich bei verdichteten Positionen um Klassifizierungen, Generalisierungen, Gruppierungen oder Aggregationen handelt, welche auch als Typkonstruktoren bezeichnet werden. Bei einer Klassifizierung werden gleichartige Entitäten zu einer Klasse zusammengefasst. Gleichartig heißt, dass die Entitäten durch gleiche Attribute beschrieben werden. Wenn mehrere Entitäten zu einer übergeordneten Entität zusammengefasst werden, bspw. Kunden und Lieferanten zu Geschäftspartnern, handelt es sich um eine Generalisierung. Das Gegenteil dazu ist die Spezialisierung, bei welcher ein Oberbegriff in Teilmengen zerlegt wird. Im Fall einer Gruppierung werden Elemente aus einer Entitätsmenge zu Gruppen zusammengefasst. Zum Beispiel verschiedene Weine zu Produktgruppen nach Rebsorte. Mittels der Aggregation beschreibt man eine Beziehung zwischen Objekten, “[...] die selbst als ein Objekt auf höherer Stufe abgebildet wird.”⁵³ Autoren schreiben zum Beispiel Artikel. Diese Beziehung wird dann als eigene Entität aufgefasst und zur Entität Literaturverzeichnis aggregiert.

⁵² In Anlehnung an: Holthuis, J. (1999), S. 126 und Schelp, J. (2000), S. 241.

⁵³ Holthuis, J. (1999), S.127.

giert.⁵⁴ Da in einer Hierarchie jede Dimensionsposition einer Hierarchieebene zugeordnet werden kann, gehören zu jeder Ebene bestimmte Eigenschaften. Mittels der Hierarchie werden die Navigationspfade, die auch als Konsolidierungs- oder Verdichtungspfade bezeichnet werden, bestimmt. Diese können für komplette Hierarchieebenen oder für einzelne Positionen festgelegt werden. Die Konsolidierungspfade bestimmen jedoch nur, welche Elemente verdichtet werden sollen, nicht jedoch nach welchem Verfahren. Dies geschieht erst im Verdichtungsalgorithmus.⁵⁵

Ein besonderer Typ der einfachen Hierarchien ist der kategorische Dimensionstyp. Dabei werden die Dimensionselemente nicht auf Grund ihrer organisatorischen Strukturen aggregiert, sondern auf Grund analyserelevanter Eigenschaften.⁵⁶

- **Hierarchien mit unterschiedlichen Pfadlängen**

Nicht immer sind Hierarchien so gleichmäßig wie im oben genannten Beispiel. Man unterscheidet zwischen ausgeglichenen und nicht ausgeglichenen Bäumen.

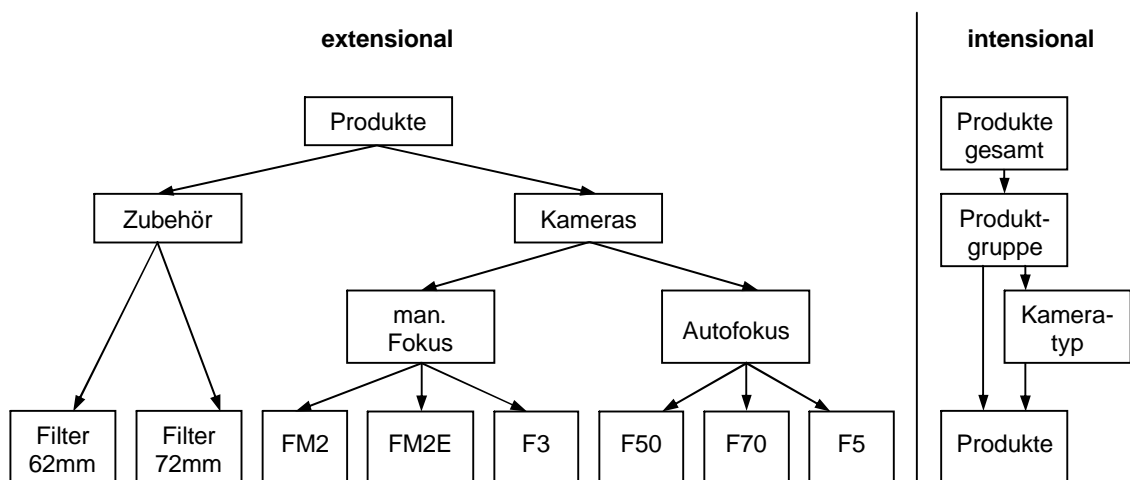


Abbildung 3-4: Hierarchie mit unterschiedlichen Pfadlängen in extensionaler und intensionaler Sicht.⁵⁷

⁵⁴ Vgl.: Schelp, J. (2000), S. 161.

⁵⁵ Vgl.: Wieken, J.-H. (1999), S. 136.

⁵⁶ Vgl.: Holthuis, J. (1999), S. 128.

⁵⁷ Quelle: Schelp, J. (2000), S. 247.

- **Heterarchien**

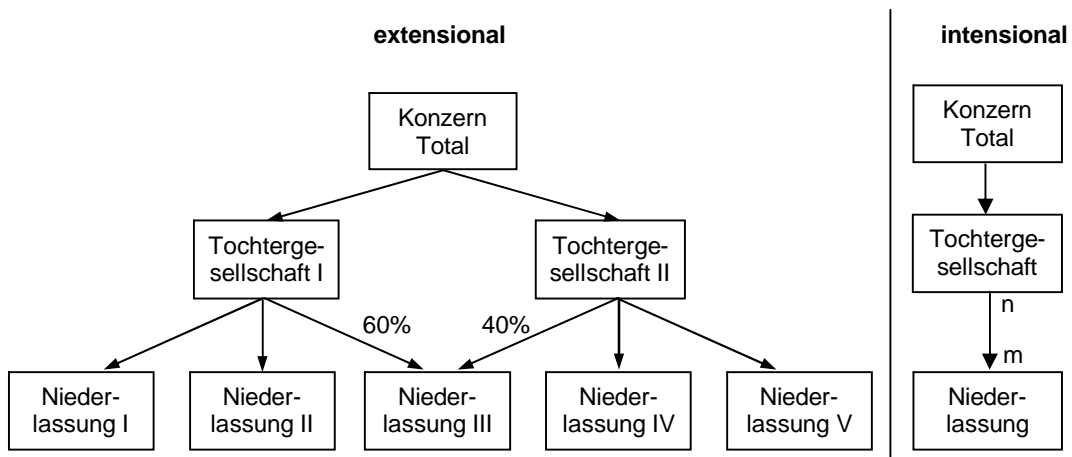


Abbildung 3-5: Heterarchie in extensionaler und intensionaler Sicht.⁵⁸

Bei Heterarchien handelt es sich um nicht mehr eindeutige Hierarchien, da in diesen eine Dimensionsposition anteilig zu mehreren Dimensionspositionen einer höheren Verdichtungsstufe verrechnet wird.⁵⁹

Es kann jedoch auch vorkommen, dass nicht alle Elemente einer Hierarchieebene verdichtet werden. Dann spricht man von einer nicht vollständigen Verdichtung.⁶⁰

- **Parallele Hierarchien**

Eine weitere Anomalie ist die parallele Hierarchie. Diese entsteht dadurch, dass Dimensionselemente nach unterschiedlichen Kriterien ohne anteilige Verrechnung gruppiert werden können.⁶¹ Wenn die verzweigenden Pfade wieder zusammengeführt werden, handelt es sich um einen alternativen Verdichtungspfad.⁶² So können Kunden beispielsweise nach Absatzregionen geordnet werden. Andererseits können sie aber auch nach Kundengruppen wie bspw. Fachhändler und Großabnehmer eingeteilt werden.⁶³ Holthuis empfiehlt im Falle einer parallelen Hierarchie, zwei eigenständige Dimensionen zu bilden, da sonst eine fehlerhafte Konsolidierung entsteht, wenn nicht überschneidungsfreie Dimensionselemente verdichtet werden. In jedem

⁵⁸ Quelle: Schelp, J. (2000), S. 244.

⁵⁹ Vgl.: Holthuis, J. (1999), S. 131.

⁶⁰ Vgl.: Herden, O. (2001), S. 21.

⁶¹ Vgl.: Schelp, J. (2000), S. 144.

⁶² Vgl.: Herden, O. (2001), S. 20.

⁶³ Vgl.: Schelp, J. (2000), S. 145.

Fall muss dem Nutzer gegenüber kenntlich gemacht werden, dass es auch alternative Verdichtungspfade gibt.

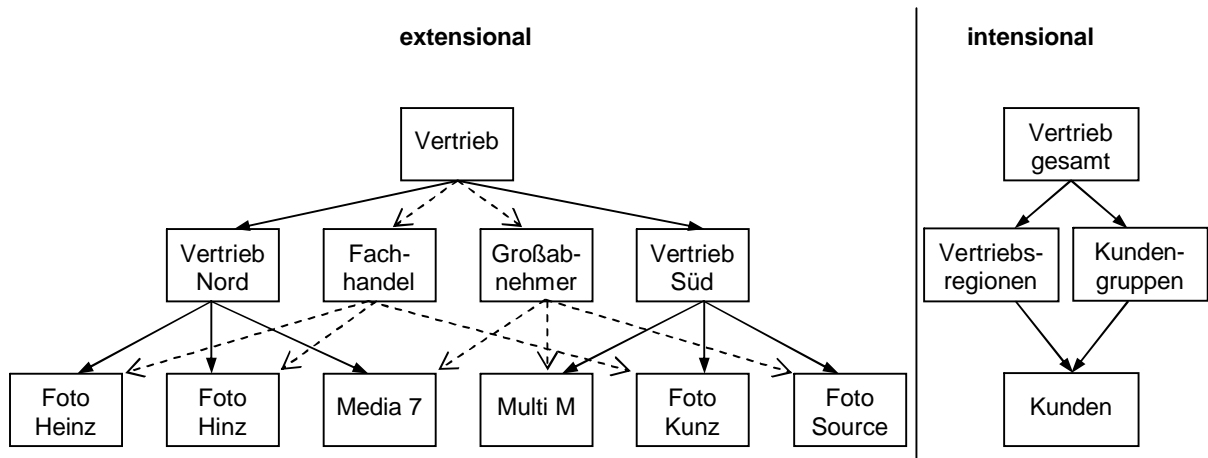


Abbildung 3-6: Parallele Hierarchie in extensionaler und intensionaler Sicht.⁶⁴

⁶⁴ Quelle: Schelp, J. (2000), S. 248.

4 Modellierungsansätze

Die hier vorzustellenden Ansätze für die multidimensionale Modellierung wurden nur teilweise speziell für die Darstellung multidimensionaler Datenstrukturen entwickelt. Manche beruhen auch auf herkömmlichen Modellierungsansätzen wie dem Entity Relationship Model (ERM) und objektorientierten Modellierungsmethoden.

Ansatz	MERM	DFM	DM	MAC	ADAPT	mUML	Kubenstruktur	SDWM
Ursprung								
ERM basierter Ansatz	x							
Multidimensionaler Ansatz		x	x	x	x		x	x
Objektorientierter Ansatz						x		

Tabelle 4-1: Überblick über die Modellierungsansätze und ihre Ursprünge.⁶⁵

4.1 Anforderungen an Modellierungsansätze

In der Literatur wird eine Reihe von Anforderungen an Modelle gestellt, die eine multidimensionale Datenstruktur beschreiben, welche sich hauptsächlich auf die angemessene Darstellung der multidimensionalen Sachverhalte konzentrieren. So muss es möglich sein, sowohl komplexe Dimensionen als auch komplexe Fakten darzustellen. Dazu gehört auch eine strikte Trennung von qualifizierenden und quantifizierenden Daten.⁶⁶

⁶⁵ In Anlehnung an: Determann, L. (2002), S. 129.

⁶⁶ Vgl.: Blaschka, M. / Sapia, C. / Höfling, G. (1999), URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/httpzSzzSzwww.forwiss.dezSz~system42zSzpublicationszSztechreport.pdf/sapia99overview.pdf> (letzter Zugriff: 20. August 2004), S. 4.

Zur Darstellung komplex strukturierter Fakten gehören folgende Anforderungen:⁶⁷

- Im Datenschema sollten beliebig viele Fakten aufgenommen werden können, um den ganzen Umfang des DWH abbilden zu können.
- Ein Fakt oder ein Datenwürfel sollte beliebig viele Kennzahlen enthalten dürfen, da zu einem Fakt meist mehrere Kennzahlen gehören. Diese sollten in einem Zusammenhang dargestellt werden.
- Beziehungen zwischen Fakten sollten abgebildet werden können, um das Verständnis für die Abhängigkeit der Daten zu vertiefen.
- Abgeleitete Kennzahlen sollten kenntlich gemacht und ihre Berechnungsvorschrift im Datenschema angegeben werden, da diese Attribute im Datenbestand häufig nicht enthalten sind und erst mittels einer Funktion aus anderen Kennzahlen berechnet werden.
- Auch die Additivität von Kennzahlen sollte explizit angegeben werden, da die Anwendung falscher Operatoren falsche Auswertungen zur Folge hat.

Zur Darstellung komplex strukturierter Dimensionen gehören folgende Anforderungen:⁶⁸

- Es muss möglich sein, Hierarchieebenen zu modellieren, damit die Daten auf einem für die spätere Analyse adäquaten Verdichtungsgrad dargestellt werden können.
- Die Darstellung von alternativen Verdichtungspfaden muss möglich sein.
- Ebenso müssen andere Beziehungen wie Generalisierung und Assoziation formuliert werden können.
- Parallele Hierarchien müssen dargestellt werden können.
- Die Darstellung von unbalancierten Hierarchien sollte möglich sein.
- Wenn es sich bei der Hierarchiestruktur um eine Heterarchie handelt, sollte die anteilige Verrechnung abgebildet werden können.
- Es sollte möglich sein, nicht-dimensionale Attribute für die Hierarchieebenen anzugeben.

Weitere Anforderungen sind:⁶⁹

- Es sollte eine Trennung zwischen Intension (Struktur) und Extension (Inhalt) bei der Modellierung vorgenommen werden, so dass bspw. Dimensionselemente auf

⁶⁷ Vgl.: Herden, O. (2001), S. 24-25.

⁶⁸ Vgl.: Herden, O. (2001), S. 24-25.

⁶⁹ Vgl.: Böhnlein, M. (2001), S. 162- 164.

der Ausprägungsebene getrennt von der Strukturierung einer Dimension betrachtet werden können.

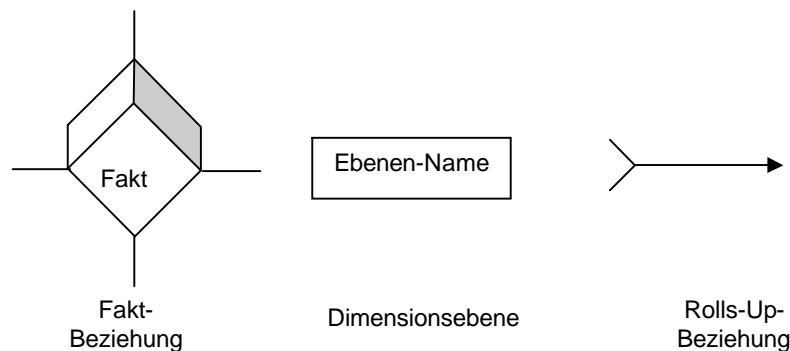
- Da Data-Warehouse-Schemata sehr komplex sein können, sollten die Modellierungsansätze über Hilfsmittel zur Beherrschung großer Schemata verfügen. Dabei kann es sich z.B. um Regeln zur Anordnung der Konstruktionselemente eines Schemas handeln, um dadurch die Lesbarkeit zu erhöhen,⁷⁰ oder um die Möglichkeit zur Sichtenbildung.
- Das entwickelte Schema sollte leicht in ein logisches überführt werden können.
- Um die Konsistenz- und Vollständigkeitsüberprüfung der Datenmodelle zu ermöglichen, sollte es ein Meta-Modell geben, da dieses über eine einheitlich vorgegebene Syntax und eine eindeutige Semantik für die Sprachkonstrukte verfügt.
- Der Modellierungsansatz sollte über ein Vorgehensmodell verfügen.
- Der Modellierungsansatz sollte durch ein Werkzeug unterstützt werden können.

4.2 Multidimensional Entity Relationship Model

Bei dem mehrdimensionalen Entity Relationship Modell (MERM) handelt es sich um eine Erweiterung des ursprünglichen ERMs nach Chen, welche von Sapia, Blaschka, Höfling und Dinter 1998 entwickelt wurde. Bei der Erweiterung des ERMs kam es den Entwicklern darauf an, dass alle neuen Elemente Spezialfälle der ursprünglichen Konstrukte sein sollten und so wenig neue Elemente wie möglich angefügt wurden. Dadurch sollten die Neuerungen schnell erlernt werden können. Trotzdem sollte es möglich sein, zwischen qualifizierenden und quantifizierenden Daten zu unterscheiden und Verdichtungen zwischen Dimensionsebenen darzustellen, da dies mit Chen's ursprünglichem Modell nicht möglich war. Daher haben sie folgende grafische Notationselemente hinzugefügt: eine Fakt-Beziehung, eine Dimensionsebene und eine Rolls-Up-Beziehung.⁷¹

⁷⁰ Vgl.: Jung, R. / Winter, R. (1998), URL: <http://www.ap.iwi.unibe.ch/publikationen/resource/mobis98.pdf> (letzter Zugriff: 20. August 2004), S. 4.

⁷¹ Vgl.: Sapia, C. / Blaschka, M. / Höfling, G. / Dinter, B. (1998), URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http:zSzzSzwww.forwiss.dezSz~system42zSzpublicationszSzdwdm98.pdf/extending-the-e-r.pdf> (letzter Zugriff: 20. August 2004), S. 5.

Abbildung 4-1: Grafische Notationselemente des MERMs.⁷²

Mittels der Rolls-Up-Beziehung wird die Verdichtung zwischen einer niedrigeren und einer höheren Dimensionsebene aufgezeigt. Dabei handelt es sich um einen gerichteten azyklischen Graphen, da ansonsten unendliche Rolls-Up-Pfade entstehen könnten. Dadurch ist es möglich, die hierarchische Strukturierung innerhalb von Dimensionen wiederzugeben. Auch können alternative Pfade und parallele Hierarchien modelliert werden. Darüber hinaus kann eine Dimensionsebene mit einer anderen Dimension geteilt werden, wodurch Redundanzen vermieden werden können. Außerdem ist es möglich, für jede Dimensionsebene zusätzliche Attribute anzugeben.⁷³

Bei der Fakt-Beziehung handelt es sich um eine Spezifikation der normalen Beziehung, welche einen Fakt repräsentiert. Sie verbindet die unterschiedlichen Dimensionsebenen, wobei diejenigen Ebenen, die direkt mit dem Fakt verbunden sind, als atomar bezeichnet werden, da sie die geringste Granularität aufweisen. Durch die Unterscheidung in Fakt-Beziehungen und Dimensionsebenen erfüllt das Modell die Forderung nach der Unterscheidung zwischen qualifizierenden und quantifizierenden Daten. Die Dimensionsebenen spiegeln die qualifizierenden Daten wider, während die quantitativen Daten durch die Attribute der Fakt-Beziehung repräsentiert werden. Allerdings gibt es keine Möglichkeit, abgeleitete Kennzahlen und die dazugehörige Regel darzustellen. Auch wird kein Symbol für die Dimension als solche angegeben, sondern nur für die einzelnen Dimensionsebenen, da die Entwickler eine spezielle Dimensionsangabe

⁷² In Anlehnung an: Sapia, C. / Blaschka, M. / Höfling, G. / Dinter, B. (1998), URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http%3A%2F%2Fwww.forwiss.de%2Fsystem42%2Fpublications%2Fszdwdm98.pdf/extending-the-e-r.pdf> (letzter Zugriff: 20. August 2004), S. 7.

für unnötig hielten, weil eine Dimension aus einer Reihe von Dimensionsebenen besteht.⁷⁴

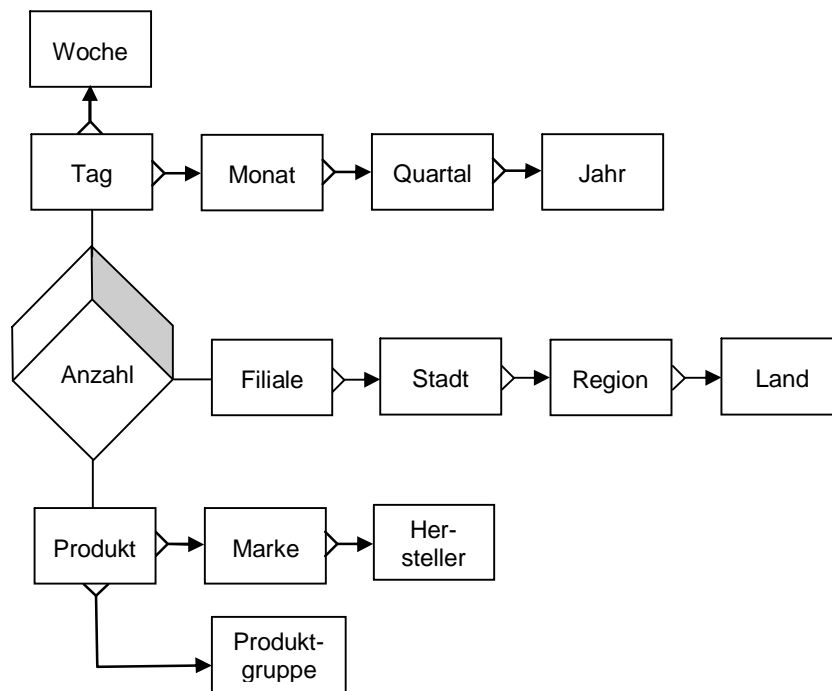


Abbildung 4-2: MERM.⁷⁵

4.3 Dimensional Fact Model

Matteo Golfarelli, Dario Maio und Stefano Rizzi haben 1998 das Dimensional Fact Model (DFM) als ein grafisches konzeptuelles Modell für Data Warehouses entwickelt.⁷⁶ Dabei wird das Multidimensionale Modell aus bereits bestehenden ER-Modellen der operativen Systeme halbautomatisch abgeleitet. Die Bausteine, aus denen sich das Modell zusammensetzt, sind Fakten mit ihren Kennzahlen und Hierarchien, welche Dimensionen und Dimensionsattribute beinhalten. Unter einem Fakt wird im DFM nicht

⁷³ Vgl.: Sapia, C. / Blaschka, M. / Höfling, G. / Dinter, B. (1998), URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http:zSzzSzwww.forwiss.dezSz~system42zSzpublicationszSzdwdm98.pdf/extending-the-e-r.pdf> (letzter Zugriff: 20. August 2004), S. 5-6.

