

Henning Baars – Bodo Rieger (Hrsg.)

**Perspektiven der betrieblichen Management- und  
Entscheidungsunterstützung**

*Forschungskolloquium Business Intelligence*

der GI Fachgruppe 5.8 Management Support Systems  
in Zusammenarbeit mit dem Lehrstuhl für  
Wirtschaftsinformatik der TU Dortmund

am 08. und 09. Oktober 2009

*im Universitätskolleg Bommerholz der  
Technischen Universität Dortmund*

**Tagungsband**

# **Perspektiven der betrieblichen Management- und Entscheidungsunterstützung**

*Forschungskolloquium Business Intelligence*

der GI Fachgruppe 5.8 Management Support Systems  
in Zusammenarbeit mit dem Lehrstuhl für Wirtschaftsinformatik der TU Dortmund

## **Vorwort**

Unter der Überschrift „Business Intelligence“ sind in vielen Unternehmen integrierte Infrastrukturen für die Managementunterstützung aufgebaut worden. Das starke Wachstum der Lösungen, ihre Ausdehnung in immer neue Anwendungsdomänen sowie die zunehmende Verzahnung mit der operativen Prozessunterstützung werfen eine Vielzahl an neuen organisatorischen, betriebswirtschaftlichen und technischen Fragen auf, die sich mit dem verfügbaren konzeptionellen Instrumentarium bislang oftmals nur unbefriedigend abdecken lassen. Die deutschsprachige Wirtschaftsinformatik mit ihrer lösungsorientierten und interdisziplinären Ausrichtung sowie ihrer langen Tradition in der Erforschung von Management Support Systemen ist bestens positioniert, um sich den entsprechenden Herausforderungen zu stellen und mit wissenschaftlicher Rigorosität neue Perspektiven aufzuzeigen.

Ziel des Kolloquiums war es, innovative Forschungsansätze und Forschungsergebnisse aus den Themenfeldern Business Intelligence (BI) und integrative Management Support Systeme (MSS) vorzustellen, zu diskutieren und zueinander in Beziehung zu setzen. Hierbei wurden bewusst auch Arbeiten in frühen Forschungsstadien berücksichtigt, um so Entwicklungslinien im Bereich der BI-Forschung aufzudecken, Kooperationspotentiale offenzulegen und für weitere Forschungsaktivitäten neue Impulse zu geben.

Der vorliegende Tagungsband stellt die diskutierten Beiträge zusammen. Die Spanne der von den elf Autoren behandelten Themen reicht dabei von primär organisatorisch-konzeptionellen Arbeiten über Modellierungsansätze bis hin zum Entwurf technischer Architekturen.

---

Dem aktuellen Thema „Operational Business Intelligence“ widmet sich der Beitrag von Herrn Dipl.-Wirt. Inform. Sebastian Neuhaus, der Operational BI von diversen verwandten Konzepten abgrenzt sowie die Möglichkeiten einer Kombination service- und ereignisorientierter Architekturen herausarbeitet.

Herr Dipl.-Kfm. Christian Schultewolter befasst sich im Anschluss mit der Fehleranfälligkeit modellgetriebener Decision Support Systeme (MDSS). Er stellt die Notwendigkeit einer systematischen Unterstützung der Modellbildung für MDSS heraus und erarbeitet Klassifizierungskriterien von Modellierungsfällen als Grundlage für eine konzeptuelle Modellierungssprache.

Der für die Wirtschaftsinformatik grundlegenden Thematik der „Vorgehensmodelle“ widmet sich Herr Prof. Dr. Stephan König in seinem Beitrag „Ein Wiki-basiertes Vorgehensmodell für Business Intelligence Projekte“. Die Zielsetzung dabei ist es, ein praxistaugliches Vorgehensmodell zu entwickeln. Das hierfür entwickelte Schema liefert auch die Grundstruktur für den Entwurf eines Wikis zur Unterstützung des Vorgehensmodells.

Mit der systematischen Erschließung des in einem BI-System generierten Wissens beschäftigen sich Herr Dipl.-Inform. Matthias Mertens, Frau Dipl.-Inform. Yvette Teiken und Herr Prof. Dr. Dr. h.c. Hans-Jürgen Appelrath. Hierzu schlagen sie den Einsatz von Ontologien vor, um eine maschinenverständliche Erfassung und Verwaltung von Metadaten zu Analyseprozessen und Analyseergebnissen sowie deren Nutzung zu ermöglichen und semantische Such- und Navigationsmöglichkeiten in einem analytischen Informationssystem bereitzustellen.