⁷⁴ Vgl.: Sapia, C. / Blaschka, M. / Höfling, G. / Dinter, B. (1998), URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http:zSzzSzwww.forwiss.dezSz~system42zSzpublicationszSzdwdm98.pdf/extending-the-e-r.pdf> (letzter Zugriff: 20. August 2004), S. 5-6.

⁷⁵ Quelle: Herden, O. (2001), S. 27.

⁷⁶ Vgl.: Schelp, J. (2000), S. 187.

nur eine Kennzahl, sondern eine Reihe von zusammengehörenden Kennzahlen,⁷⁷ die einen betriebswirtschaftlich relevanten Betrachtungsgegenstand beschreiben,⁷⁸ verstanden. Die Fakten werden alle separat betrachtet und den entsprechenden Dimensionen zugeordnet. Allerdings ist es auch möglich, Fakten ohne Faktattribute darzustellen. Dadurch können Beziehungen zwischen dimensional Attributen, bspw. beim Wetter die Zeit und der Ort, modelliert werden.⁷⁹

Bei den Dimensionen handelt es sich nicht um eine Menge von Attributen und Hierarchien, sondern um das Attribut niedrigster Granularität innerhalb einer Hierarchie.⁸⁰

Aggregiert werden die Kennzahlen im Normalfall mittels der Addition. Jedoch ist dies nicht entlang aller Dimensionen möglich. Man unterscheidet diese Attribute in semi-additive und nicht-additive Faktattribute. Dabei sind semi-additive Attribute entlang einer oder mehrerer Dimensionen nicht addierbar, während nicht-additive Attribute entlang keiner Dimension summierbar sind.⁸¹ Sind die Faktattribute jedoch auf andere Weise aggregierbar, dann wird dies durch eine gestrichelte Linie zwischen Kennzahl und Dimension und der Angabe des alternativen Operators dargestellt.

Ein Fakt wird als zweigeteiltes Rechteck modelliert, in dessen oberen Bereich der Faktname und im unteren Bereich die Kennzahlen stehen. Von diesem Fakt zweigen die Dimensionen in einer Baumstruktur ab. Diese und ihre Attribute werden als Kreise dargestellt, welche die Hierarchieebenen verkörpern.⁸² Die Dimensionsattribute bilden dabei einen Pfad aufsteigender Granularität und werden durch ungerichtete Kanten verbunden, die eine 1:n Beziehung symbolisieren.⁸³ Wenn eine Beziehung zwischen zwei Dimensionsattributen nicht für jede ihrer Ausprägungen gilt, dann kann dies mittels einer optionalen Beziehung dargestellt werden. Dabei handelt es sich um einen Querstrich auf der verbindenden Kante. Außer den dimensional Attributen existieren noch die nicht- dimensional Attribute. Diese werden durch einen Strich dargestellt und können nur als Blätter im Dimensionsbaum vorkommen. Sie beschreiben Zusatzinformationen, die zur Hierarchiebildung ungeeignet sind.⁸⁴

⁷⁷ Vgl.: Determann, L. (2002), S. 138.

⁷⁸ Vgl.: Schelp, J. (2000), S. 188.

⁷⁹ Vgl.: Determann, L. (2002), S. 138.

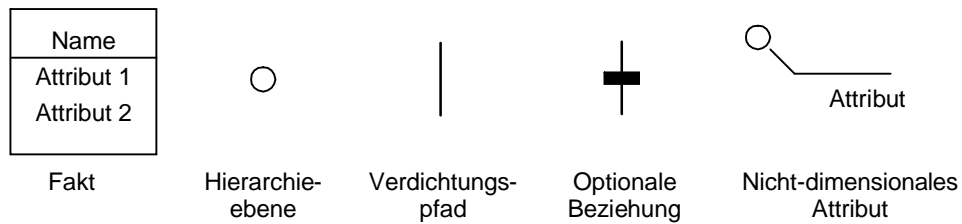
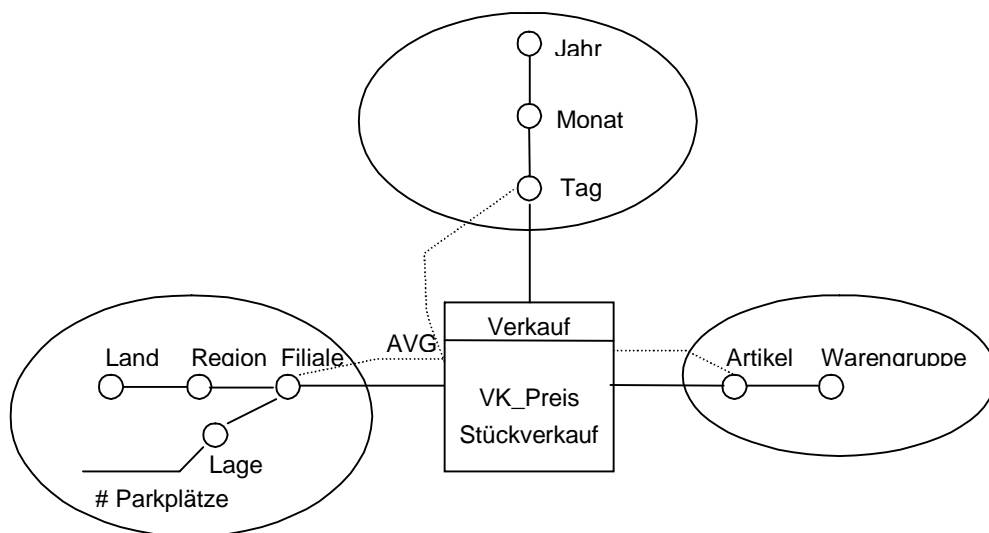
⁸⁰ Vgl.: Determann, L. (2002), S. 138.

⁸¹ Vgl.: Schelp, J. (2000), S. 189.

⁸² Vgl.: Herden, O. (2001), S. 32.

⁸³ Vgl.: Hettler, D. / Preuss, P. / Niedereichholz, J. (2003), S. 100.

⁸⁴ Vgl.: Herden, O. (2001), S. 32.

Abbildung 4-3: Grafische Notationselemente des DFM.⁸⁵Abbildung 4-4: Dimensional Fact Model.⁸⁶

4.4 Multidimensional Data Model

Die Entwickler des Multidimensional Data Model (MD) Luca Cabibo und Riccardo Torlone bezeichnen ihr Modell als ein logisches Modell.⁸⁷ Alberto Abelló, José Samos und Fèlix Saltor rechnen es jedoch zu den konzeptionellen Modellen, da es u.a. unab-

⁸⁵ Quelle: Herden, O. (2002), S. 32.

⁸⁶ Quelle: Determann, L. (2002), S. 139.

⁸⁷ Vgl.: Cabibbo, L / Torlone, R. (1998), URL: <http://cabibbo.dia.uniroma3.it/pub/pdf/edbt1998.pdf> (letzter Zugriff: 20. August 2004), S. 1.

hängig von einem bestimmten Datenbanksystem ist und darüber hinaus einen höheren Abstraktionsgrad besitzt als das Star Schema.⁸⁸

Die wichtigsten Modellierungsobjekte sind Dimensionen und f-Tabellen. Bei den Dimensionen handelt es sich hierbei tatsächlich um Kategorien, die es ermöglichen, Informationen aus unterschiedlichen Blickwinkeln zu betrachten. Jede Dimension besteht aus Ebenen, die, der Granularität der enthaltenen Daten entsprechend, hierarchisch angeordnet sind. Innerhalb einer Dimension können diese Ebenen mittels einer Roll-Up-Funktion verbunden werden. Auch ist es möglich, sie über Attribute näher zu beschreiben. In den f-Tabellen werden die betriebswirtschaftlichen Kennzahlen symbolischen Koordinaten, bzw. ihren entsprechenden Dimensionsebenen zugeordnet.⁸⁹

In der grafischen Darstellung werden die f-Tabellen durch einen Knoten definiert und wiedergegeben. Gleichzeitig verbindet er die Variablen und die Dimensionen. Die hierarchischen Beziehungen zwischen den Ebenen werden durch Pfeile dargestellt, welche die Roll-Up-Beziehung symbolisieren.⁹⁰



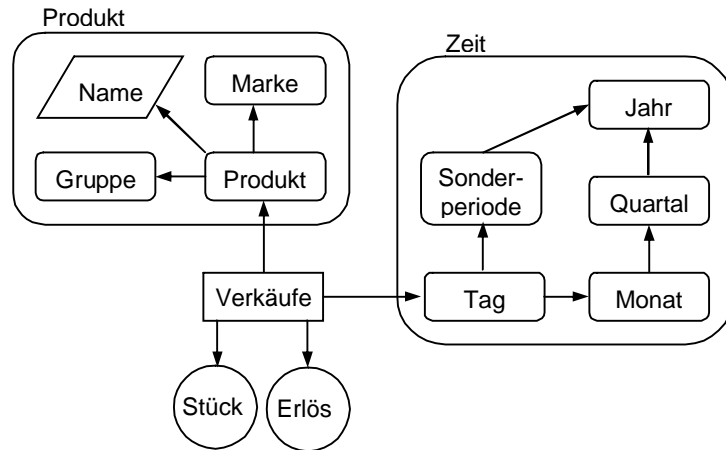
Abbildung 4-5. Grafische Notationselemente des MD.⁹¹

⁸⁸ Vgl.: Abello, A. / Samos, J. / Saltor, F. (2000), URL: <http://www-lsi.ugr.es/~bdf/Trabajos/lsi0006ugr.pdf> (letzter Zugriff: 20. August 2004), S. 8.

⁸⁹ Vgl.: Cabibbo, L / Torlone, R. (1998), URL: <http://cabibbo.dia.uniroma3.it/pub/pdf/edbt1998.pdf> (letzter Zugriff: 20. August 2004), S. 2.

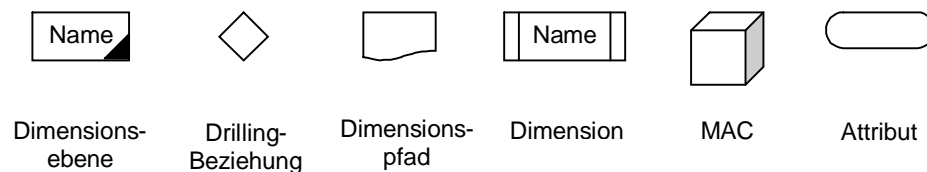
⁹⁰ Vgl.: Schelp, J. (2000), S. 196.

⁹¹ In Anlehnung an: Schelp, J. (2000), S. 196.

Abbildung 4-6: MD-Modell.⁹²

4.5 Multidimensional Aggregation Cube

Das Multidimensional Aggregation Cube (MAC) Modell wurde 2001 von Aris Tsois, Nikos Karayannidis und Timos Sellis als eine konzeptuelle und ausdrucksstarke Modellierungsmethode für das multidimensionale Modell vorgestellt. Das Modell setzt sich aus folgenden Elementen zusammen: Dimensionsebenen (Dimension Levels), Drilling-Beziehungen (Drilling Relationships), Dimensionspfaden (Dimension Paths), Dimensionen (Dimensions), Würfeln (Cubes) und Attributen (Attributes).⁹³

Abbildung 4-7: Grafische Notationselemente des MAC.⁹⁴

Eine Ebene besteht aus einer Reihe von Dimensionselementen (Dimension Members), welche die detailliertesten Elemente des Modells darstellen. Zu den einzelnen Ebenen können Attribute definiert werden. Dabei bildet eine Untermenge dieser Attribute einen

⁹² In Anlehnung an: Schelp, J. (2000), S. 196.

⁹³ Vgl.: Tsois, A. / Karavannidis, N. / Sellis, T. (2001), URL: <http://www.dbnet.ece.ntua.gr/pubs/uploads/TR-2001-5.pdf> (letzter Zugriff: 20. August 2004), S. 4.

⁹⁴ Quelle: Herden, O. (2001), S. 35.

Schlüssel für die Dimensionsebene. Die einzelnen Ebenen können über Drilling-Beziehungen miteinander verbunden werden. Das heißt, ein Element der höheren Ebene (Parent Level) wird mit einem oder mehreren Elementen der niedrigeren Ebene (Child Level) verbunden. Eine Reihe von Drilling-Beziehungen bildet einen Dimensionspfad. In einer einfachen Form gehört zu jeder Drilling-Beziehung jeweils ein Child Level, so dass jedes Child Level, außer dem letzten, gleichzeitig ein Parent Level ist. Mit diesen Pfaden werden die Wege festgelegt, über die spätere Drill-Down oder Roll-Up-Operationen ausgeführt werden. Das heißt, selbst wenn von einem Parent Level mehrere Drilling-Beziehungen ausgehen, wird die Drill-Down-Operation nur über die Beziehung ausgeführt, die Teil des Dimensionspfades ist, über welchen die Analyse stattfindet. Ein oder mehrere Dimensionspfade zusammen bilden eine Dimension, solange sie mindestens eine Ebene mit mindestens einem weiteren Pfad gemeinsam haben und eine semantische Beziehung zwischen den Pfaden besteht. Dabei müssen die gemeinsamen Ebenen auch in beiden Pfaden die gleiche Bedeutung haben. Die Anzahl aller möglichen Dimensionswerte wird als dimension domain bezeichnet.⁹⁵

Der Multidimensional Aggregation Cube verbindet letztendlich die Domänen der einzelnen Dimensionen miteinander. Bei den Attributen dieser Beziehung handelt es sich um die Kennzahlen des MACs. Eine Instanz dieser Beziehung wird als Zelle bezeichnet; in dieser befinden sich die Kennzahlen. Dabei kann es sich um Kennzahlen unterschiedlicher Granularität handeln, die auch voneinander abhängig sein können. Die Dimensionswerte, welche die Zelle verbindet, werden Koordinaten genannt. Die Kennzahlen des MACs sind funktional abhängig von ihren Koordinaten.⁹⁶

Eine Besonderheit des MAC-Modells ist die Möglichkeit, Analysepfade (Analysis Paths) zu definieren, „[...]die das Aufspüren der Dimensionshierarchien während der konzeptionellen Modellierung erleichtern soll.“⁹⁷ Außerdem existiert eine „ALL-Ebene“, die das Zusammenführen nicht vollständiger Verdichtungspfade ermöglicht.⁹⁸

⁹⁵ Vgl.: Tsois, A. / Karavannidis, N. / Sellis, T. (2001), URL: <http://www.dbnet.ece.ntua.gr/pubs/uploads/TR-2001-5.pdf> (letzter Zugriff: 20. August 2004), S. 4-7.

⁹⁶ Vgl.: Tsois, A. / Karavannidis, N. / Sellis, T. (2001), URL: <http://www.dbnet.ece.ntua.gr/pubs/uploads/TR-2001-5.pdf> (letzter Zugriff: 20. August 2004), S. 8-9.

⁹⁷ Herden, O. (2001), S. 20.

⁹⁸ Vgl.: Herden, O. (2001), S. 39.

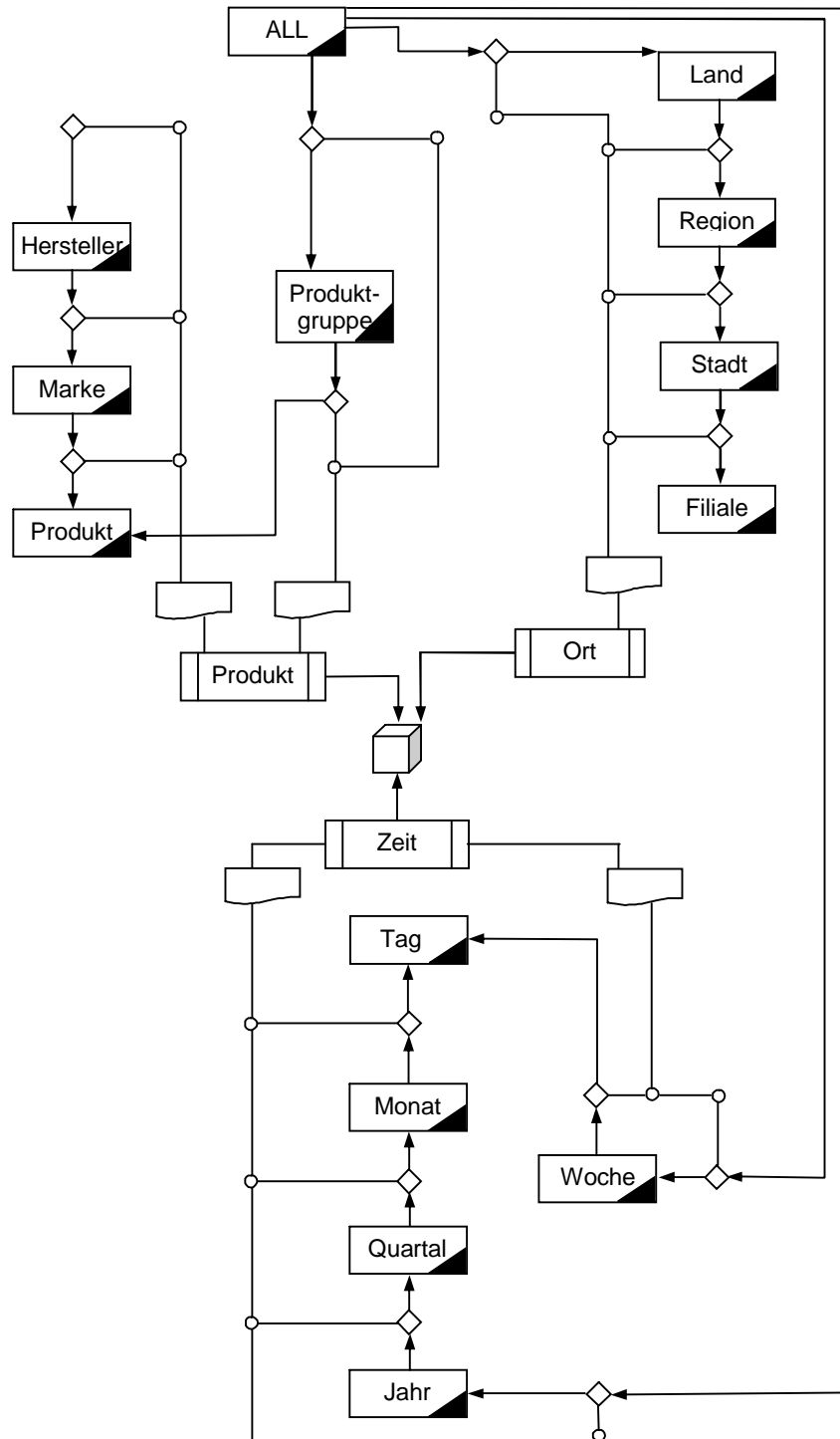


Abbildung 4-8: MAC Modell.⁹⁹

⁹⁹ Quelle: Herden, O. (2001), S. 36.

4.6 Application Design for Analytical Processing Technologies

Die 1996 von Dan Bulos vorgestellte Modellierungsmethode Application Design for Analytical Processing Technologies ADAPT, die speziell für Data-Warehouse-Anwendungen entwickelt wurde, wurde in der Vergangenheit häufig auf Grund ihrer umfangreichen Modellierungselemente und der dadurch bedingten schweren Verständlichkeit kritisiert. In der aktuellen Version 3.0 wurde die Anzahl der Elemente daher verringert.¹⁰⁰

Die einzelnen Elemente werden folgenden Gruppen zugeordnet: Kernobjekte (Core Objects), Dimensionsobjekte (Dimension Objects), Würfelobjekte (Cube Objects), Beziehungsobjekte (Relationship Objects) und Teilmengenobjekte (Subset Objects).¹⁰¹

Die Kernobjekte des Modells sind Würfel (Cubes) und Dimensionen (Dimensions). Teilweise werden auch Formeln (Models) dazugezählt.¹⁰² Der obere Teil des Würfels enthält entweder den Namen oder die Kennzahlen des Würfels, während der untere Teil die zugehörigen Dimensionen beinhaltet. Dadurch können Kennzahlen, welche die gleichen Dimensionen nutzen, in einem gemeinsamen Würfel modelliert werden.¹⁰³ Mittels Formeln lassen sich Berechnungen darstellen. Allerdings wird nicht die exakte Berechnungsformel angegeben.¹⁰⁴

Die Dimensionen legen die Struktur der Daten fest. In der ursprünglichen ADAPT-Version wurden die Dimensionen in sechs unterschiedliche Dimensionstypen eingeteilt: Aggregierende, Sequentielle, Eigenschafts-, Versions-, Kennzahlen- und Tupeldimension.¹⁰⁵ Diese Elemente sind in der aktuellen Version nicht mehr enthalten. Über Hierarchien, welche zu den Dimensionsobjekten zählen, werden die Aggregationsbeziehungen innerhalb von Dimensionen angegeben, wobei die Hierarchieebenen (Levels) die Reihenfolge innerhalb einer Dimension bestimmen. Dabei enthält die oberste Ebene die Daten der höchsten Aggregationsstufe und die niedrigste die Daten mit der geringsten Granularität. Zusätzlich können über den Dimensionswert (Member) einzelne Ausprä-

¹⁰⁰ Vgl.: Hettler, D. / Preuss, P. / Niedereichholz, J. (2003), S. 99.

¹⁰¹ Vgl.: o. A. (o. D.), URL: www.symcorp.com (letzter Zugriff: 20. August 2004).

¹⁰² Dora Hettler, Peter Preuss und Joachim Niedereichholz zählen die Formeln in ihrem Artikel „Vergleich ausgewählter Ansätze zur semantischen Modellierung von Data- Warehouse- Systemen“ zu den Kernelementen. In der ADAPT-Hilfdatei werden sie allerdings den Dimensions- und den Würfelobjekten zugeordnet.

¹⁰³ Vgl.: Hettler, D. / Preuss, P. / Niedereichholz, J. (2003), S. 99.

¹⁰⁴ Vgl.: Bulos, D. / Forsman, S. (1998), URL: http://www.symcorp.com/downloads/ADAPT_white_paper.pdf (letzter Zugriff: 20. August 2004), S. 13.

¹⁰⁵ Vgl.: Determann, L. (2002), S. 143.

gungen einer Hierarchieebene dargestellt werden. Es ist möglich, die einzelnen Hierarchieebenen durch nicht-dimensionale Attribute (Attributes) näher zu beschreiben.¹⁰⁶ Außerdem können Dimensionswerte auch in einer Teilmenge einer Dimension (Scope) zusammengefasst werden. Auch für die Teilmenge eines Hypercubes ist ein Modellierungsobjekt vorgesehen, welches zur Gruppe der Würfelobjekte gehört.¹⁰⁷

So vielfältig wie die Modellierungsobjekte der Daten sind auch die Beziehungstypen, die sie verbinden. Ein Connector, der durch einen einfachen Pfeil dargestellt wird, verbindet ein Objekt mit einem untergeordneten Objekt. Mittels eines Doppelpfeils wird eine so genannte Strict-Precedence-Beziehung beschrieben. Damit wird jeder Dimensionswert der niedrigeren Hierarchieebene direkt der übergeordneten zugeordnet. Trifft dies nicht zu, verwendet man das Symbol der Loose Precedence. Dadurch lassen sich variierende Pfandlängen, bzw. unbalancierte Hierarchien veranschaulichen. Mit der Self-Precedence wird eine Hierarchie beschrieben, die keine Ebenen hat. Darüber hinaus kann mit der Depends-On-Beziehung angezeigt werden, dass ein bestimmtes Objekt auch von einer anderen Dimension abhängt. Eine ähnliche Verbindung ist die Varies-By-Beziehung. Damit wird dargestellt, dass ein Objekt unterschiedliche Ausprägungen annehmen kann. Diese wiederum hängen von der betrachteten Dimension ab. Elemente, welche in einer Formel verwendet werden, werden über die Used-By-Beziehung verbunden.¹⁰⁸

Damit die Beziehung zwischen einem Objekt und seiner Teilmenge exakt angezeigt werden kann, gibt es vier Subset-Objekte. Mittels des Fully Exclusive Subsets wird angezeigt, dass die Obermenge vollständig auf die Teilmengen aufgeteilt wird, wobei jede Teilmenge einzigartig ist, da Elemente, welche in einer Teilmenge enthalten sind, nicht in der anderen verwendet werden dürfen. Diese Form der Aufteilung wird auch als vollständige disjunkte Aufteilung bezeichnet. Es ist auch möglich die Obermenge vollständig, aber nicht disjunkt, aufzuteilen. Dabei kann ein Element einer Teilmenge auch Teil einer anderen Teilmenge sein. In der ADAPT-Notation wird dafür das Fully Overlapping Subset Symbol verwendet. Wird die Obermenge nicht vollständig aufgeteilt, verwendet man die Notationselemente Partially Exclusive Subset und Partially Overlapping Subset, wobei ersteres eine disjunkte und letzteres eine nicht-disjunkte Aufteilung symbolisiert.¹⁰⁹ In der ursprünglichen ADAPT-Version gab es noch weitere Elemente, die

¹⁰⁶ Vgl.: Hettler, D. / Preuss, P. / Niedereichholz, J. (2003), S. 99-100.

¹⁰⁷ Vgl.: o. A. (o. D.), URL: www.symcorp.com (letzter Zugriff: 20. August 2004).

¹⁰⁸ Vgl.: o. A. (o. D.), URL: www.symcorp.com (letzter Zugriff: 20. August 2004).

¹⁰⁹ Vgl.: o. A. (o. D.), URL: www.symcorp.com (letzter Zugriff: 20. August 2004).

zur Darstellung von logischen und physischen Aspekten verwendet wurden, wie z.B. ein SQL-Durchgriff. Diese wurden aus der aktuellen Version entfernt.¹¹⁰

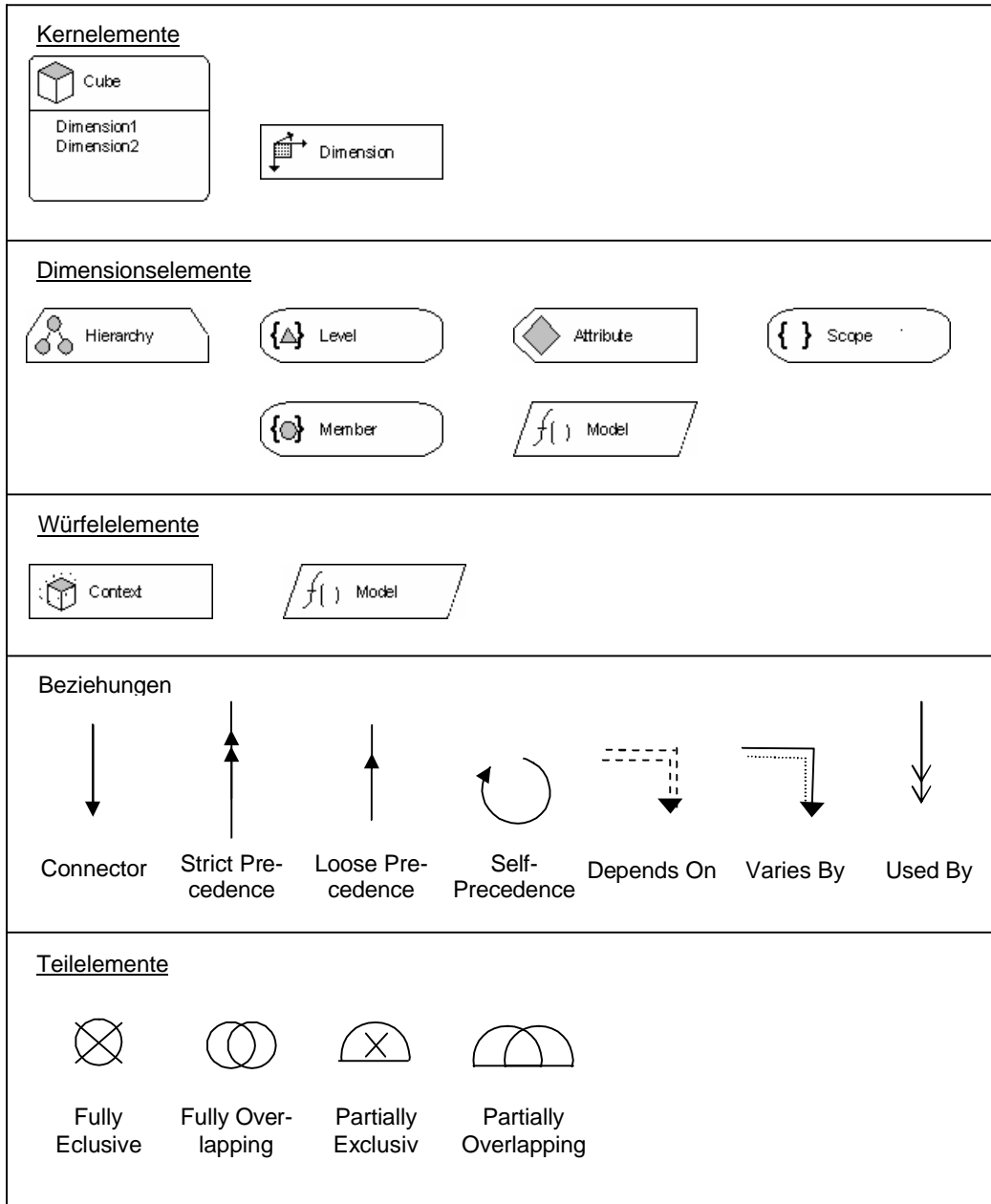


Abbildung 4-9: Grafische Notationselemente von ADAPT.¹¹¹

¹¹⁰ Vgl.: Hettler, D. / Preuss, P. / Niedereichholz, J. (2003), S. 99-100.

¹¹¹ In Anlehnung an o. A. (o.D.), URL: www.symcorp.com (letzter Zugriff: 20. August 2004).

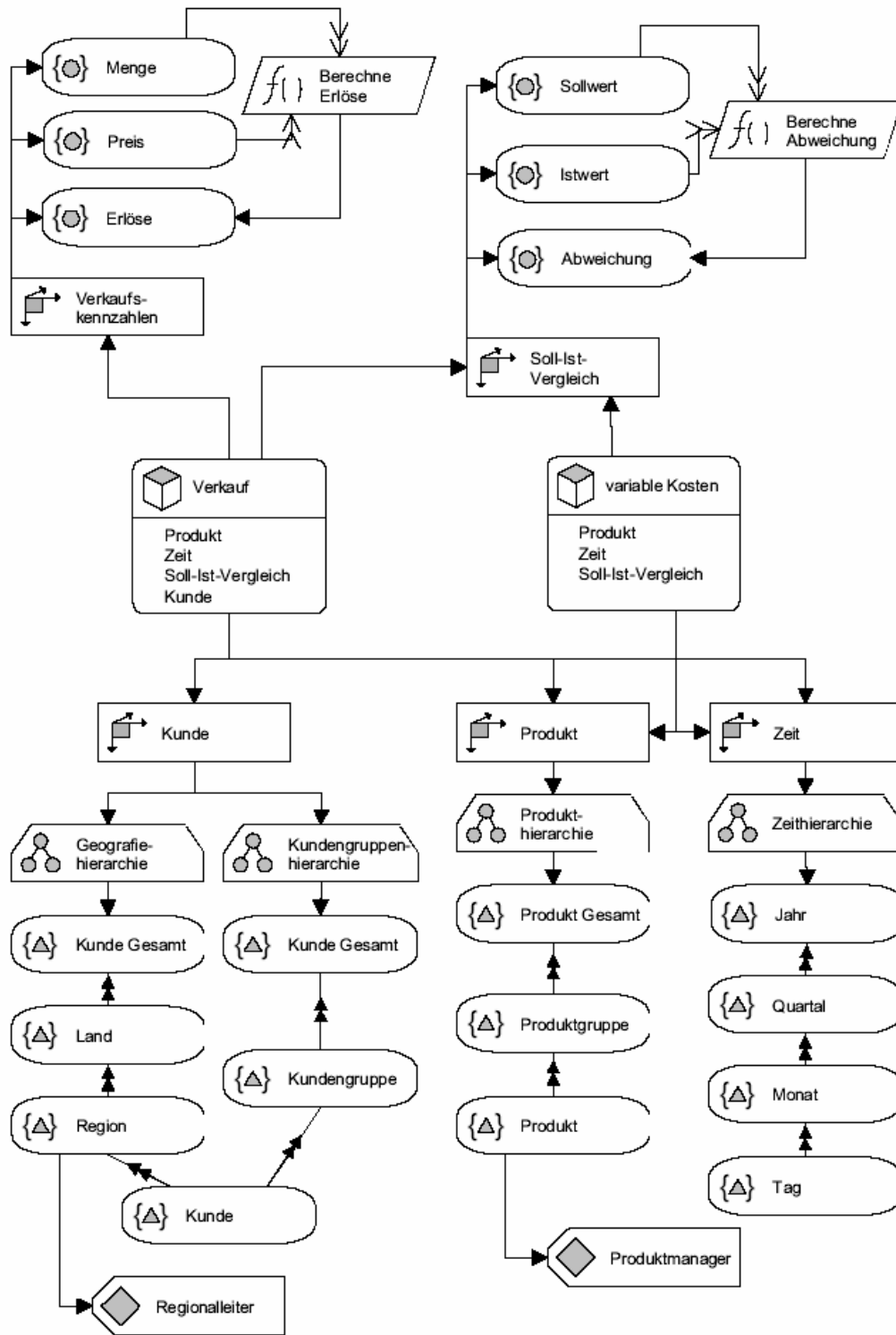


Abbildung 4-10: ADAPT Modell.¹¹²

¹¹² Quelle: Hettler, D. / Preuss, P. / Niedereichholz (2003), S. 101.

4.7 Multidimensional Unified Modeling Language

Die Multidimensional Modeling Language (MML), die an dieser Stelle nicht näher erklärt wird,¹¹³ wurde 1999 von Arne Harren als eine multidimensionale Sprache für die konzeptionelle Data Warehouse Datenmodellierung vorgestellt. Da es für die grafische Notation der MML jedoch keine Werkzeugunterstützung gibt, werden die Schemata mittels einer Erweiterung der UML, der *mUML*, dargestellt. Diese wird durch das CASE Werkzeug Rational Rose unterstützt.¹¹⁴

Das *mUML*-Diagramm wird durch das static structure diagram der UML, also dem Klassendiagramm dargestellt. Dimensionsebenen und Fakten werden als Dimensional-Classes und Fact-Classes modelliert, wobei ihre Zugehörigkeit zu diesen Metaklassen durch Stereotypen definiert werden, die dem Namen der Metaklasse entsprechen. Ihre Attribute, die in der MML in gesonderten Metaklassen angegeben wurden, werden hierbei in einem speziellen Teil des Klassensymbols in der Form `Attributname [Multiplizität] : Typ` eingetragen. Handelt es sich um ein optionales Attribut, so wird die Multiplizität mit „0..1“ angegeben. Der Typ leitet sich aus der Data-Element Instanz ab, die einem Attribut als Datentyp zugeordnet wurde. Abgeleitete Kennzahlen werden mit einem vor ihren Namen gesetzten Schrägstrich markiert. Außerdem werden ihre Berechnungsformel und die Parameterliste aus der Computation-Metaklasse durch die Elementeeigenschaften `formula` und `parameter` wiedergegeben. Es ist also möglich, in einer Fact-Class mehrere Kennzahlen anzugeben

Von den fünf verschiedenen Verbindungen, welche die UML vorgibt, werden lediglich die Assoziation, die Komposition und die Generalisierung in der *mUML* verwendet. Durch Assoziationen stellt man die normale Beziehung zwischen zwei Klassen dar. Dabei wird an den Enden der Assoziation jeweils die Rollenbezeichnung der beteiligten Klassen und die Multiplizität angegeben. Außerdem ist es möglich, durch die Elementeeigenschaft `attribute`, zusätzliche Informationen aufzunehmen. Bei den Dimension- und Roll-Up-Beziehungen handelt es sich um eine gerichtete zyklenfreie Beziehung. Um sie von normalen Assoziationen zu unterscheiden, werden die drei Stereotypen `Dimension`, `Roll-Up` und `Shared-Roll-Up` verwendet. Multiplizitätsangaben und Rollenbezeichnungen werden mit Ausnahme bei der Verbindung `Shared-Roll-Up`, bei welcher ein `multiplicity`-Attribut für die type Seite existiert, nicht in das *mUML* Schema aufgenommen.

¹¹³ Für eine ausführliche Beschreibung Vgl.: Harren, A. (1999): Konzeptuelles Data Warehouse- Design.

¹¹⁴ Vgl.: Harren, A. (1999), S. 117.

Da Shared-Roll-Up Verbindungen Informationen über die anteilige Zuordnung der beteiligten Objekte besitzen sollten, verfügen sie über die Elementeigenschaft `portion`, deren Wert den Namen der Data-Element Instanz, in welcher sich diese Informationen befinden, enthält.

Die Kompositionsbeziehung, die nur zwischen Klassen mit dem gleichen Stereotyp erlaubt ist, wird durch eine Linie zwischen dem besitzenden und dem abhängigen Element dargestellt, wobei auf der Seite des besitzenden Objekts eine schwarz gefüllte Raute eingezeichnet ist. Dahingegen befinden sich auf der Seite der abhängigen Klasse eine Rollenbezeichnung und eine Multiplizitätsangabe. Ebenso wie die Komposition sind Generalisierungskonstrukte nur zwischen Klassen des gleichen Stereotyps erlaubt. Die Freigabemöglichkeit für Operatoren wird in einer Elementeigenschaft angezeigt, deren Wert eine Liste von Attribut-Operator-Zuordnungen enthält. Dadurch wird festgelegt, für welches Attribut welcher Operator erlaubt ist.¹¹⁵

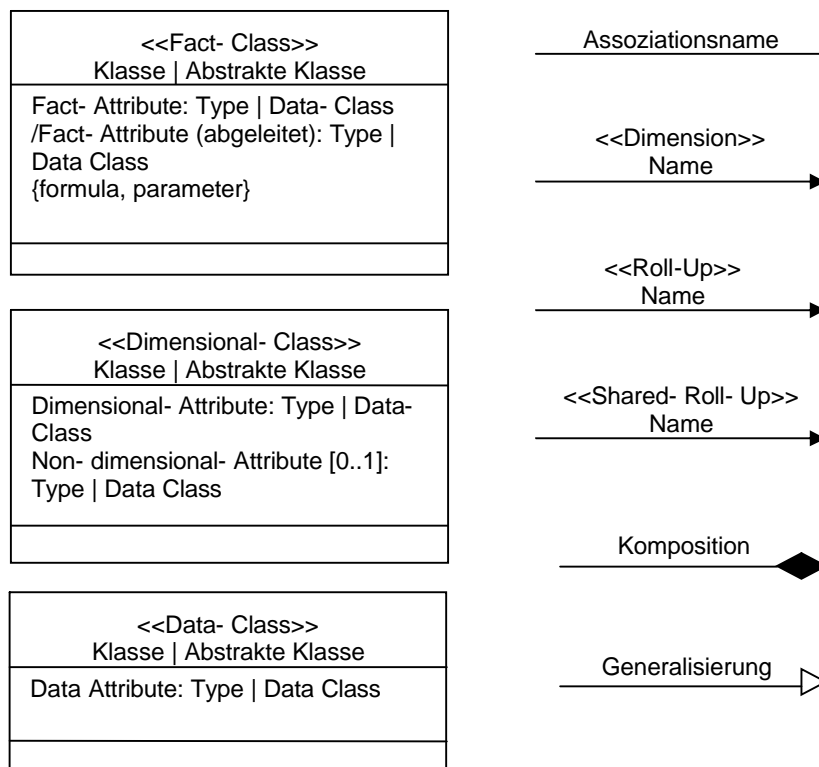
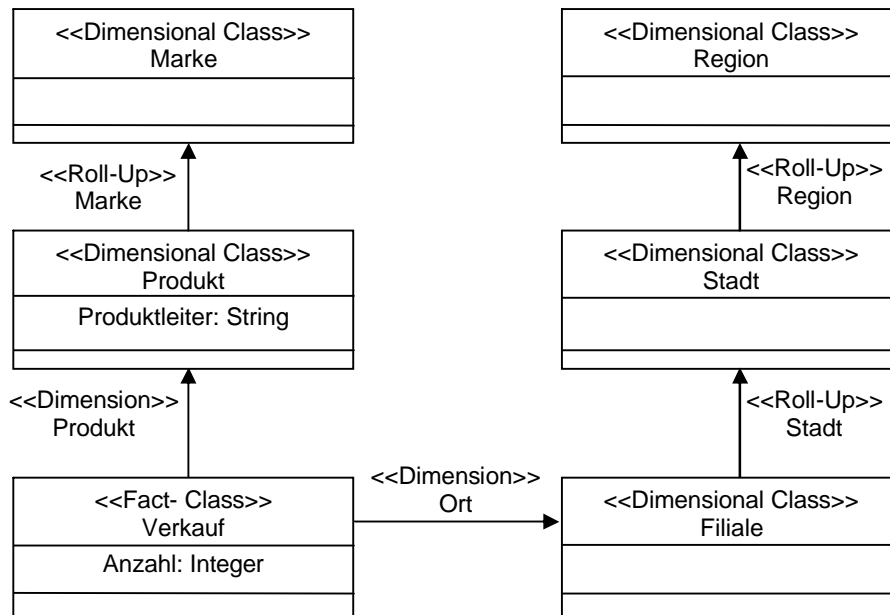


Abbildung 4-11 Grafische Notationselemente der mUML.¹¹⁶

¹¹⁵ Vgl.: Harren, A. (1999), S. 109-115.

¹¹⁶ In Anlehnung an: Böhnlein, M (2001), S. 226.

Abbildung 4-12: mUML Modell.¹¹⁷

4.8 Kubenstrukturmodell

Das Kubenstruktursystem, welches von J. Schelp u.a. mit dem Ziel entwickelt wurde, die unterschiedlichen Dimensionsstrukturen und ihre Kennzahlen, bzw. Kennzahlensysteme vollständig abzubilden, unterscheidet sich von den bisher vorgestellten Modellen dadurch, dass kein zentrales Faktenelement, um welches die zugehörigen Dimensionen angeordnet werden, existiert. Die betriebswirtschaftlichen Variablen, bzw. die Kennzahlensysteme werden in einer eigenen Dimension dargestellt. Die n Dimensionen werden kreisförmig um einen zentralen Ring angeordnet, welcher in n Segmente eingeteilt wird, in welchen die Dimensionsbeschreibungen angegeben werden. Die typisierten Dimensionselemente, welche die höchste Verdichtungsstufe in der Dimensionshierarchie darstellen, werden mit dem ihnen zugehörigen Segment verbunden. In der atomaren Sicht können die Dimensionselemente auch vollständig angegeben werden, wobei dies für die Kennzahlendimension auch in der verdichteten Sicht gilt. Darüber hinaus ist es möglich Attribute zu den jeweiligen Dimensionselementen anzugeben. Dadurch können z.B. in der Kennzahlensicht die Berechnungsvorschriften oder der Gültigkeitszeitraum in das

¹¹⁷ In Anlehnung an: Hettler, D. / Preuss, P. / Niedereichholz (2003), S. 105.

Modell aufgenommen werden. Diese Attribute werden neben ihren Elementen eingetragen. Neben dem Festhalten des Gültigkeitszeitraums empfiehlt Schelp auch eine Beschreibung der einzelnen Schemabestandteile hinsichtlich Quellen, begrifflicher Abgrenzungen und Verwendungsmöglichkeiten. Zusätzlich sollte der organisatorisch für das entsprechende Element Verantwortliche in die Attribute aufgenommen werden, um Inkonsistenzen zu vermeiden. Außerdem können die Kardinalitäten auf den Kanten zwischen den Elementen angegeben werden.¹¹⁸

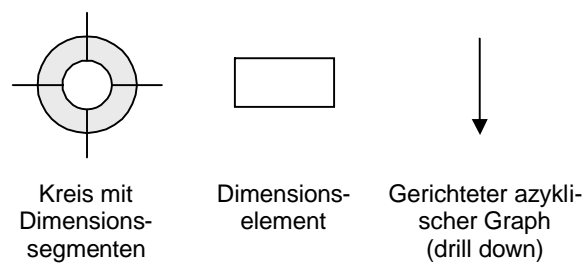
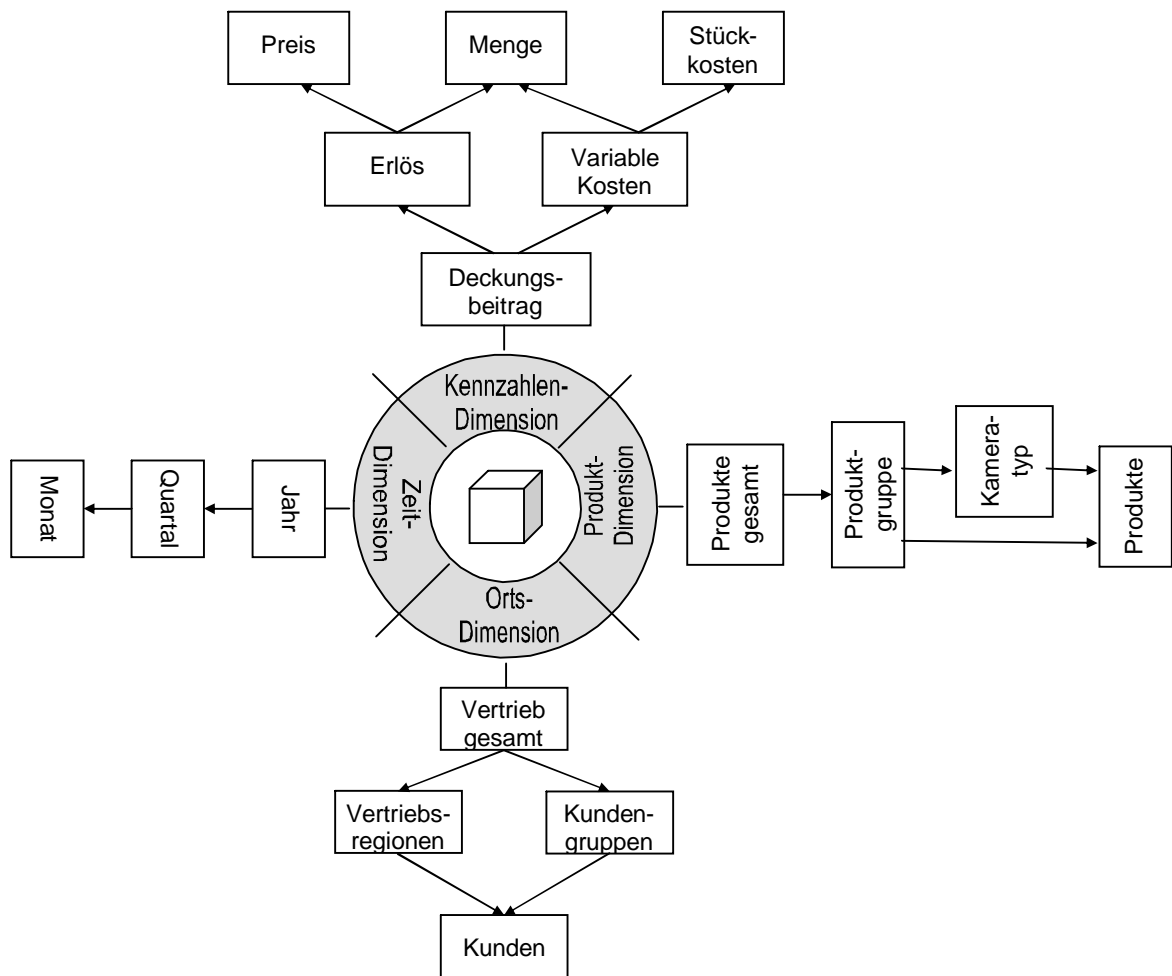


Abbildung 4-13: Grafische Notationselemente des Kubenstrukturmodells.¹¹⁹

¹¹⁸ Vgl.: Schelp, J. (2000), S. 238, 251-256.

¹¹⁹ In Anlehnung an: Schelp, J. (2000), S. 252.

Abbildung 4-14: Das Kubenstrukturmodell.¹²⁰

4.9 Semantisches Data Warehouse-Modell

Das Semantische Data Warehouse-Modell (SDWM), welches Michael Böhnlein 2001 in seiner Dissertation vorgestellt hat, wurde speziell als konzeptuelles Datenmodell für die Modellierung multidimensionaler Datenstrukturen im Bereich Data Warehouse entwickelt. Ein großer Unterschied zu den bisher vorgestellten Modellen ist, dass das SDWM die Bildung von Sichten auf Teile des Gesamtmodells ermöglicht, um somit das Verständnis komplexer Datenmodelle zu erleichtern. Es liegt ein Meta-Modell vor, an Hand dessen die Konsistenz und Vollständigkeit des Datenmodells überprüft werden

¹²⁰ Quelle: Schelp, J. (2000), S. 252.

kann. Darüber hinaus wird eine explizite Trennung zwischen Struktur und Inhalt vorgenommen, so dass intensionale und extensionale Betrachtungen von Dimensionen möglich sind. Auch die Kennzahlen und Dimensionen werden explizit unterschieden.

Böhnlein unterscheidet zwischen fünf verschiedenen Sichten: der integrierten Gesamtansicht, der Dimensionssicht, der Basiskennzahlensicht, der Kennzahlensystemsicht und der Datenstruktursicht.¹²¹

- **Dimensionssicht**

Die Objekte der intensionalen Dimensionssicht sind Dimensionen, Dimensionshierarchiestufen, Aggregationsbeziehungen und Dimensionsschnittstellen. Zu jeder Dimension gehören mindestens zwei Hierarchiestufen, die über eine Aggregationsbeziehung verbunden sind, welche die niedrigere Stufe zu einer höheren verdichtet. Die höhere Stufe steht mit einer niedrigeren in einer 1:n Beziehung, die durch die Angabe von min- und max-Kardinalitäten genauer beschrieben werden kann. Außerdem können für jede Hierarchiestufe nicht-dimensionale Attribute¹²² angegeben werden. Die Verbindungen der Dimensionen zu einer Basiskennzahl werden durch die Dimensionsschnittstelle dargestellt. Wenn eine Dimension mehrere Schnittstellen hat, können diese durch Rollennamen voneinander unterschieden werden.

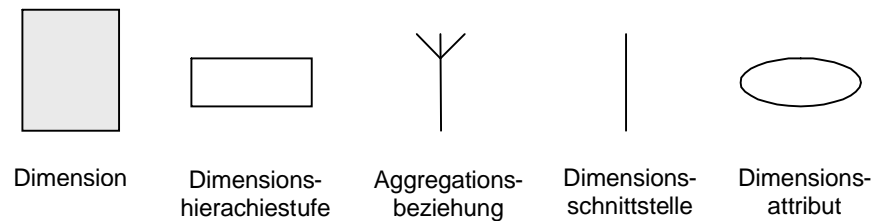
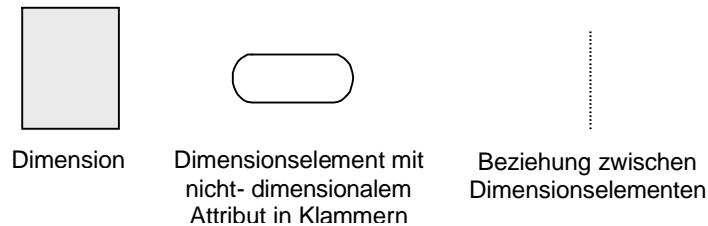
In der intensionalen Sicht werden die Dimensionen als grau unterlegtes Rechteck dargestellt, welches die anderen Objekte umgibt. Die Hierarchiestufen werden durch weiße Rechtecke, die Aggregationsbeziehungen über einen Krähenfuß, Attribute durch Ellipsen und die Schnittstellen durch eine senkrechte Linie symbolisiert.

In der Instanzsicht werden Dimensionen auf die gleiche Weise dargestellt wie in der Typisierung. Die anderen Modellierungselemente werden nicht verwendet. Stattdessen wird ein Symbol für Dimensionselement, ein abgerundetes Rechteck, und ein neues Symbol für Beziehung, eine unterbrochene Linie, eingeführt. Außerdem ist es möglich, die Ausprägungen der beschreibenden Attribute in Klammern innerhalb der jeweiligen Dimensionselemente anzugeben.¹²³

¹²¹ Vgl.: Böhnlein, M. (2001), S. 257-262.

¹²² Böhnlein spricht von einem dimensionalen Attribut. Aufgrund der Begriffsbezeichnungen, die in dieser Arbeit verwendet werden, handelt es sich jedoch um ein nicht-dimensionales Attribut.

¹²³ Vgl.: Böhnlein, M. (2001), S. 264-272.

Abbildung 4-15: Grafische Notationselemente der intensionalen Sicht.¹²⁴Abbildung 4-16: Grafische Notationselemente der extensionalen Sicht.¹²⁵

- **Basiskennzahlensicht**

Für die Basiskennzahlensicht werden nur Basiskennzahlen, bzw. Grundkennzahlen verwendet und keine abgeleiteten Kennzahlen, wie z.B. Verhältniszahlen oder Kennzahlen, die mittels einer mathematische Operation durch Basiskennzahlen gebildet werden. Die Bausteine dieser Perspektive sind Dimensionen, ihre Schnittstellen und die Basiskennzahlen selbst. Die Symbole für Dimensionen und Schnittstellen entsprechen ihren Symbolen aus der Struktursicht, während für die Kennzahl das gleiche Symbol wie für ein Dimensionselement eingesetzt wird. Da eine Kennzahl nicht entlang aller Dimensionen uneingeschränkt verdichtet werden kann, wird ihre Semiaggregierbarkeit durch eine Kantenbeschriftung des Dimensionsschnittstellensymbols angezeigt.¹²⁶

¹²⁴ Quelle: Böhnlein, M. (2001), S. 268.

¹²⁵ Quelle: Böhnlein, M. (2001), S. 268.

¹²⁶ Vgl.: Böhnlein, M. (2001), S. 273-276.

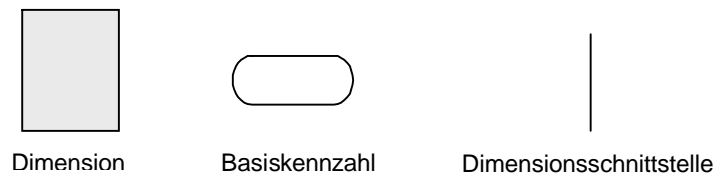


Abbildung 4-17: Grafische Notationselemente der Basiskennzahlensicht.¹²⁷

- **Sicht auf das Kennzahlensystem**

Bei der Sicht auf das Kennzahlensystem werden die Kennzahlen über die Grenzen ihrer Würfel hinweg in einem Kennzahlensystem betrachtet, da ein Kennzahlensystem eine größere Aussagekraft besitzt als einzelne Kennzahlen und somit der betrachtete Gegenstand vollständiger erfasst werden kann. In diesem System werden Basiskennzahlen und abgeleitete Kennzahlen miteinander verbunden. Die häufigsten abgeleiteten Kennzahlen sind Verhältniskennzahlen, die den drei verschiedenen Typen Gliederungszahlen, Beziehungszahlen und Indexzahlen zugeordnet werden.

Durch eine Gliederungszahl wird eine Teilgröße zu einer Gesamtgröße ins Verhältnis gesetzt. Bei einer Beziehungszahl werden inhaltlich gleichwertige, aber statistisch ungleichartige Maße in einen sinnvollen Zusammenhang gesetzt. So ergibt sich zum Beispiel die Exmatrikulationsquote einer Hochschule aus den ungleichartigen statistischen Massen ‚Anzahl der eingeschriebenen Studierenden‘ und ‚Anzahl der exmatrikulierten Studierenden‘. Durch Indexzahlen wird das Verhältnis gleichartiger, aber zeitlich oder örtlich verschiedener Zahlen angegeben.

Dargestellt werden die Basiskennzahlen durch das gleiche Symbol wie in der Basiskennzahlensicht. Das formgleiche Beschreibungselement wird auch für Kennzahlen verwendet, allerdings ist der Rahmen mehrfach unterbrochen. Die Kennzahlenbeziehung wird durch einen gerichteten Pfeil, der von der Basiskennzahl ausgeht, beschrieben.¹²⁸

¹²⁷ Quelle: Böhnlein, M. (2001), S. 275.

¹²⁸ Vgl.: Böhnlein, M. (2001), S. 276-280.

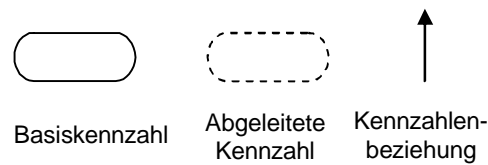


Abbildung 4-18: Grafische Notationselemente des Kennzahlensystems.¹²⁹

Um eine genauere Beschreibung der Kennzahlen zu ermöglichen sollten diese zusätzlich mit folgenden Eigenschaften in einer Tabelle dargestellt werden.

- **Lange Bezeichnung:** Eine ausführliche Bezeichnung, die mit der Bezeichnung im Diagramm übereinstimmt.
- **Kurzbezeichnung:** Eine kurze Beschreibung der Kennzahl, um diese eindeutig zu identifizieren
- **Beschreibung:** Eine ausführliche Beschreibung der inhaltlichen Semantik einer Kennzahl
- **Art der Kennzahl:** Angabe, ob es sich um eine Basis- oder abgeleitete Kennzahl handelt. Trifft letzteres zu, sollte auch der Typ angegeben werden.
- **Definition:** In ihr werden die Berechnungsformeln der abgeleiteten Kennzahl angegeben.
- **Wertebereich:** Darin wird der Wertebereich der Kennzahl angegeben
- **Maßeinheit:** Zu jeder Kennzahl gehört eine Maßeinheit
- **Skala:** Dadurch wird bestimmt, ob einer Kennzahl eine Nominal-, Ordinal-, Intervall- oder Verhältnisskala zu Grunde liegt.
- **Aggregierbarkeit:** Darin wird angegeben, ob eine Kennzahl vollständig oder semi-aggregierbar ist.
- **Kennzahlentyp:** Der Kennzahlentyp gibt an, ob es sich um eine Bestands- oder Bewegungskennzahl handelt.
- **Aggregationsfunktion:** Damit wird festgelegt, ob es sich um eine distributive, algebraische oder holistische Funktion handelt.¹³⁰

¹²⁹ Quelle: Böhnlein, M. (2001), S. 279.

¹³⁰ Vgl.: Böhnlein, M. (2001), S. 280-282.

- **Datenstruktursicht**

Gegenstand der Datenstruktursicht ist die multidimensionale Struktur also der eigentliche Hyperwürfel des Data Warehouse. Die Elemente sind Basis- und abgeleitete Kennzahlen, Dimensionen, deren Schnittstellen zu den Basiskennzahlen, und die Kennzahlenbeziehungen. Da die abgeleiteten Kennzahlen nur indirekt über ihre Basiskennzahlen mit den Dimensionen verbunden sind, müssen sich alle abgeleiteten Kennzahlen auf die gleichen Dimensionen beziehen und können auch nur anhand dieser analysiert werden. Aufgrund der Abhängigkeiten zwischen abgeleiteter Kennzahl und Basiskennzahl gilt auch, dass die Aggregierbarkeit der Basiskennzahlen die Aggregierbarkeit der abgeleiteten Kennzahl bestimmt. Ist z.B. eine Basiskennzahl entlang einer Dimension nur eingeschränkt verdichtbar, so gilt dies auch für die abgeleiteten Kennzahlen. Die Symbole der Datenstruktursicht entsprechen den zuvor beschriebenen.¹³¹

- **Gesamtsicht**

In der Gesamtsicht werden alle Sichten zusammengeführt. Sie ermöglicht es den Modellierern, die verschiedenen Hyperwürfel im Zusammenhang zu betrachten. Dabei können zwei Hyperwürfel zueinander in Beziehung gesetzt werden, wenn sie mindestens eine Dimension gemeinsam haben. Das grafische Notationssymbol für eine solche Dimension ist ein längliches Rechteck mit unterbrochenem Rahmen.¹³²

¹³¹ Vgl.: Böhnlein, M. (2001), S. 282-284.

¹³² Vgl.: Böhnlein, M. (2001), S. 287-291.

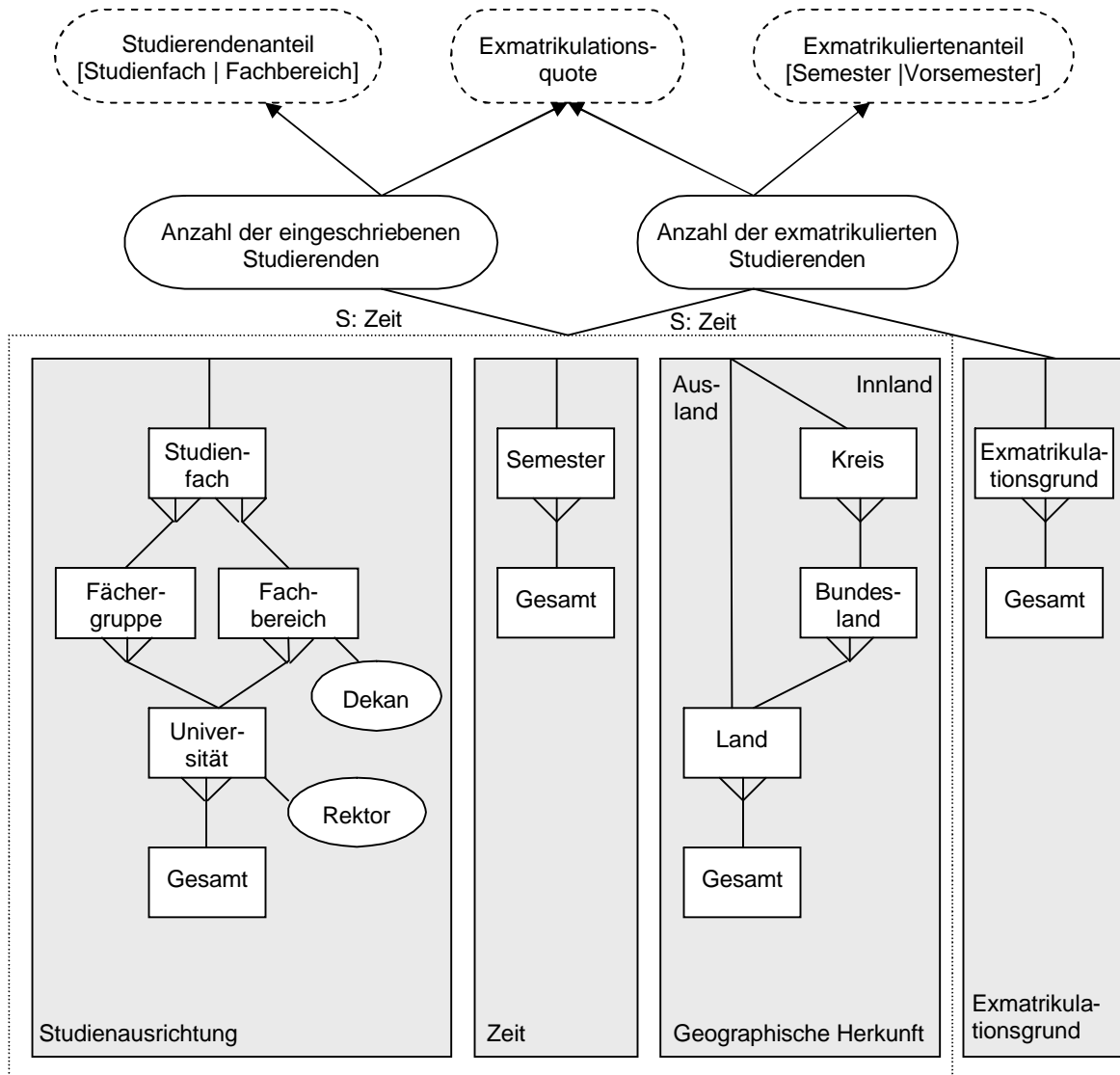


Abbildung 4-19: Integrierte Gesamtansicht des SDWM.¹³³

¹³³ Quelle.: Böhnlein, M. (2001), S. 290.

4.10 Vergleich der Modellierungsansätze

		MERM	ADAPT	DFM	MD	MAC	mUML	Kubenstruktur	SDWM
Allgemeine Anforderungen	Vorgehensmodell	-	-	+	+	-	+	+	+
	Metamodell	+	-	- ¹³⁴	-	-	+	-	+
	Unterstützung der Komplexitätsbeherrschung durch Sichtenbildung	-	-	-	-	-	-	-	+
	Trennung intensionale und extensionale Struktur	+	- ¹³⁵	+	+	+	+	+	+
	Trennung qualitative und quantitative Daten	+	+	+	+	+	+	- ¹³⁶	+
	Überführung in logisches Schema	+	-	+	-	-	+	+	+
	Werkzeugunterstützung	+	+ ¹³⁷	-	-	-	+	-	+
Anforderungen Fakten	Mehrere Fakten im Schema	+	+	-	-	+	+	-	+
	Mehrere Kennzahlen in einem Fakt	+	+	+	+	+	+	+	+
	Beziehungen zwischen Fakten	-	-	-	-	-	+	-	+
	Explizite Angabe abgeleiteter Attribute und ihrer Berechnungsvorschrift	-	+	-	-	-	+	+	+ ¹³⁸
	Angabe der Additivität	-	-	+	-	-	+	-	+
Anforderungen Dimensionen	Hierarchieebene	+	+	+	+	+	+	+	+
	Parallele Hierarchie	+	+	+	+	+	+	+	+
	unbalancierte Hierarchie	-	-	-	-	+	-	+	+
	Alternative Verdichtungs-pfade	+	+	-	+	+	+	+	+
	Heterarchien	-	-	-	-	-	+	+	-
	nicht-dimensionale Attribute		+	+	+	+	+	+	+
	Sonstige Beziehungen	-	-	-	-	-	+	-	-

-: nicht unterstützt/vorhanden

+: unterstützt/vorhanden

Tabelle 4-2: Vergleich der Modellierungsansätze.¹³⁹

¹³⁴ Anmerkung: Die meisten Modelle verfügen nicht über ein grafisches Metamodell, sondern nur über ein formales Modell.

¹³⁵ ADAPT mischt die beiden Sichten, da bspw. einzelne Dimension Members angegeben werden können.

¹³⁶ Die Kennzahlen werden ebenfalls in einer Dimension dargestellt, wodurch die Unterscheidung zwischen Kennzahl und Merkmal erschwert wird.

¹³⁷ Die Notationselemente existieren als VISIO Template.

¹³⁸ Abgeleitete Attribute werden im Schema gekennzeichnet. Ihre Berechnungsvorschrift wird jedoch erst im Datenblatt angegeben.

5 Vorgehensmodelle zur Entwicklung von Data Warehouse Schemata

Bei der Entwicklung von Data Warehouse Schemata gibt es verschiedene Herangehensweisen, um die Daten für das konzeptionelle Modell zu gewinnen. Im folgenden Kapitel werden die Analyse der Modelle operativer Quellsysteme, die Analyse der Geschäftsprozessmodelle und die Informationsbedarfsanalyse vorgestellt.

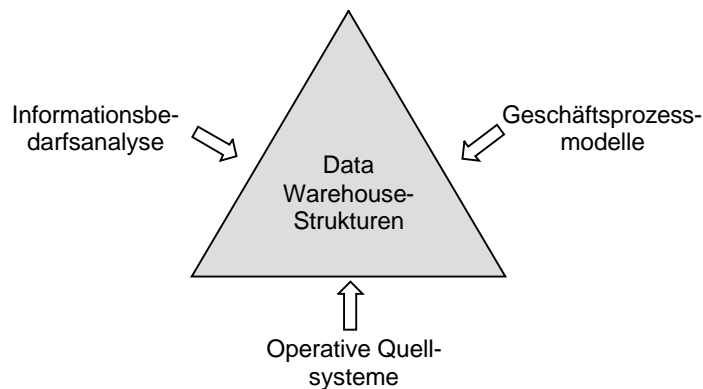


Abbildung 5-1: Vorgehensweisen bei der Entwicklung von Data Warehouse Strukturen.¹⁴⁰

5.1 Analyse der operativen Systeme

Ziel der Analyse der operativen Quellsysteme ist es, festzustellen, welche Daten für die Entscheidungsunterstützung bereits verfügbar sind. Stellvertretend für verschiedene Vorgehensmodelle aus der Literatur wird hier das Vorgehensmodell nach Golfarelli, Maio und Rizzi beschrieben.

¹³⁹ In Anlehnung an: Böhnlein, M. (2001), S. 213 und Herden, O.(2001), S. 38-39.

¹⁴⁰ Quelle: Böhnlein, M. (2001), S. 307.



Abbildung 5-2: Entwicklung des Data Warehouse Schemas aus den operativen Modellen.¹⁴¹

Neben dem grafischen und formalen Modell haben Golfarelli et. al. auch ein Vorgehensmodell zur Erstellung des DFM aus den ER-Modellen der operativen Systeme entwickelt. Dazu werden in einem ersten Schritt die Fakten identifiziert, welche aus geeigneten Entity- oder Beziehungstypen gebildet werden. Anschließend sind für jeden Fakt folgende Schritte zu durchlaufen:

1. Erstellen der Attributbäume

Jeder Baum muss „[...] mit einem der Attribute des Fakt in Beziehung stehen und alle von der Wurzel abhängigen Attribute müssen vom betreffenden Fakt funktional abhängig sein.“¹⁴²

2. Beschneiden der Attributbäume

Da nicht alle überführten Attribute für eine Analyse relevant sind, wird der Attributbaum beschnitten (pruning) und aufgepfropft (grafting). Die nicht relevanten Zweige werden abgeschnitten. Allerdings ist es möglich, dass nicht alle Hierarchieebenen des abgeschnittenen Zweigs uninteressant sind. In diesem Fall werden die interessanten Ebenen auf den noch bestehenden Teil des Baumes aufgepfropft.¹⁴³

3. Bestimmen der Dimensionen

In den bisher erstellten Attributbäumen müssen nun die Dimensionen bestimmt und die Aggregationspfade festgelegt werden.¹⁴⁴ Dadurch wird auch die Granularität der Fakten bestimmt, da diese durch die Wahl der Dimensionen festgelegt wird.¹⁴⁵

¹⁴¹ In Anlehnung an: Böhnlein, M. (2001), S. 310; Determann, L. (2002), S. 139-141.

¹⁴² Schelp, J. (2000), S. 192.

¹⁴³ Vgl.: Schelp, J. (2000), S. 192.

¹⁴⁴ Vgl.: Schelp, J. (2000), S. 192.

¹⁴⁵ Vgl.: Determann, L. (2002), S. 140.

4. Definieren der Faktattribute

An dieser Stelle werden die konkreten Kennzahlen des Faktas abgeleitet. „Diese sind entweder als numerische Werte im ERM enthalten oder werden durch Zählen von Objektmengen (z.B. Anzahl Kunden) hinzugefügt.“¹⁴⁶ Auch die abgeleiteten Fakten müssen berücksichtigt werden. Die Berechnungsvorschriften für die Datenverdichtung sollten jedoch separat notiert werden.¹⁴⁷

5. Generieren der Hierarchien

Die Hierarchien werden aus den zuvor entwickelten Attributbäumen generiert und gegebenenfalls um weitere Strukturierungen ergänzt.¹⁴⁸ Dabei muss besonders auf die Kardinalitäten der Beziehungen geachtet werden, da in dieser Baumstruktur Vorgänger und Nachfolger nur über eine 1:n-Beziehung verbunden werden dürfen. Die für die Aggregation unrelevanten Attribute werden als nicht-dimensionale Attribute dargestellt.¹⁴⁹

Dieses Vorgehen bezieht sich jedoch nur auf ER-Diagramme und lässt somit weitere Informationsbedürfnisse und andere Informationsquellen des Nutzers unberücksichtigt. Außerdem wird vorausgesetzt, dass tatsächlich ein korrektes Modell der operativen Systeme existiert. Sollte kein Diagramm existieren, das den Datenbestand des Unternehmens beschreibt, ist es nicht empfehlenswert, dieses im Zuge einer DWH-Entwicklung zu erstellen. Dieses würde wahrscheinlich ansonsten, auf Grund des hohen Aufwands, der mit der Erstellung verbunden ist, ohne weitere Modifikationen als Data Warehouse Schema übernommen.¹⁵⁰

¹⁴⁶ Determann, L. (2002), S. 140.

¹⁴⁷ Vgl.: Determann, L. (2002), S. 140.

¹⁴⁸ Vgl.: Determann, L. (2002), S. 140.

¹⁴⁹ Vgl.: Schelp, J. (2000), S. 192.

¹⁵⁰ Vgl.: Wirtz, K.(2002), S. 132.

5.2 Analyse der Geschäftsprozesse

Böhnlein hat ein System entwickelt, in welchem die Datenstrukturen mittels der Semantischen Objektorientierten Methode (SOM) aus den Geschäftsprozessen des Unternehmens in sechs Schritten abgeleitet werden können.

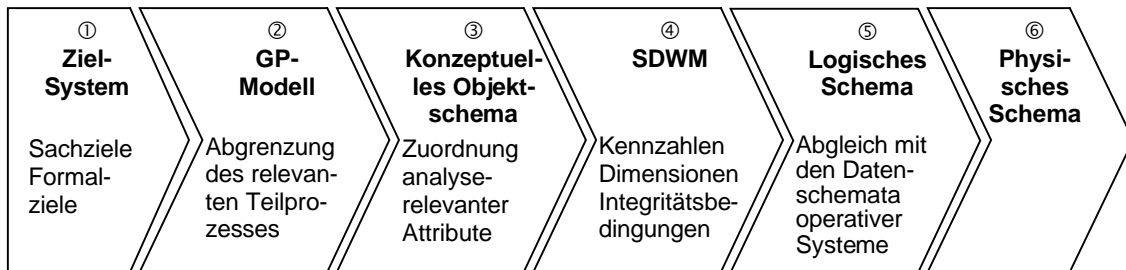


Abbildung 5-3: Entwicklung des Data Warehouse Schemas aus den Geschäftsprozessen.¹⁵¹

Dabei dienen die Schritte 1-3 der Identifikation von Data Warehouse relevanten Daten, während die Schritte 5-6 den Modellierungsebenen entsprechen.¹⁵²

5.2.1 Bestimmung des Zielsystems

Im ersten Schritt werden die Ziele des Unternehmens festgelegt. Dabei unterscheidet man zwischen Sach- und Formalziel. Sachziele geben Art und Zweck der Leistungserstellung an und bestimmen somit, was erstellt werden soll. Formalziele hingegen legen Art und Umfang der Sachzielerreichung fest und damit, wie eine Leistung erstellt wird. Bei der Zielfestlegung ist vor allem darauf zu achten, dass diese quantifizierbar und somit operationalisierbar sind. Gleichzeitig sollten die Ziele konsistent und kompatibel sein. Nachdem die Zielfindung abgeschlossen ist, liegt ein hierarchisches Zielsystem der Sachziele aus originären und abgeleiteten Zielen vor, wobei die Ziele der unteren Planungs- und Entscheidungsebenen zu den oberen in einem Zweck-Mittelverhältnis stehen, wodurch die Ausrichtung aller Unterziele auf das Oberziel gewährleistet wird. Die Formalziele werden soweit möglich den jeweiligen Sachzielen zugeordnet. Ein Ziel gilt dann als ausreichend operationalisiert, wenn es möglich ist, den Istwert mit dem Planwert zu vergleichen. Gemessen werden diese Werte über Kennzahlen. Zusätzlich

¹⁵¹ In Anlehnung an: Böhnlein, M. (2001), S. 310.

¹⁵² Vgl.: Böhnlein, M. (2001), S. 308-310.

muss für jedes Zielobjekt ein Managementobjekt, das für dessen Erreichung zuständig ist, bestimmt werden.¹⁵³

5.2.2 Analysieren des Geschäftsprozessmodells

Geschäftsprozesse stellen eine Anleitung zur Umsetzung der Ziele dar. Deshalb wird für jedes operationalisierte Sachziel der dazugehörige Teilprozess, der aus dem Prozessgefüge des Managementobjekts abgeleitet wird, herausgestellt. Durch die Präzisierung des Teilprozesses mittels Interaktionsschemata (IAS) und Vorgangs-Ereignis-Schemata (VES) wird das Verständnis für den Anwendungsbereich verbessert. Außerdem können so zu den Zielen gehörende Teilleistungen entdeckt werden. Für die Präzisierung der Prozesse gelten die Prinzipien der SOM-Methodik. Das heißt, dass die Prozesse zunächst nach dem Verhandlungsprinzip zerlegt werden, um somit die Lenkung von Transaktionen aufzuzeigen. Anschließend werden die Transaktionen, die eine zusammengesetzte Leistung weiter geben, in einzelne Teilleistungen zerlegt. Erst danach werden Objekte, die zusammengesetzte Leistungen erstellen, in Teilobjekte aufgegliedert. Anschließend wird ein zum IAS passendes VES erstellt.¹⁵⁴

5.2.3 Konzeptuelles Objektschema

Durch das Konzeptuelle Objektschema (KOS) können die Datenstrukturen eines Geschäftsprozesses beschrieben werden, aus welchen dann die Datenstrukturen des DWH abgeleitet werden können. Das KOS ist eine abgewandelte Form des Semantischen Entity Relationship Modell (SERM), die um objektorientierte Elemente erweitert wurde. Dabei handelt es sich um konzeptuelle Objekttypen (KOTs), die untereinander in Beziehung stehen, und durch Namen, Attribute und Operatoren näher beschrieben werden. Man unterscheidet zwischen drei Beziehungen. Durch die Interacts_with-Beziehung werden Interaktionskanäle zwischen Objekttypen dargestellt. Eine is_a-Beziehung beschreibt die Generalisierung oder Spezialisierung von Objekttypen. Durch die is_part_of-Beziehung werden Aggregationen von Objekttypen modelliert.

Da es sich beim KOS um eine Erweiterung des SERM handelt, ist es möglich, Existenzabhängigkeiten zwischen KOTs zu visualisieren. Dabei werden die existenzunabhängigen KOTs im Diagramm links angeordnet, während rechts die existenzabhängigen

¹⁵³ Vgl.: Böhnlein, M. (2001), S. 327-332.

¹⁵⁴ Vgl.: Böhnlein, M. (2001), S. 337-345.

abgebildet werden, so dass der Grad der Existenzabhängigkeit von links nach rechts zunimmt. Daraus ergibt sich eine hierarchische Struktur.

Das KOS wird aus den im Schritt zuvor entwickelten IAS und VES unter Beachtung einiger Regeln abgeleitet. Dabei werden die betrieblichen Objekte als objektspezifische KOTs abgebildet. Da sie existenzunabhängig sind, werden sie in der linken Spalte angeordnet. Betriebliche Transaktionen sind von ihren entsprechenden Objekten abhängig und werden als transaktionsspezifische KOTs dargestellt. Diese sind über eine `interacts_with`-Beziehung mit ihren objektspezifischen KOTs verbunden. Auch die Reihenfolgenbeziehungen im VES werden zur Abbildung der Existenzabhängigkeiten verwendet. So ergeben sich aus einer Reihenfolge von Transaktionen entsprechende Sequenzen von Existenzabhängigkeiten zwischen den transaktionsspezifischen KOTs. Zusätzlich kann das KOS verfeinert werden, indem analyserelevante Attribute zu den KOTs hinzugefügt werden, komplexe Objekt durch `is_part_of`-Beziehungen zerlegt und Generalisierungen durch `is_a`-Beziehungen ermöglicht werden. Außerdem ist es möglich, KOTs mit den gleichen Attributen zusammenzufassen und nicht-analyserelevante Objekte zu entfernen. Des Weiteren können die Kardinalitäten der Beziehungen zwischen KOTs angegeben werden.¹⁵⁵

5.2.4 Semantisches Data Warehouse Schema

Im vierten Schritt werden nun die Kennzahlen und Dimensionen identifiziert. Um geeignete Kennzahlen zu finden, muss festgestellt werden, wie die Leistung eines Geschäftsprozesses, die dieser zur Erfüllung des Sachziels erbringt, gemessen werden kann. Dafür eignen sich meist mehrere Kennzahlen, wobei zunächst die Basiskennzahlen identifiziert werden. Anschließend werden diese Kennzahlen einem Objekttyp des KOS zugeordnet, um das Aufspüren der Dimensionen vorzubereiten. Dabei handelt es sich meist um transaktionsorientierte KOTs, da jede Transaktion an der Erstellung oder Übergabe einer Leistung beteiligt ist. Von diesen KOT's sind insbesondere diejenigen Objekte, die aus Durchführungstransaktionen des Geschäftsprozessmodells abgeleitet wurden und die KOTs, welche aus Vereinbarungstransaktionen entwickelt wurden, besonders geeignet.

Um Dimensionen zu bestimmen, werden diejenigen KOTs und ihre Attribute, von denen das KOT, welchem die Kennzahl zugeordnet wurde, existenziell abhängig ist, näher

¹⁵⁵ Vgl.: Böhnlein, M. (2001), S. 345-351.

betrachtet. Diese KOTs befinden sich links im Diagramm und sind über Kanten direkt oder indirekt mit dem Kennzahlen KOT verbunden. „Um den Teil des Schemas zu ermitteln, der die Eigenschaft der Existenzvoraussetzung für einen gegebenen KOT erfüllt, ist die sog. Hülle der Existenzvoraussetzungen C_{EX} zu bestimmen.“¹⁵⁶ Durch diesen Algorithmus werden ausgehend vom Kennzahlen KOT, die Kanten schrittweise nach links durchlaufen und die verbundenen KOTs in C_{EX} aufgenommen. Durch eine Repeat-Until-Schleife wird angegeben, wenn keine weiteren KOTs mehr hinzufügbare sind. Abschließend wird der Start-Objekttyp aus dem Ergebnis entfernt, „[...] da er keine Existenzvoraussetzung für sich selbst darstellt.“¹⁵⁷

Durch diesen Algorithmus kann der Suchraum für Domänen eingeschränkt werden. Das Erkennen und Strukturieren der Dimensionen ist jedoch die Aufgabe von Experten. Einige Regeln erleichtern dies. So sind (0,*)- und (1,*)-Beziehungen zwischen zwei und mehr aufeinander folgenden Objekttypen ein Hinweis auf hierarchische Strukturen innerhalb einer Dimension, wobei (0,*)-Beziehungen auf unbalancierte Hierarchien hindeuten. Wenn mehrere Kanten auf einen Objekttyp zulaufen, könnte es sich um eine parallele Hierarchie handeln.

Auch bei den analyserelevanten Attributen, die den KOTs in Schritt drei zugeordnet wurden, kann es sich um Dimensionen handeln. Deshalb werden die nicht normalisierten Werte weiter zerlegt, um als KOS-Teilausschnitt in das Schema eingefügt zu werden, so dass die Regeln auch auf sie anwendbar sind. Zusätzlich kann das Formalziel zur Identifizierung relevanter Dimensionen herangezogen werden.

Abschließend müssen die Kennzahlen ihren entsprechenden Dimensionen zugeordnet und die Integritätsbedingungen definiert werden. Dabei muss überprüft werden, ob die Kennzahlen bezüglich aller ihrer Dimensionen aggregierbar sind. Wenn dies nicht zutrifft, ist es erforderlich, diesen Tatbestand im Schema zu vermerken. Auch sollten gemeinsame Dimensionen identifiziert werden.¹⁵⁸

¹⁵⁶ Böhnlein, M. (2001), S. 354.

¹⁵⁷ Vgl.: Böhnlein, M. (2001), S. 354.

¹⁵⁸ Vgl.: Böhnlein, M. (2001), S. 351-363.

5.2.5 Logisches Data Warehouse Schema

In diesem Schritt wird das konzeptionelle Schema in drei Schritten in ein logisches Schema überführt. Zunächst wird das SDWM anhand einiger Ableitungsregeln in das Star/Galaxy oder in das Snowflake Schema transformiert.

SDWM	Star/Galaxy Schema
Dimensionsattribut	Attribut einer Dimensionstabelle
Dimensionshierarchiestufe	Attribute (ID und Bezeichnung) einer Dimensionstabelle
Dimension	Dimensionstabelle
Dimensionsschnittstelle	Tabellenbezeichnung
Basiskennzahl	Fakttabelle und Attribute einer Fakt-tabelle
Abgeleitet Kennzahl	Attribute einer Fakttabelle

Tabelle 5-1: Abbildungsregeln zwischen SDWM und Star/ Galaxy Schema.¹⁵⁹

Im Anschluss erfolgt die Überführung in das Relationenmodell, um den physischen Entwurf vorzubereiten.

Star/Galaxy Schema	Relationenmodell
Attribut einer Dimensionstabelle	Attribut
Attribut einer Fakttabelle	Attribut
Dimensionstabelle	Relationstyp
Fakttabelle	Relationstyp

Tabelle 5-2: Abbildungsregeln zwischen Star/ Galaxy Schema und Relationenmodell.¹⁶⁰

Bevor jedoch das physische Schema entwickelt wird, müssen zunächst in einem Soll-Ist-Vergleich die Daten der operativen Systeme mit den Daten der mehrdimensionalen Strukturen abgeglichen werden, um Informationslücken aufzudecken, die eine Umsetzung des Sollkonzepts verhindern könnten. Dafür kann bspw. das Vorgehen von Golfarelli et. al. zur Identifikation von Data-Warehouse-Strukturen aus den operativen Systemen verwendet werden.¹⁶¹

¹⁵⁹ Quelle: Böhnlein, M (2001), S. 366.

¹⁶⁰ Quelle: Böhnlein, M (2001), S. 371.

¹⁶¹ Vgl.: Böhnlein, M (2001), S 363-371.

5.2.6 Physisches Data Warehouse Schema

Zum Abschluss wird das logische mittels eines physischen Schemas auf ein konkretes Datenbanksystem übertragen. Dabei wird zur Erzeugung des physischen Schemas bspw. SQL als Datendefinitionssprache verwendet, um Anweisungen zu generieren. So wird für jeden Relationstyp auf der logischen Ebene mittels des Befehls CREATE TABLE eine eigene Tabelle im Zieldatenbanksystem generiert.

Das Hauptaugenmerk des physischen Entwurfs liegt auf der Performanceoptimierung. Diese wird über Indexstrukturen, Fragmentierungsstrategien und Pre-Aggregationen gewährleistet.¹⁶²

Bei der Anwendung dieser Methode könnte das auf diesem Modell basierende Data Warehouse zur Unterstützung von Geschäftsprozessen genutzt werden. Für ein rein analytisches System wäre der Ansatz jedoch weniger geeignet, „[...]da der einem analytischen Informationssystem zugrunde liegende Prozess nicht operational strukturierbar ist.“¹⁶³

5.3 Informationsbedarfsanalyse

Unter einer Informationsbedarfsanalyse wird in der Literatur meist ein nachfrageorientierter Ansatz verstanden, dessen Ziel es ist, den Informationsbedarf der zukünftigen Nutzer eines Systems zu ermitteln. Dafür werden verschiedene Methoden wie Interviews¹⁶⁴, oder auch die Ermittlung kritischer Erfolgsfaktoren empfohlen.¹⁶⁵

Problematisch bei diesen Ansätzen ist allerdings, dass es für die Anwender schwierig ist, ihren objektiven und auch ihren noch nicht gedeckten Informationsbedarf zu formulieren. Zumal die Anwender sich nicht bewusst sind, welche neuartigen Informationen durch ein Data Warehouse bereitgestellt werden können.

¹⁶² Vgl.: Böhnlein, M (2001), S. 373-375.

¹⁶³ Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004); S. 10.

¹⁶⁴ Vgl.: Poe, V. / Reeves, L. (1997), S. 118-128.

¹⁶⁵ Vgl.: Holthuis, J. (1999), S. 25.

An der Universität St. Gallen wurde in Zusammenarbeit mit mehreren Kooperationspartnern ein vierphasiges Vorgehensmodell zur Informationsbedarfsanalyse entwickelt, welches auf in der Praxis gesammelten Erfahrungen basiert.¹⁶⁶

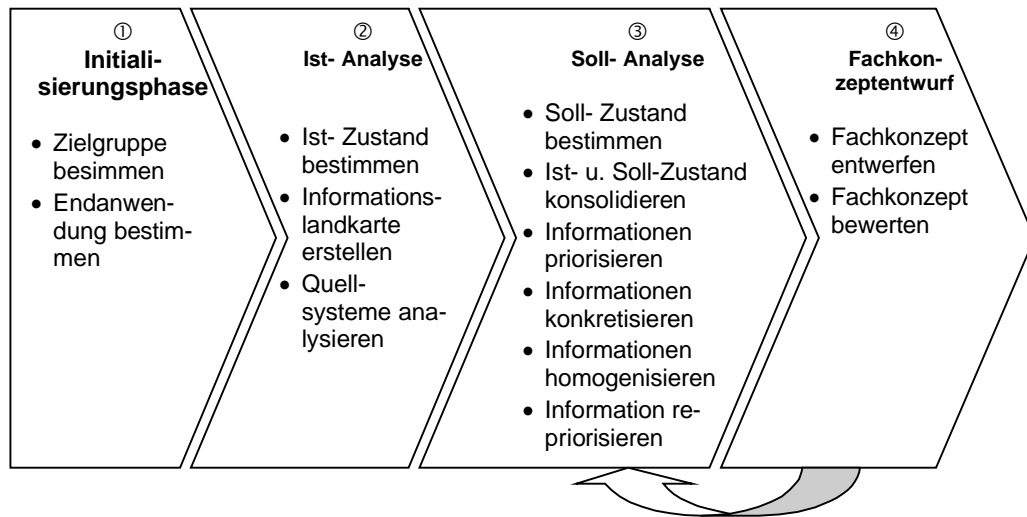


Abbildung 5-4: Entwicklung des Data Warehouse Schemas aus der Informationsbedarfsanalyse.¹⁶⁷

5.3.1 Initialisierungsphase

In dieser Phase wird zunächst die Zielgruppe ermittelt und anschließend die Endanwendung bestimmt, da die Benutzer und der Nutzungszweck den Typ des Systems beeinflussen.¹⁶⁸

5.3.2 Ist-Analyse

Um den Ist-Zustand der Informationsversorgung im Unternehmen zu ermitteln, empfehlen die Autoren B. Strauch und R. Winter eine Inventur des Berichtswesens, da dadurch sowohl eine Informationsüberflutung der zukünftigen Nutzer vermieden wird, als auch eine bereits vorhandene Sammlung von Metadaten für die Entwicklung des DWH ge-

¹⁶⁶ Vgl.: Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004), S. 10-11.

¹⁶⁷ In Anlehnung an: Böhnlein, M. (2001), S. 310; Strauch, B. (2002): S. 172.

¹⁶⁸ Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004), S. 12-13.

nutzt werden kann.¹⁶⁹ Dafür erstellen die Verantwortlichen der operativen Berichtssysteme zunächst eine Liste der von ihnen produzierten Berichte, in welcher eine Anzahl von Metadaten zu den jeweiligen Berichten angegeben wird, wie z.B. Berichtspositionen, Produzenten, etc. Wenn möglich werden diese Metadaten in einer Datenbank abgelegt, um somit später automatisch eine Informationslandkarte generieren zu können. Anschließend wird eine Liste aller Berichtspositionen erstellt, um potentielle Synonyme zu identifizieren. Der Abgleich der einzelnen Positionen kann jedoch nur durch Mitarbeiter der Fachbereiche vorgenommen werden. Gleichzeitig werden Begriffspositionen mit ähnlicher Bedeutung zu Clustern zusammengefasst, um einen Überblick über die Informationsversorgung zu geben. Das Ergebnis der Analyse sollte in einer Informationslandkarte abgebildet werden, die den Ist-Zustand der Informationen aus unterschiedlichen Blickwinkeln darstellt. So kann z.B. dargestellt werden, welche Organisationseinheit welche Information nutzt.¹⁷⁰ Anschließend werden die ermittelten Datenquellen hinsichtlich ihrer Qualität analysiert, da in der Praxis mangelnde Datenqualität häufig eine sinnvolle Umsetzung von Teilen des Data Warehouse verhindert.¹⁷¹

5.3.3 Soll-Analyse

Der Soll-Zustand wird mittels Business Questions erhoben. Dabei handelt es sich um typische Fragestellungen, die mit der Tätigkeit der Benutzer zusammenhängen und durch die bisher vorhandenen Systeme nicht oder nur unzureichend beantwortet werden können.¹⁷² Aus den Fragen werden dann die Kennzahlen und ihre Dimensionen mit ihren Hierarchiestufen abgeleitet und konsolidiert.¹⁷³ Anschließend wird geprüft, welche Teilbestände des Ist-Zustands in das DWH aufgenommen werden und mit dem Sollzustand abgeglichen. Darüber hinaus müssen die Daten priorisiert und konkretisiert werden. Um die Komplexität von Projekten einzuschränken, werden die Informationen schrittweise durch das DWH bereitgestellt. Mögliche Kriterien für die Priorisierung wären z.B. die Unterscheidung in Informationen, die enthalten sein müssen und solche,

¹⁶⁹ Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004), S. 13.

¹⁷⁰ Vgl.: Strauch, B. (2002): S. 180-186.

¹⁷¹ Vgl.: Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004), S. 14-15.

¹⁷² Vgl.: Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004), S. 15.

¹⁷³ Vgl.: Strauch, B. (2002): S. 187-190.

die bereitgestellt werden sollten, Datensicherheitsaspekte, Kosten und Zeit. Bei der Konkretisierung werden die Berechnungsvorschriften mit der Unterstützung der betroffenen Organisationseinheiten und die dafür benötigten weiteren Attribute aus den Datenquellen bestimmt. Auch muss die Granularität und die Periodizität der Daten festgehalten werden. Zusätzlich müssen in der Phase der Soll-Analyse die verschiedenen Begriffe vereinheitlicht werden.¹⁷⁴ Nachdem bei der Generierung der Informationslandkarte bereits potentielle Synonyme entdeckt wurden, ist nun zu überprüfen, ob es sich bei den Informationen, die in der Priorisierungsphase als wichtig eingestuft wurden, um echte Synonyme handelt. Ist dies der Fall, müssen die Fachbereiche den künftig zu verwendenden Begriff festlegen. Dieser wird dann weiter mit den bestimmten Soll-Kennzahlen verglichen. Analog dazu erfolgt die Bereinigung von Homonymen. Werden im Zuge der Bereinigung mehrere Begriffe zusammengefasst, müssen die gewünschte Periodizität, die Quellsysteme und ihre verantwortlichen Organisationseinheiten sowie Dimensionen und Hierarchiestufen neu bestimmt werden.¹⁷⁵ Nachdem in Phase zwei bereits eine Analyse der Datenquellen vorgenommen wurde, muss der Informationsbedarf neu priorisiert werden, „da nicht vorhandene Daten oder mangelnde Datenqualität dazu führen, dass Informationen nicht zur Verfügung gestellt werden können.“¹⁷⁶ Um die Qualität der Daten zu analysieren, muss man zunächst überprüfen, ob die Daten bereits im Data Warehouse auf Grund der iterativen Entwicklung verfügbar sind, und ob die Anforderungen bezüglich der Dimensionen und Hierarchiestufen erfüllt sind. Ist dies nicht der Fall, müssen die operativen Systeme, die bereits bei der Konkretisierung ermittelt wurden, dahingehend überprüft werden, ob in ihnen Daten enthalten sind, die zur Berechnung der Kennzahl benötigt werden, und ob die Qualität dieser Daten dafür ausreichend ist. Diese Beurteilung sollte von Personen vorgenommen werden, die eine Schnittstellenfunktion zwischen Fachbereichen und Informatikabteilungen innehaben. Bei der Beurteilung, ob die Kennzahl umsetzbar ist oder nicht, sollten die Bedingungen einer Integration von externen Daten und der Bereinigung der Datenquellen angegeben werden.¹⁷⁷

¹⁷⁴ Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf), S. 15-16.

¹⁷⁵ Vgl.: Strauch, B. (2002), S. 200-202.

¹⁷⁶ Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf), S. 16.

¹⁷⁷ Vgl.: Strauch, B. (2002), S. 205-206.

5.3.4 Fachkonzeptentwurf

Das Ergebnis der Informationsbedarfsanalyse sollte ein Fachkonzeptentwurf sein, der durch ein multidimensionales semantisches Datenmodell repräsentiert wird. Dieser muss von den beteiligten Organisationseinheiten bewertet werden. Entspricht er deren Anforderungen nicht, so muss das Soll-Konzept erneut erstellt werden.¹⁷⁸

¹⁷⁸ Vgl.: Strauch, B. / Winter, R. (2002), URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004), S. 17.

6 Business Content

Beim Aufbau eines Data Warehouse spielt die Entwicklung des Datenmodells eine wichtige Rolle. Der Aufbau eines eigenen Datenmodells benötigt jedoch viel Zeit. Nicht selten dauert die Modellierung zwei bis drei Jahre, das Data Warehouse wird aber eigentlich sofort benötigt. In diesem Fall bietet es sich an, auf so genannte ‚generic data models‘ zurückzugreifen. Dabei handelt es sich um ein Datenmodell, welches speziell für eine Branche entwickelt wurde. Da Firmen, die sich in der gleichen Branche betätigen, ähnliche Daten verwenden und vergleichbare Informationsbedürfnisse haben, kann ein Modell, das einmal für eine solche Firma entwickelt wurde, als Template für andere Firmen verwendet werden.¹⁷⁹

Die SAP AG liefert als Add-On zum Business Information Warehouse den Business Content aus. Dabei handelt es sich um „[...] vorkonfigurierte, auf konsistenten Metadaten basierende rollen- und aufgabenbezogene Informationsmodelle.“¹⁸⁰ Zu diesen betriebswirtschaftlichen Lösungen gehören u.a. Rollen, Arbeitsmappen, Queries, InfoSources, InfoCubes, ODS-Objekte, Kennzahlen, Merkmale, Fortschreibungsregeln sowie Extraktoren für SAP R/3, mySAP.com Business Applications und andere Anwendungen.¹⁸¹ Manche Objekte sind branchenspezifisch, andere bereichs- und funktionspezifisch.¹⁸² Der Business Content kann entweder mit oder ohne Anpassungen übernommen werden oder als Beispiel für selbst erstellten Business Content dienen.¹⁸³

6.1 Komponenten des Datenmodells

Das Datenmodell des BW setzt sich aus unterschiedlichen BW-Objekten zusammen, die für verschiedene Aufgaben verwendet werden. Dabei handelt es sich um InfoObjekte, InfoCubes, ODS Objekte, InfoSets und MultiProvider. Man fasst diese Objekte unter dem Oberbegriff InfoProvider zusammen.¹⁸⁴ Dies sind Komponenten, die in Queries

¹⁷⁹ Vgl.: Inmon, W.H. (2000b), URL: <http://www.inmoncif.com/library/whiteprs/earlywp/ttgendm.pdf> (letzter Zugriff: 20. August 2004), S. 1.

¹⁸⁰ Seemann, A. / Schmalzriedt, B. / Lehmann, P. (2001), S. 197.

¹⁸¹ Seemann, A. / Schmalzriedt, B. / Lehmann, P. (2001), S. 197.

¹⁸² o. A. (2002b), URL: <http://www.sap.com/germany/media/50058307.pdf> (letzter Zugriff: 20. August 2004).

¹⁸³ Seemann, A. / Schmalzriedt, B. / Lehmann, P. (2001), S. 197.

¹⁸⁴ Vgl.: Mehrwald, C. (2004), S. 57.

verwendet werden können. Teilweise werden in diesen Objekten die Daten physisch vorgehalten. Andere stellen lediglich eine logische Sicht auf die Daten dar.¹⁸⁵

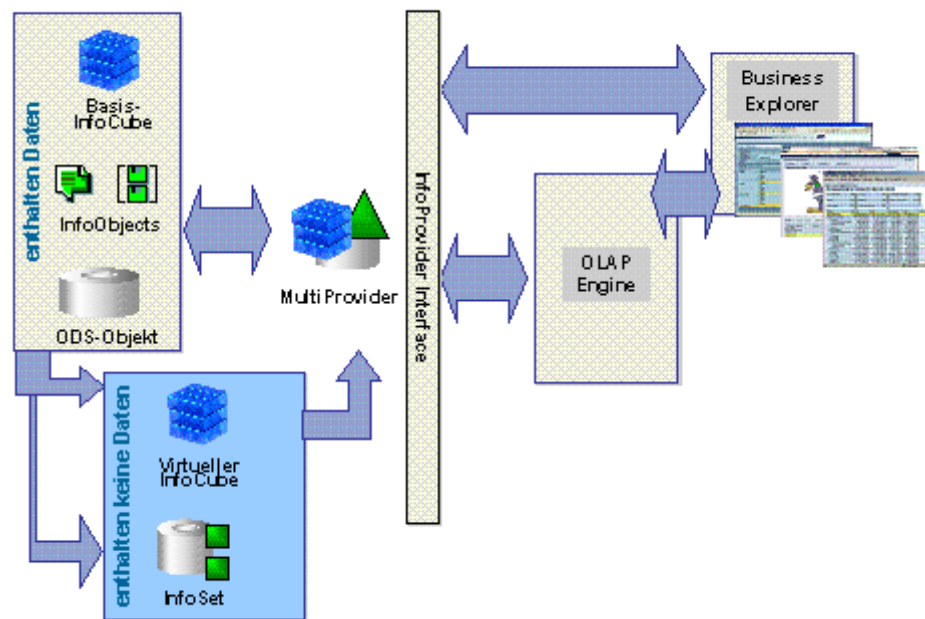


Abbildung 6-1: Übersicht über InfoProvider.¹⁸⁶

6.1.1 InfoObjekte

InfoObjekte stellen die Grundlage für alle weiteren InfoProvider. Diese können nur dann angelegt werden, wenn bereits ein entsprechendes InfoObjekt existiert. InfoObjekte sind durch einen systemweit eindeutigen technischen Namen und durch ihre Bezeichnung definiert. Es gibt vier verschiedene InfoObjekt-Typen: Merkmale, Zeiten, Einheiten und Kennzahlen.¹⁸⁷

- **Merkmale**

„Merkmale sind Bezugsgrößen mit betriebswirtschaftlicher Bedeutung nach denen eine sinnvolle Gruppierung von Kennzahlen möglich ist.“¹⁸⁸ Im nicht erweiterten

¹⁸⁵ Vgl.: o. A. (2004b), URL: http://help.sap.com/saphelp_nw04/helpdata/de/a7/d50f395fc8cb7fe1000000a11402f/frameset.htm (letzter Zugriff: 20. August 2004).

¹⁸⁶ Quelle: o. A. (2004b), URL: http://help.sap.com/saphelp_nw04/helpdata/de/a7/d50f395fc8cb7fe1000000a11402f/frameset.htm (letzter Zugriff: 20. August 2004).

¹⁸⁷ Vgl.: Mehrwald, C. (2004), S. 57-60.

¹⁸⁸ Mehrwald, C. (2004), S. 48.

Starschema werden Dimensionen durch einen Schlüssel identifiziert und die Merkmale in den Dimensionen angelegt, so dass keine weitere Identifikation nötig ist. Dieses Konzept gilt nicht im BW. Alle Merkmalswerte werden durch eine Stammdaten-Identifikationsnummer (SID) eindeutig bezeichnet. Die Merkmalswerte werden also nicht mehr in der Dimensionstabelle sondern in den damit verknüpften SID-Tabellen angegeben. Dadurch sind die Merkmale vollständig von den anderen BW-Objekten abgekapselt und somit kann dieselbe SID-Tabelle von mehreren Objekten genutzt werden.¹⁸⁹ Die Stammdaten sind also übergreifend gültig und stehen über alle InfoCubes zur Verfügung.¹⁹⁰ Da die SID-Tabelle von mehreren Objekten referenziert wird, ist das Löschen der Stammdaten nicht zulässig, wenn aus einem Objekt heraus noch auf den Eintrag verwiesen wird. Um eine unzulässige Löschung zu verhindern, wird ein entsprechendes Flag gesetzt, sobald die SID von weiteren Objekten verwendet wird. Allerdings wird das Flag nicht automatisch gelöscht, wenn die SID nicht mehr verwendet wird, so dass zeitaufwändige Überprüfungen notwendig sind. Der Name der SID hängt vom Namen des InfoObjekts ab.¹⁹¹ Ein InfoMerkmal kann mit Referenz auf ein anderes Merkmal angelegt werden. Somit gelten für dieses Merkmal die gleichen technischen Einstellungen wie für das referenzierte Merkmal. Der Unterschied liegt in der betriebswirtschaftlichen Semantik. So können bspw. die Beschreibung, die Darstellung und die Berechtigungsrelevanz vom Original abweichen. Diese Methode eignet sich insbesondere für das Anlegen von Partnerrollen, wenn diese die gleichen Stammdaten verwenden. Gäbe es die Möglichkeit der Referenzierung nicht, müsste für jede Rolle ein eigenes InfoObjekt angelegt und mit den gleichen Stammdaten versorgt werden.¹⁹²

Merkmale können über Texte und Stammdaten (Attribute), so genannte Master Data, verfügen. Texte werden als beschreibende Informationen für Merkmalsausprägungen (z.B. Kundenname) verwendet. Für diese Texte, die aus den Quellsystemen bezogen werden, gibt es unterschiedliche Längen und sie können auch sprachabhängig sein. Es ist möglich, die Texte stichtagsbezogen darzustellen. Allerdings bedeutet dies, dass bei jedem Laden ein neuer Gültigkeitszeitraum angefügt wird, so dass sich die Datensätze schnell vervielfachen. Attribute sind InfoObjekte, die einem Merkmal logisch untergeordnet sind. Man unterscheidet zwischen so genannten

¹⁸⁹ Vgl.: Mehrwald, C. (2004), S. 63-64.

¹⁹⁰ Vgl.: Hahne (2004), S. 3.

¹⁹¹ Vgl.: Mehrwald, C. (2004), S. 66.

¹⁹² Vgl.: Mehrwald, C. (2004), S. 82-83.

Anzeigeattributen und Navigationsattributen. Navigationsattribute können in einer Query selektiert werden. Beide Attributsformen können sowohl zeitkonstant als auch zeitabhängig sein. Die Anzeigeattribute werden in Klarform und nicht in Form der SID in den entsprechenden Stammdatentabellen gespeichert, während Navigationsattribute in Form der SID hinterlegt werden. Wenn es sich bei den Attributen um InfoObjekte handelt, die ebenfalls Attribute besitzen, so können diese auch als Attribut des Master-Data-InfoObjekts verwendet werden.¹⁹³

- **Zeiten**

Zeiten werden, obwohl sie die gleichen Funktionen erfüllen wie Merkmale, als eigene InfoObjekttypen betrachtet. Dabei können jedoch ausschließlich die Zeitmerkmale, die durch den Business Content zur Verfügung gestellt werden, verwendet werden, wobei ihre Bezeichnung an die individuellen Bedürfnisse einer Firma angepasst werden können. Wenn weitere Zeitmerkmale benötigt werden, kann der Modellierer diese als normale Merkmale anlegen. Allerdings können dann die Vorteile echter Zeitmerkmale, wie z.B. die Konvertierung beim Staging nicht wahrgenommen werden.¹⁹⁴

- **Kennzahlen**

Ein weiterer InfoObjekttyp ist die Kennzahl. Dabei handelt es sich um Fluss- oder Bestandsgrößen. Während Flussgrößen zeitraumabhängig sind, sind Bestandsgrößen zeitpunktbezogen. Abhängig vom Typ der Kennzahl existiert zu ihnen eine Einheit oder nicht. So haben Kennzahlen vom Typ Zahl, Integer, Datum oder Zeit keine Einheit, während für Kennzahlen vom Typ Betrag oder Menge fixe oder variable Einheiten definiert werden. Dies spielt insbesondere für die Datenanalyse eine große Rolle, wenn die Kennzahlen in Bezug zu einer Einheit gesetzt oder in andere Einheiten umgerechnet werden. Verfügt eine Kennzahl über eine fixe Einheit, so wird diese zusammen mit der Definition der Kennzahl gespeichert. Soll eine Kennzahl über unterschiedliche Einheiten verfügen, so muss zur Kennzahl ein weiteres InfoObjekt angelegt werden, das überall dort, wo die Kennzahl verwendet wird, zusätzlich die Einheit ablegt.¹⁹⁵

¹⁹³ Vgl.: Mehrwald, C. (2004), S. 72-81.

¹⁹⁴ Vgl.: Mehrwald, C. (2004), S. 66-67.

¹⁹⁵ Vgl.: Mehrwald, C. (2004), S. 68-71.

6.1.2 Hierarchien

Über externe Hierarchien werden Werte eines Merkmals gegliedert und nach bestimmten Kriterien zusammengefasst. Dabei besteht jede Hierarchie aus einem obersten Hierarchieknoten, der Wurzel und darunter liegenden Hierarchieknoten, welche die Struktur vorgeben, so dass unterschiedliche Hierarchieebenen abgebildet werden. Es gibt sowohl bebuchbare als auch nicht bebuchbare Hierarchieknoten. Nicht bebuchbare Knoten geben die Struktur der Hierarchie vor. Bei bebuchbaren Knoten handelt es sich um die Merkmalsausprägungen des InfoObjekts. „Bei der Datenanalyse werden die Kennzahlwerte aller Merkmale, die unter einem Hierarchieknoten eingeordnet sind, auf diesen Knoten aggregiert.“¹⁹⁶

Die nicht bebuchbaren Knoten können unterschiedlich tief verästelt sein. Dadurch kann die Struktur der Hierarchie auf Grund fachlicher Vorgaben gestaltet werden, ohne einer Einschränkung durch technische Bedingungen zu unterliegen. Es gibt zwei Arten von nicht bebuchbaren Knoten: Textknoten und fremde Merkmalsknoten. Textknoten bestehen aus einem beliebigen Knotennamen und einer Beschreibung. Sie werden zur Beschreibung von Hierarchiestrukturen verwendet. Es ist jedoch auch möglich, statt einer Bezeichnung, die Textinhalte anderer InfoObjekte zu hinterlegen. Dann spricht man von fremden Merkmalsknoten. Auf diese Weise sind die Bezeichnungen immer aktuell, da sie bei jeder Nutzung aus der Texttabelle des InfoObjekts ausgelesen werden. Allerdings muss dann zu jedem Hierarchieknoten das zugehörige InfoObjekt und ein Merkmalswert angegeben werden.

Über bebuchbare Knoten werden die Merkmalswerte eines InfoObjekts in die Hierarchie eingeordnet. Eine Sonderform stellen dabei Hierarchieintervalle dar. Bei einem normalen bebuchbaren Knoten muss ein exakter Merkmalswert definiert werden, bei einem Hierarchieintervall kann ein Nummernkreis von Werten definiert werden.¹⁹⁷

¹⁹⁶ Mehrwald, C. (2004), S. 86.

¹⁹⁷ Vgl.: Mehrwald, C. (2004), S. 85-88.

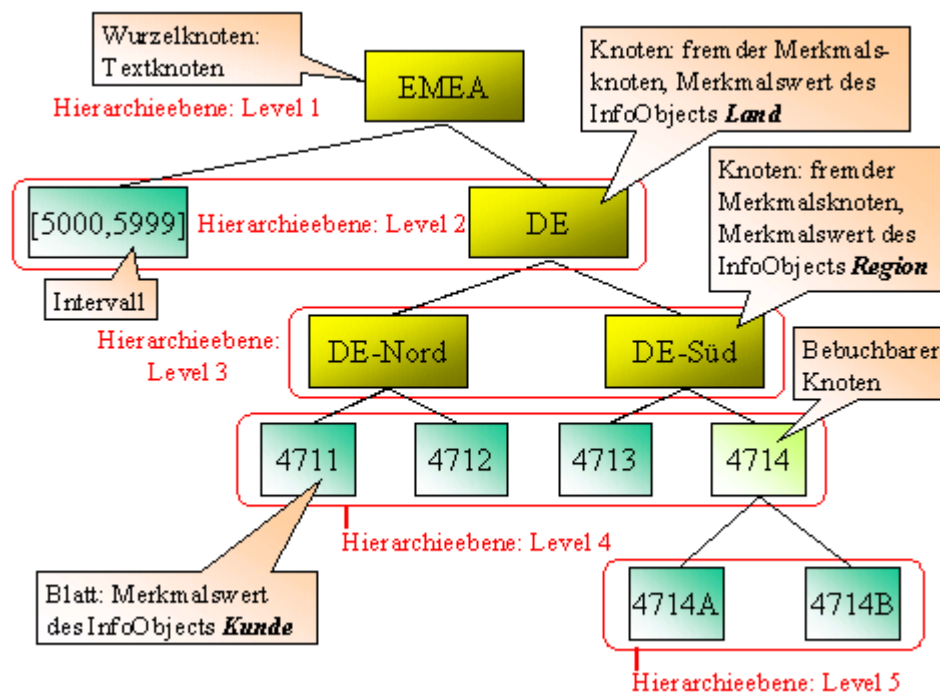


Abbildung 6-2: Unbalancierte Hierarchie mit bebuchbaren und nichtbebuchbaren Knoten.¹⁹⁸

Eine Hierarchie kann in mehreren Versionen existieren. Diese Versionen können für Planungs- und simulationsartige Reportings verwendet werden. Hierarchien sind jedoch nicht nur in mehreren Versionen vorhanden, sondern auch zeitabhängig. Die Beziehung der Knoten zueinander ändert sich also für bestimmte Zeitpunkte. Die Zeitabhängigkeit kann sich entweder auf die Gesamthierarchie oder auf die Hierarchiestruktur beziehen. Trifft ersteres zu, wird die Abhängigkeit über die Wurzel auf alle darunter liegenden Knoten übertragen. Bei der zeitabhängigen Hierarchiestruktur kann der Zeitraum pro Knoten festgelegt werden. Dadurch kann z.B. im Rahmen einer Umstrukturierung ein Mitarbeiter zu verschiedenen Zeitpunkten verschiedenen Kostenstellen zugeordnet werden.

Obwohl externe Hierarchien von den Master Data eines InfoObjekt losgelöst sind, beziehen sie sich auf genau ein InfoObjekt. Jedes InfoObjekt kann mehrere externe Hierarchien besitzen, wobei jedoch die Eigenschaften der Hierarchien in der InfoObjekt

¹⁹⁸ Quelle: O.A: (2004b), URL: http://help.sap.com/saphelp_nw04/helpdata/de/c1/493d3854129f68e1000009b38f8cf/content.htm (letzter Zugriff: 20. August 2004).

Pflege festgelegt werden. Daher sind auch alle Eigenschaften der Hierarchien, in denen das Objekt verwendet wird, identisch.¹⁹⁹

Hierarchien können auch über die Attribute eines Merkmals dargestellt werden. Zu einem InfoObjekt 0MATERIAL²⁰⁰ könnten dann die Produktstufen 1, 2, 3... angelegt werden. Allerdings könnten so nur balancierte Hierarchien dargestellt werden und Versionierungen sind auch nicht möglich. Darüber hinaus ist es möglich, die Hierarchien erst in der Dimension der InfoCubes anzugeben. Dadurch können die Attribute auch historisch dargestellt werden.²⁰¹

6.1.3 InfoCubes

Für die Speicherung mehrdimensionaler Daten im BW werden InfoCubes verwendet. Dabei werden die Daten nach dem Prinzip des erweiterten Sternschemas angeordnet. Sie beschreiben meist einen geschlossenen Datenbestand eines betriebswirtschaftlichen Bereichs und dienen der Speicherung und Analyse der Bewegungsdaten in Fakten- und Dimensionstabellen. Sie bieten die meisten analytischen Funktionen und ermöglichen auch eine gute Performance.²⁰² Man unterscheidet bei der physischen Datenablage zwischen Standard BasisCubes und transaktionalen BasisCubes. Als virtuelle Datenablage werden Remote und Virtuelle Cubes verwendet. Die Unterscheidung der Typen ist jedoch für das Reporting irrelevant.²⁰³ InfoCubes können in InfoAreas übersichtlich angeordnet werden.²⁰⁴

¹⁹⁹ Vgl.: Mehrwald, C. (2004), S. 88-89.

²⁰⁰ Anmerkung: die 0 bedeutet, dass es sich um ein mit dem Business Content ausgeliefertes Objekt handelt.

²⁰¹ Vgl.: Mehrwald, C. (2004), S. 148-150.

²⁰² Vgl.: Mehrwald, C. (2004), S. 90-91.

²⁰³ Vgl.: o. A. (2004b), URL: http://help.sap.com/saphelp_nw04/helpdata/de/8d/2b4e3cb7f4d83ee1000000a114084/content.htm (letzter Zugriff: 20. August 2004).

²⁰⁴ Vgl.: Seemann, A. / Schmalzriedt, B. / Lehmann, P. (2001), S. 136.

Beim Datenmodell der BasisCubes handelt es sich um ein erweitertes Starschema.

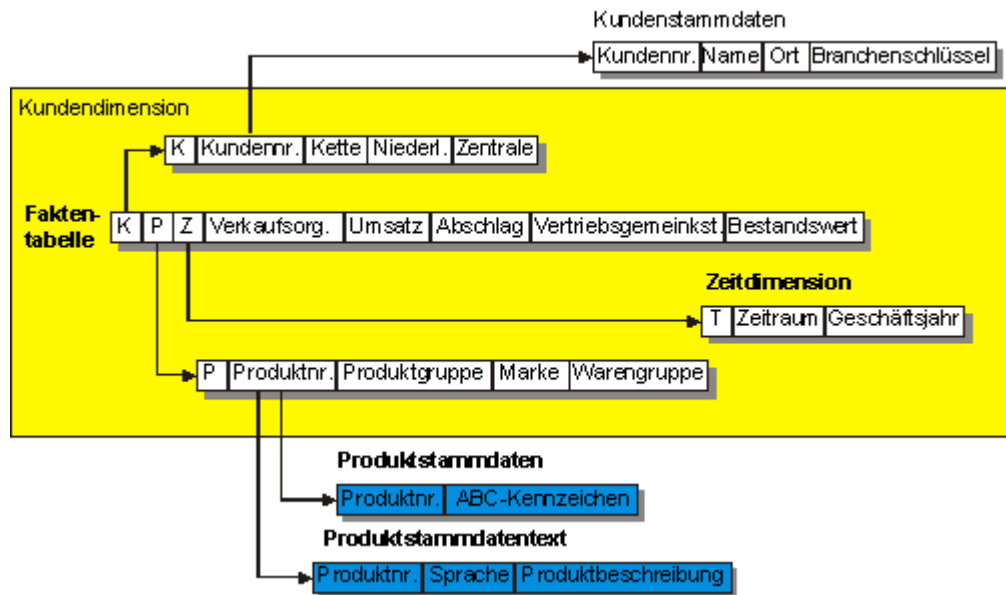


Abbildung 6-3: Erweitertes Sternschema des InfoCubes.²⁰⁵

Mittelpunkt des Schemas ist die Faktentabelle, in der die Kennzahlen des InfoCubes enthalten sind. Mehrere Dimensionstabellen, welche die Merkmale des InfoCubes beinhalten, umgeben diese. Die Merkmale bestimmen auch die Granularität der Kennzahlen. Verknüpft werden die Dimensionstabellen und die Faktentabellen über identifizierende Nummern (Dimension-ID's), die einen Teil der Schlüssel der jeweiligen Tabellen bilden. Über die Dimensionstabellen werden die Faktentabellen mit den Stammdatentabellen, in denen sich die eigentlichen Dimensionswerte befinden, verknüpft.²⁰⁶

Der Zusammenhang zwischen Stammdaten und Dimensionstabellen wird über die SID hergestellt, die in den Dimensionstabellen abgelegt werden. Dadurch können auch Funktionen der Master Data, Navigationsattribute, Texte, Zeitabhängigkeiten und externe Hierarchien genutzt werden.

Jeder Basis Cube verfügt über maximal 16 Dimensionen, wobei pro Dimension 253 InfoObjekte aufgenommen werden können. Drei Dimensionen davon sind bereits fest vorgegeben. Dabei handelt es sich um die Dimensionen Paket, Zeit und Einheit.

Eine besondere Dimensionsform ist die Line-Item-Dimension. Dabei wird die SID direkt in die Faktentabelle übernommen und somit die Performance verbessert. Allerdings bestehen solche Dimensionen nur aus einem InfoObjekt.

²⁰⁵ Quelle: o. A. (2004b), URL: http://help.sap.com/saphelp_nw04/helpdata/de/4c/89dc37c7f2d67ae10000009b38f889/content.htm (letzter Zugriff: 20. August 2004).

Eine wichtige Funktion der InfoCubes in Bezug auf die Datenanalyse sind komplexe Aggregationsverfahren. Eine typische Ausnahmeaggregation ist bspw. der durchschnittliche Lagerbestand. Andere InfoProvider sind lediglich in der Lage, Standardaggregationen wie Summenbildung oder die Ermittlung von Minimal- und Maximalwerten durchzuführen.

Eine besondere Form der Ausnahmeaggregation ist die Bestandsveränderung. Dabei werden die Bestandskennzahlen nicht im Datenmodell gespeichert, sondern auf Basis von Veränderungen berechnet. Man unterscheidet zwei Arten von Bestandskennzahlen. Zum einen die Bestandskennzahl mit Bestandsveränderung. Dabei berechnet sich die Bestandskennzahl aus genau einer anderen Kennzahl, die sowohl Zugänge als auch Abgänge darstellt. Die Bestandskennzahl mit Zu- und Abgängen errechnet sich aus zwei Kennzahlen, nämlich den Zu- und Abgängen. Eine weitere Bestandskennzahl ist, dass genau ein zeitliches Bezugsmerkmal und mehrere sachliche Bezugsmerkmale definiert werden. Erst im Anschluss an die Standard- und die Ausnahmeaggregation wird die Währungsaggregation durchgeführt, sofern diese benötigt wird.²⁰⁷

6.1.4 ODS-Objekte

In ODS-Objekten werden wie in Basis Cubes auch Bewegungsdaten gespeichert. Der Unterschied liegt jedoch darin, dass sich die Daten in den ODS Objekten nachträglich verändern können, während die Daten in InfoCubes nicht volatil sind. Sie werden in erster Linie für operatives Reporting und die Analyse volatiler Transaktionsdaten eingesetzt. Die Daten werden auf der Belegebene gespeichert und sonst nicht aggregiert.²⁰⁸

6.1.5 InfoSets

Info-Sets werden für kurzfristige und einmalige Analysen verwendet, für die kein BasisCube definiert ist. Dabei handelt es sich um Metaobjekte ohne eigene Datenhaltung. Im Falle einer Analyse werden die Daten des Info-Sets direkt aus den Datenquellen gelesen. Ihr Nachteil ist, dass nur die Abfrageergebnisse angezeigt werden. Navigation,

²⁰⁶ Vgl.: Hahne, M. (2004), S. 3.

²⁰⁷ Vgl.: Mehrwald, C. (2004), S. 92-100.

²⁰⁸ Vgl.: Mehrwald, C. (2004), S. 101-103.

die Verwendung externer Hierarchien oder die Nutzung von Stammdaten ist nicht möglich.²⁰⁹

6.1.6 MultiProvider

Für Datenanalysen, die sich über mehrere InfoProvider erstrecken wurden so genannte MultiProvider (ehemals MultiCubes) entwickelt. Der MultiProvider selbst enthält keine Daten, sondern bezieht seine Daten aus den zu Grunde liegenden InfoProvidern, deren Daten durch eine ‚Union Operation‘ zusammengefasst werden. Alle Werte der InfoProvider werden also zusammengeführt. Darin besteht ein wichtiger Unterschied zum InfoSet. Bei diesen werden mittels Joins nur die Werte zusammengeführt, die in beiden Tabellen vorhanden sind.²¹⁰

6.2 Grafische Modellierungsmethode

Für den konzeptionellen Entwurf gibt es eine Reihe von Modellierungsmethoden. Ein standardisiertes Verfahren für die logische grafische Modellierung der Datenbasis des BW existierte bislang nicht. Michael Hahne hat, um diese Lücke zu schließen, eine logische mehrdimensionale grafische Modellierungsmethode für das BW entwickelt, deren Bausteine als Templates für Microsoft Visio existieren. Problematisch ist jedoch, dass noch keine geeignete Transformationslogik zur Überführung eines semantischen multidimensionalen Modells in ein logisches BW-Modell existiert.²¹¹

6.2.1 Darstellung von Dimensionen

Zur Abbildung von Dimensionen und ihren Bestandteilen existiert das Objekt Dimension, welches den Dimensionsnamen beinhaltet. Merkmale werden durch das Objekt Characteristic dargestellt. Die Beschreibungen der Merkmalsausprägungen werden im Objekt Texts aufgenommen, zuzüglich der Information, ob es sich um einen kurzen, mittellangen oder langen Text handelt und ob er sprachabhängig ist. Da Merkmale außer Texten noch zusätzlich über Attribute verfügen können, werden diese über die Objekte Dis-

²⁰⁹ Vgl.: Mehrwald, C. (2004), S. 104-105.

²¹⁰ Vgl.: o. A. (2004b), URL: http://help.sap.com/saphelp_nw04/helpdata/de/52/1ddc37a3f57a07e10000009b38f889/content.htm (letzter Zugriff:20.08.2004).

²¹¹ Vgl.: Hahne, M. (2004), S. 1, 13.

play Attribute und Navigation Attribute dargestellt, wobei für die Darstellung von zeitabhängigen Attributen das Objekt Navigation Attribute Time-Dep entworfen wurde.²¹²

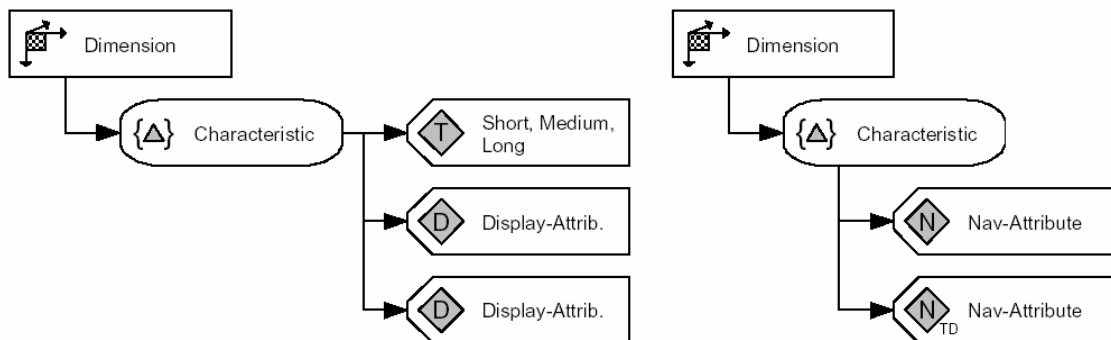


Abbildung 6-4: Logisches Modell der Dimension im SAP BW.²¹³

6.2.2 Hierarchien

Für die Darstellung von Hierarchien gibt es im BW drei Möglichkeiten. Zum einen können die Hierarchien in den Dimensionen des InfoCube angelegt werden. Andererseits können sie aber auch in den Stammdaten des Merkmals angegeben oder aber als externe Hierarchien gebildet werden.

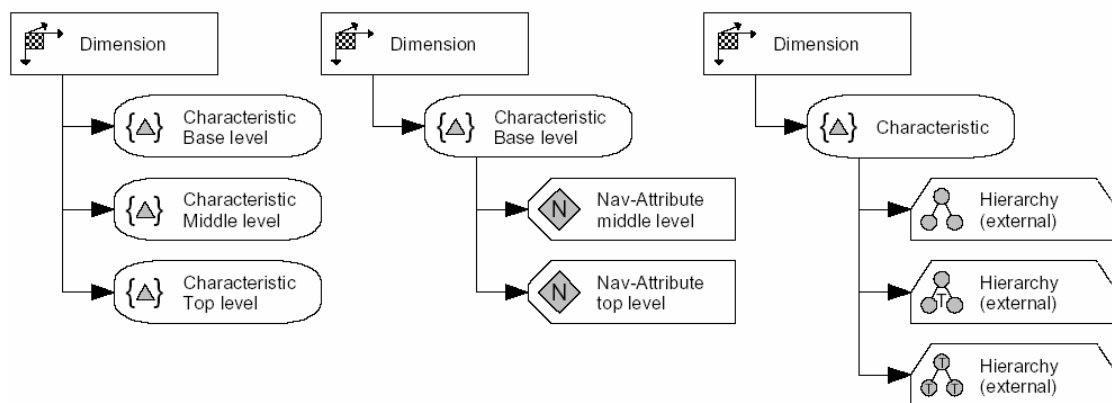


Abbildung 6-5: Logisches Modell von verschiedenen Hierarchien im BW.²¹⁴

²¹² Vgl.: Hahne, M. (2004), S. 8-9.

²¹³ Quelle: Hahne, M. (2004), S. 9.

²¹⁴ Quelle: Hahne, M. (2004), S. 9.

6.2.3 Darstellung von InfoCubes und MultiCubes

Um das Objekt BasisCube können bis zu dreizehn frei wählbare Dimensionen angegeben werden. Die Auflistung der Fakten erfolgt wie in ADAPT auch direkt im BasisCube. Auch die Dimensionen können ausführlich dargestellt werden. Aufgrund der Schema-komplexität kann diese jedoch auch getrennt erfolgen.

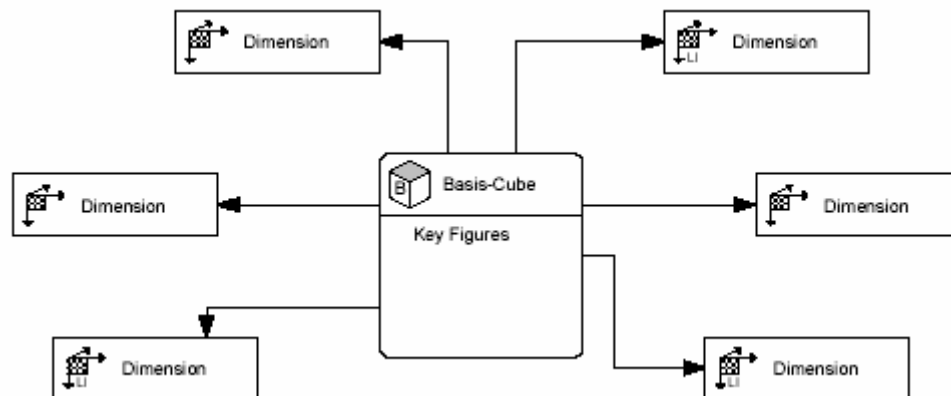


Abbildung 6-6: Logisches Modell des Basis Cube's mit Dimension.²¹⁵

Die Verbindung eines MultiCubes mit seinen InfoProvidern erfolgt über einen based upon Verbindungs Pfeil.²¹⁶

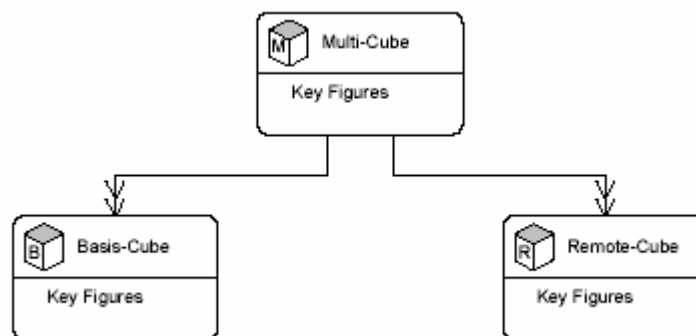


Abbildung 6-7: Logisches Modell des Multi Cubes.²¹⁷

²¹⁵ Quelle: Hahne, M. (2004), S. 10.

²¹⁶ Vgl.: Hahne, M. (2004), S. 10.

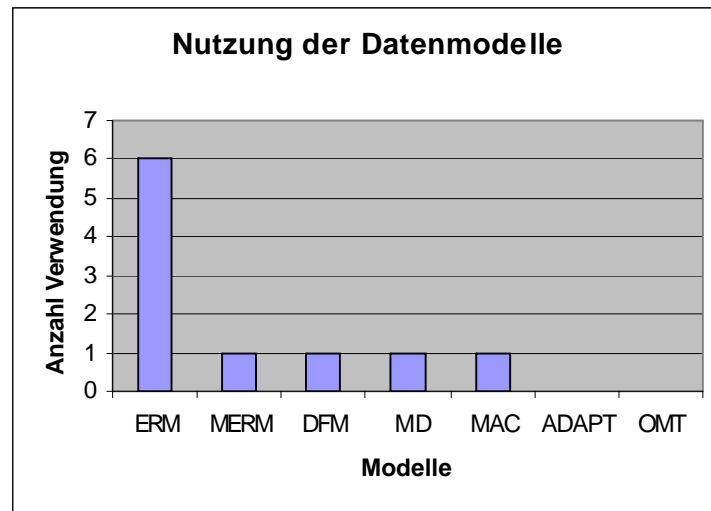
²¹⁷ Quelle: Hahne, M. (2004), S. 10.

7 Auswertung der Umfrage

Grundlage der Auswertung bilden neun teilweise nicht ganz vollständig ausgefüllte Fragebögen. Lediglich drei Projekte dieser Grundmenge beschäftigen sich mit dem Thema Enterprise Data Warehousing in dem Sinne, dass versucht wird, eine einheitliche Datenbasis für alle reportingrelevanten Daten im Unternehmen zu erstellen. Die anderen Projekte entwickeln Data Warehouses lediglich für einen bestimmten Bereich. Darüber hinaus verfügen die Data Warehouses von drei der Befragten nicht über einen so genannten ‚Single Point of Truth‘ und entsprechen somit nicht der idealtypischen Architektur eines Data Warehouses.

Vor diesem Hintergrund ist es verständlich, dass lediglich zwei Umfrageteilnehmer angeben, ein konzeptionelles unternehmensweites Datenmodell, welches auf den Daten der operativen Systeme basiert, bereits vor der Einführung eines Datenmodells besessen, und dieses für die Modellierung der DWH Basis verwendet zu haben. Überraschend ist jedoch, dass dies von zwei Teilnehmern angekreuzt wurde, welche nur bereichsspezifische Data Warehouses aufbauen. Eine mögliche Erklärung dafür wäre, dass in der Praxis unter ‚Unternehmensweit‘, nicht die Zusammenführung aller reportingrelevanten Daten des Gesamtunternehmens verstanden wird, sondern lediglich die konzernweit anerkannte Datenbasis für einen bestimmten Bereich.

Mit Ausnahme von zwei Projekten wurde die Datenbasis des DWH zunächst konzeptionell modelliert. Teilweise wurde die Datenbasis mit mehreren Modellen dargestellt. Kombiniert wurden das ERM und das MD, das ERM und MAC, das ERM, das MERM und das DFM. Das meistgenutzte Modell ist dabei das ERM.

Abbildung 7-1: Nutzung der Datenmodelle.²¹⁸

Die in der Literatur als wichtig deklarierten multidimensionalen Modelle finden also in der Praxis kaum Verwendung, außer in Kombination mit dem ERM, welches in der Literatur als ungenügend für die Modellierung der Data-Warehouse-Datenbasis angesehen wird.

Ansonsten wurde einmal die SAP Methode ASAP (Accelerated SAP) angegeben. Da es sich dabei jedoch um eine Methode handelt, die speziell für das BW konzipiert ist, handelt es sich hierbei nicht um eine konzeptionelle Entwurfsmethodik, da diese von jeglichem System unabhängig ist.²¹⁹

Zur Modellierung der Datenbasis verwenden die meisten Befragten mehrere Modellierungstools. Einer der Befragten modelliert auf ‚herkömmliche‘ Weise mit Papier und Bleistift. Am häufigsten wurden PowerPoint, Excel und Visio verwendet. Dies hat mich überrascht, da es in PowerPoint keine Funktion gibt, um bspw. die syntaktische Konsistenz eines Modells zu gewährleisten. Darüber hinaus ist es ungewöhnlich, dass zwar Visio verwendet wird, die Modellierungsmethode ADAPT, für die es spezielle Visio Vorlagen gibt, jedoch nicht.

In der Literatur wurden verschiedene Anforderungen erhoben, die ein konzeptionelles Modell erfüllen sollte. Für jede dieser Forderungen konnte in der Umfrage auf einer

²¹⁸ Eigene Darstellung.

Skala von 1 bis 5 (im Unfragebogen unwichtig bis sehr wichtig) bewertet werden, ob sie in der Praxis wirklich relevant ist. Berechnet wurde der Wert, indem die Punkte addiert wurden. Bei neun Teilnehmern konnte eine Forderung maximal 45 Punkte (9 Teilnehmer und Höchstpunktzahl 5 (sehr wichtig)) erreichen. In der Praxis wurden die Forderungen nach einer Möglichkeit, Beziehungen zwischen Fakten abzubilden und Hierarchieebenen darzustellen, als am wichtigsten erachtet.

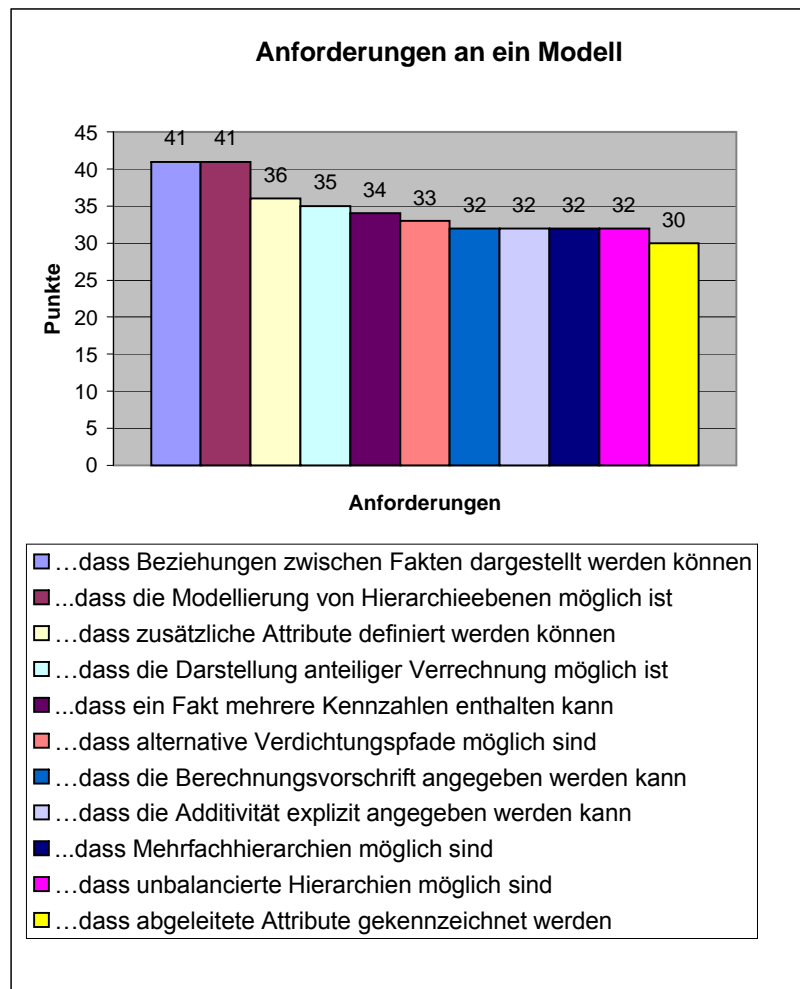


Abbildung 7-2: Anforderungen an ein Modell.²²⁰

Nach den Angaben der Befragten sollten im Datenmodell noch Binnen-Umsatz-Eliminierung, Zeitabhängigkeiten, Zeitabhängige Stammdaten, Konsolidierungseffekte

²¹⁹ Für eine genaue Beschreibung Vgl.: o. A. (2000): Multi- Dimensional Modeling with BW. ASAP for BW Accelerator.

²²⁰ Eigene Darstellung.

und die Datenhaltung mit unterschiedlichen Granularitäten im gleichen Cube dargestellt werden können.

Als Probleme bei der Datenmodellierung wurden Brüche zwischen Operativen Systemen und Data Warehouse Applikation Integration, Unbalancierte Hierarchien, Datenqualität und Datenlücken, die Flexibilität bei Aggregationsverhalten und Probleme bei der Umsetzung von Berechtigungskonzepten genannt.

Lediglich fünf der Befragten haben ein Referenzmodell verwendet, und nur drei der Umfrageteilnehmer haben ein Vorgehensmodell zur Modellierung. Dabei wurde in zwei Projekten, in welchen ein Vorgehensmodell zur Modellierung existiert, auch ein Referenzmodell verwendet. Ob dabei ein zwingender Zusammenhang besteht, kann jedoch auf Grund der statistisch geringen Masse nicht festgestellt werden. Die Pflege des Datenmodells ist bei sechs Teilnehmern organisatorisch geregelt. Auffällig daran ist, dass alle Beteiligten, die ein Referenzmodell verwendet haben, auch die Pflege des Datenmodells organisatorisch festgelegt haben, so dass in diesem Fall vermutet werden kann, dass ein Zusammenhang besteht. Die Verantwortung für das Modell ist unterschiedlich verteilt. In einigen Projekten sind die Bereiche, für die das DWH entwickelt wird, zuständig. In anderen Projekten ist entweder die IT-Abteilung eines Unternehmens verantwortlich oder die Projektgruppe, die für die gesamte Entwicklung des DWH zuständig ist.

Die Konsistenz wird zum einen über technische Überprüfungen wie Summen- und Plausibilitätskontrollen sichergestellt, zum anderen über Berechtigungskonzepte. So können bspw. nur die Entwickler, oder nur diejenigen Mitarbeiter, die auch die Verantwortung für das Modell tragen, nach Absprache mit den Usern Änderungen durchführen. Darüber hinaus wurden teilweise zunächst unternehmensweit gültige Definitionen für Merkmale und Kennzahlen festgelegt und das Datenmodell in Kooperation zwischen der IT-Abteilung und den Geschäftsbereichen definiert. In einem anderen Projekt wurde ein Qualitätssicherungsteam mit der Wahrung der Konsistenz beauftragt. Außerdem wurde der SAP Business Content verwendet. Lediglich bei zwei Befragten ist die Konsistenz nicht sichergestellt und bei einem weiteren Projekt wird das Datenmodell nur im Falle eines Software-Updates nachgepflegt.

Die Datenmodellierung wurde häufig modular vorgenommen. Um die Informationsbedürfnisse der Anwender zu erfahren, wurden Untersuchungen in den entsprechenden Geschäftsbereichen vorgenommen. Dabei wurde sowohl das Informationsbedürfnis der Mitarbeiter erfragt, als auch die bereits in den operativen Systemen vorhandenen Daten verwendet. Allerdings wurden bereits hier teilweise Performanceaspekte in die Modellierung mit einbezogen. In einem Projekt erfolgte die Modellierung auch in Zusammenarbeit mit dem Softwarehersteller.

Bis auf einen Befragten waren sich alle Umfrageteilnehmer der besonderen Bedeutung der semantischen Definition von Kennzahlen und Merkmalen bewusst. In einem Projekt wurde alleine daran ein Jahr gearbeitet. In einem anderen Projekt wurden speziell zu diesem Thema Schulungen für die Anwender veranstaltet.

Fünf der Befragten nutzen das SAP BW. Ein weiterer plant die Nutzung des BW, des Business Content, des Repositories und eine Dokumentation der Änderungen. Zwei der fünf Nutzer verwenden den Business Content jedoch nicht. In den übrigen drei Fällen werden Extraktoren genutzt, in zwei Fällen auch die InfoCubes und einer der Befragten verwendet auch die Queries. Zwei der Befragten nutzen das BW Repository.

Bis auf einen dokumentieren alle BW Nutzer die Änderungen ihres Datenmodells. Auch die Nicht-SAP-Nutzer dokumentieren ihre Änderungen. Acht der neun Befragten wünschen sich ein grafisches Tool, um den Business Content zu visualisieren.

8 Fazit

Wie in dieser Arbeit deutlich wurde, gibt es eine Reihe von Modellierungsansätzen, die eine adäquate Darstellung der multidimensionalen Datenstrukturen ermöglichen. Je mächtiger aber eine solche Methodik ist, umso komplizierter ist sie auch. Außerdem werden teilweise wie bspw. in der $mUML$, Aspekte der konzeptionellen und der logischen Modellierung (bspw. Angabe eines Datentyps) gemischt.

Auch das St. Galler Vorgehensmodell, welches in Teilen bereits in der Praxis erprobt ist, stellt eine gute Anleitung zur Gewinnung der analyserelevanten Daten dar, da es sowohl die Anforderungen der späteren Anwender als auch die Daten der operativen Systeme und ihre Qualität berücksichtigt.

Was in der Praxis jedoch fehlt ist ein Tool, welches durchgängig die Modellierung aller drei Entwurfsebenen unterstützt. Dies wird auch durch die Umfrage bestätigt. Die meisten Befragten verwendeten zur Modellierung ihrer Daten verschiedene Tools. Eventuell wäre es für Nutzer des BW sinnvoll, ADAPT als konzeptionelles Modell zu verwenden, welches dann über noch fehlende geeignete Transformationsschritte auf das logische Modell von Michael Hahne übertragen würde. ADAPT eignet sich deshalb gut, da Hahne ähnliche Symbole zur grafischen Darstellung verwendet und beide Modellierungsansätze als Templates für Visio existieren. Ideal wäre es, wenn es noch möglich wäre, über XML die entsprechenden Datenstrukturen auch automatisch im BW anzulegen.

Dass zwischen Theorie und Praxis ein großer Unterschied besteht, wurde durch die Ergebnisse der Umfrage deutlich. Obwohl in der Forschung eine Reihe von Modellierungsansätzen und prototypartigen Tools zur Modellierung entworfen wurden, die die multidimensionale Abbildung von Daten ermöglichen, werden diese in der Praxis kaum verwendet. Tatsächlich wird der konzeptionellen Modellierung und der Pflege dieses Datenmodells teilweise nur eine geringe Bedeutung beigemessen.

Literaturverzeichnis

- Abello, A. / Samos, J. / Saltor, F. (2000): A Data Warehouse Multidimensional Data Models Classification. URL: <http://www-lsi.ugr.es/~bdf/Trabajos/lsi0006ugr.pdf> (letzter Zugriff: 20. August 2004)
- Blaschka, M. / Sapia, C. / Höfling, G. (1999): REPORT. An Overview of Multidimensional Data Models for OLAP. URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http:zSzzSzwww.forwiss.dezSz~system42zSzpublicationszSztechreport.pdf/sapia99overview.pdf/> (letzter Zugriff: 20. August 2004)
- Blaschka, M. / Sapia, C. / Höfling, G. / Dinter, B. (1998): Finding Your Way through Multidimensional Data Models. URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http:zSzzSzwww.forwiss.dezSz~system42zSzpublicationszSzdwdot.pdf/blaschka98finding.pdf/> (letzter Zugriff: 20. August 2004)
- Böhnlein, M. (2001): Konstruktion semantischer Data-Warehouse-Schemata. Wiesbaden, Deutscher Universitäts-Verlag, 2001.
- Böhnlein, M. / Ulbrich-vom-Ende, A. (2000): Grundlagen des Data Warehousing. Modellierung und Architektur. URL: <http://www.seda.wiai.uni-bamberg.de/ceus/publikationen/downloads/BoUI2000.pdf> (letzter Zugriff: 20. August 2004)
- Bulos, D. / Forsman, S. (1998): Getting Started with ADAPT™. OLAP Database Design. URL: http://www.symcorp.com/downloads/ADAPT_white_paper.pdf (letzter Zugriff: 20. August 2004)
- Cabibbo, L / Torlone, R. (1998): A Logical Approach to Multidimensional Databases. URL: <http://cabibbo.dia.uniroma3.it/pub/pdf/edbt1998.pdf> (letzter Zugriff: 20. August 2004)
- Determann, L. (2002): Modellierung Analytischer Informationssysteme: Ein Konzept zur multidimensionalen Datenstrukturierung. Wiesbaden: Deutscher Universitäts-Verlag, 2002.
- Gabriel, R. / Gluchowski, P (1997): Semantische Modellierungstechniken für multidimensionale Datenstrukturen. In: HMD-Theorie und Praxis der Wirtschaftsinformatik, Heft 195 / 1997, S. 18-37.
- Gabriel, R. / Gluchowski, P. (1998): Grafische Notationen für die semantische Modellierung multidimensionaler Datenstrukturen in Management Support Systemen. In: Wirtschaftsinformatik 40, Heft 6 / 1998, S. 493-502.

- Golfarelli, M. / Maio, D. / Rizzi, S. (1998):** Conceptual Design of Data Warehouses from E/R Schemes. URL: <ftp://ftp-db.deis.unibo.it/pub/stefano/hicss98.pdf> (letzter Zugriff: 20. August 2004)
- Golfarelli, M. / Maio, D. / Rizzi, S. (1998):** The Dimensional Fact Model: A Conceptual Model For Data Warehouses. URL: <http://www-db.deis.unibo.it/~srizzi/PDF/ijcis98.pdf> (letzter Zugriff: 20. August 2004)
- Hahne, M. (2003):** Logische Datenmodellierung zur Abbildung mehrdimensionaler Datenstrukturen im SAP Business Information Warehouse. URL: <http://www.btw2003.de/proceedings/paper/IP8.pdf> (letzter Zugriff: 20. August 2004)
- Hahne, M. (2004):** Grafische Repräsentation mehrdimensionaler Datenmodelle des SAP Business Information Warehouse. URL: http://www.cundus.de/downloads/MKWI2004_Hahne_paper.pdf (letzter Zugriff: 20. August 2004)
- Harren, A. (1999):** Konzeptuelles Data Warehouse-Design. Diplomarbeit, Universität Oldenburg, Fachbereich Informatik, 1999.
- Heinrich, L. / Heinzl, A. / Roithmayr, F. (2004):** Wirtschaftsinformatik-Lexikon. München; Wien: Oldenbourg, 2004.
- Herden, O. (2001):** Eine Entwurfsmethodik für Data Warehouses. Dissertation, Universität Oldenburg, Fachbereich Informatik, 2001. URL: <http://deposit.ddb.de/cgi-bin/dokserv?idn=964097281> (letzter Zugriff: 20. August 2004)
- Hettler, D. / Preuss, P. / Niedereichholz, J. (2003):** Vergleich ausgewählter Ansätze zur semantischen Modellierung von Data-Warehouse-Systemen. In: HMD-Theorie und Praxis der Wirtschaftsinformatik, Heft 231 / 2003, S. 97-107.
- Holthuis, J. (1999):** Der Aufbau von Data Warehouse-Systemen: Konzeption - Datenmodellierung - Vorgehen. Wiesbaden: Deutscher Universitätsverlag, Gabler, 1999.
- Inmon, W.H. (1997):** Iterative Development Using A Data Model. URL: <http://www.inmoncif.com/library/whiteprs/techtopic/tt03.pdf> (letzter Zugriff: 20. August 2004)
- Inmon, W.H. (2000a):** Creating The Data Warehouse Data Model From The Corporate Data Model. URL: <http://www.inmoncif.com/library/whiteprs/earlywp/ttdwdmod.pdf> (letzter Zugriff: 20. August 2004)
- Inmon, W.H. (2000b):** Using the Generic Data Model. URL: <http://www.inmoncif.com/library/whiteprs/earlywp/ttgendm.pdf> (letzter Zugriff: 20. August 2004)

- Jung, R. / Winter, R. (1998):** Data-Warehouse-Design: Ansätze für eine effiziente Modellierung komplexer, dynamischer Schemata. URL: <http://www.ap.iwi.unibe.ch/publikationen/resource/mobis98.pdf> (letzter Zugriff: 20. August 2004)
- Mehrwald, C. (2004):** SAP Business Information Warehouse 3: Architektur, Konzeption, Implementierung. Heidelberg: dpunkt Verlag, 2004.
- Mucksch, H. / Behme, W.(2000):** Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik. In: Mucksch, H. / Behme, W. (Hrsg.) (2000): Das Data Warehouse-Konzept: Architektur-Datenmodelle-Anwendungen. Wiesbaden: Gabler, 2000. S. 3-80
- o. A (2000):** Multi Dimensional Modeling with BW: ASAP for BW Accelerator. URL: http://help.sap.com/bestpractices/crossindustry/businessintelligence/v133/documentation/Multi-dimensional_modeling_EN.doc (letzter Zugriff: 20. August 2004)
- o. A. (2002):** Data Warehousing mit mySAP™ Business Intelligence. URL: <http://www.sap.com/germany/media/50058307.pdf> (letzter Zugriff: 20. August 2004)
- o. A. (2004a):** Unternehmen vernachlässigen ihre taktischen Inforamtionen: Kaum ein CEO traut internen Daten – Amerikaner haben mehr Probleme als Europäer. In: Computer Zeitung Nr.19 / 2004, S. 2
- o. A. (2004b):** SAP Bibliothek-Modellierung. URL: http://help.sap.com/saphelp_nw04/helpdata/de/a7/d50f395fc8cb7fe10000000a11402f/frameset.htm (letzter Zugriff: 20. August 2004)
- o. A. (o. D.):** Adapt Help. URL:www.symcorp.com(letzter Zugriff: 20. August 2004)
- Poe, V. / Reeves, L. (1997):** Aufbau eines Data Warehouse. München: Prentice Hall, 1997.
- Preuschhoff, S. (2003):** Enterprise Data Warehousing. Gegenstand, Ansätze, Technologien. Diplomarbeit, Fachhochschule Stuttgart – Hochschule der Medien, Fachbereich Information und Kommunikation, 2003.
- Rosemann, M. (1998):** Die Grundsätze ordnungsmäßiger Modellierung-Intention, Entwicklung, Architektur und Multiperspektivität. In: Maicher, M. / Scheruhn, H.-J.: Informationsmodellierung – Referenzmodelle und Werkzeuge. Wiesbaden: Gabler, 1998. S. 1-21.

- Sapia, C. / Blaschka, M. / Höfling, G. / Dinter, B. (1998): Extending the E/R Model for the Multidimensional Paradigm. URL: <http://citeseer.ist.psu.edu/cache/papers/cs/21200/http:zSzzSzwww.forwiss.dezSz~system42zSzpublicationszSzdw98.pdf/extending-the-e-r.pdf> (letzter Zugriff: 20. August 2004)
- Schelp, J. (2000): Modellierung mehrdimensionaler Datenstrukturen analysierender Informationssysteme. Wiesbaden: Deutscher Universitäts-Verlag, Gabler, 2000.
- Seemann, A. / Schmalzriedt, B. / Lehmann, P. (2001): SAP Business Information Warehouse. Bonn: Galileo Press GmbH, 2001.
- Strauch, B. (2002): Entwicklung einer Methode für die Informationsbedarfsanalyse im Data Warehousing. Dissertation, Universität St. Gallen. Bamberg: Difo-Druck GmbH, 2002.
- Strauch, B. / Winter, R. (2002): Vorgehensmodell für die Informationsbedarfsanalyse im Data Warehouse. URL: [http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/\\$file/dw2002.pdf](http://verdi.unisg.ch/org/iwi/iwi_pub.nsf/wwwPublRecentGer/173010B2CD2D63E9C1256D090050B55F/$file/dw2002.pdf) (letzter Zugriff: 20. August 2004)
- Totok, A. (2000): Modellierung von OLAP- und Data Warehouse Systemen. Wiesbaden: Gabler, 2000.
- Tsois, A. / Karavannidis, N. / Sellis, T. (2001): MAC: Conceptual Data Modeling for OLAP. URL: <http://www.dbnet.ece.ntua.gr/pubs/uploads/TR-2001-5.pdf> (letzter Zugriff: 20. August 2004)
- Verlag für Deutsche Wirtschaft AG (2004): www.zitate.de- Ihr perfektes Zitat 2004. URL: <http://www.zitate.de> (letzter Zugriff: 20. August 2004)
- Vossen, G. (2000): Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme. München; Wien: Oldenbourg, 2000.
- Wieken, J.-H. (1999): Der Weg zum Data warehouse: Wettbewerbsvorteile durch strukturierte Unternehmensinformationen. München: Addison-Wesley, 1999.
- Wirtz, K. (2002): Der Data-Warehouse-Rahmenplan: Entwicklung eines konzeptionellen Schemas. Wiesbaden: Deutscher Universitäts-Verlag, 2002.

Copyrightverweis

SAP Business Information Warehouse, SAP R / 3, mySAP und SAP Business Content sind eingetragene Marken der SAP AG.

Microsoft Visio, EXCEL und PowerPoint sind eingetragene Marken der Microsoft Corporation.

ADAPT ist ein eingetragenes Warenzeichen der Symmetry Corporation.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Diplomarbeit selbständig angefertigt habe. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht.

Ort, Datum

Unterschrift