Eine domänenspezifische Abschätzung der Potentiale von BI-Lösungen liefern Herr Prof. Dr. Andreas Hilbert und Herr Dipl.-Wirtsch.-Inf. Alexander E. Müller in „Business Intelligence bei Entscheidungen im Human Resource Management“. Hierfür systematisieren die Autoren Entscheidungsprobleme im Personalbereich und diskutieren die Eignung unterschiedlicher Systemklassen für ihre Unterstützung.

Einen neuen Ansatz zur Gestaltung analytischer Informationssysteme präsentieren Herr Dipl. Wirt. Inf. Christian Kurze und Herr Prof. Dr. Peter Gluchowski. Die Autoren demonstrieren mit ihrem Konzept sowie einem darauf aufsetzenden Prototypen, wie bei einer Portierung des Ansatzes der „Model Driven Architecture“ eine Erhöhung des Automatisierungsgrades bei der Erstellung und Analyse multidimensionaler Analysesysteme erzielt werden kann.

Herr Dr. Ralf Knackstedt und Herr mgr Łukasz Lis charakterisieren in ihrem Beitrag die Modellierungstechnik H2 for Reporting und leiten auf dieser Basis Erweiterungspotenzia-

le und Forschungsbedarfe ab. Ihre Ausführungen illustrieren die Autoren an dem innovativen Anwendungsgebiet von Forschungsportalen mit analytischen Funktionen.

Im Umfeld des „Web 2.0“ einzuordnen ist die Arbeit von Herrn Prof. Andreas Hilbert und Herrn Dipl.-Wirt.-Inf. Andreas Schieber, die den Stand der Forschungen im Bereich der Analyse von Weblogs beleuchten und eine Konzeption für ein entsprechendes Analysesystem ableiten.

Den Bereich des Managements von BI-Organisationen strukturiert und diskutiert Herr Dipl. Inf. Walid Mehanna. Er entwirft hierfür ein Rahmenkonzept in Form eines Referenzmodells, dessen Implementierung durch ein zugehöriges Dokumentationswerkzeug unterstützt wird.

Herr Prof. Dr. Dr. h. c. mult. Peter Mertens skizziert in seinem Beitrag die Vision eines Systems zur Automatisierung der Berichterstattung von Kontrollgremien (Aufsichtsräte, Verwaltungsräte, Aufsichtsbehörden). Er geht dabei auf die Bedeutung sowie die besonderen Herausforderungen einer solchen Anwendung ein und entwirft erste konkrete Vorschläge zur Realisierung.

Mit dem Multi-Channel-Management im Kontext des Customer Relationship Management setzen sich Herr Prof. Dr. Andreas Hilbert und Herr Dipl. Wirt.-Inf. Stefan Sommer auseinander. Sie entwickeln eine Systematisierung für den elektronischen Kaufprozess, mit der die Unterschiede des Kundenbeziehungsmanagement kanal- und phasenübergreifend im Zeitverlauf verdeutlicht werden können.

Die Vielschichtigkeit der diversen Beiträge verdeutlicht den aktuellen Stand der BI- und MSS-Forschung, die unseres Erachtens eine interessante Wegmarke erreicht hat: Dadurch, dass sich in der Praxis bereits grundlegende Lösungsansätze für die Managementunterstützung etabliert haben, kann bereits auf umfangreiches Erfahrungswissen und ausgereifte Technologien aufgesetzt werden. Dies ist vor allem für die Wirtschaftsinformatik als Wissenschaft von Bedeutung, da die Forschung in diesem Feld nicht mehr auf Spekulationen zum Bedarf sowie auf Projektionen technologischer Grundlagenforschung angewiesen ist, sondern reale Lösungskonzepte für ebenso reale Probleme liefern kann. Mit dem weiter andauernden Wachstum des BI-Bereichs sowie dem zunehmenden Bedürfnis zur Erschließung neuer Anwendungsbereiche gehen gleichzeitig auch neue Forschungsprobleme einher. Es besteht von daher weiterhin Bedarf für die Exploration innovativer Lösungen. Diese zwei Gesichter des BI-Bereichs – realer Bedarf und innovative Herausforderungen – lassen auch in Zukunft Raum für spannende Forschungsvorhaben.

An dieser Stelle sei daher allen Autoren für ihre Beiträge und ihr Engagement bei der Veranstaltung gedankt. Ebenfalls richtet sich unser Dank an die weiteren beteiligten Gutachter, Herr Prof. Dr. Peter Gluchowski, Herr Prof. Dr. Hans-Georg Kemper sowie Herr Prof. Dr. Richard Lackes. Die ausführlichen und konstruktiven Gutachten haben unseres Erachtens wesentlich zur hohen Qualität der Veranstaltung beigetragen. Unser Dank richtet sich weiterhin an Frau Stefanie Wenk und Herrn Onur Bazarkaya für die redaktionelle Unterstützung. Schließlich möchten wir uns bei der Universität Dortmund für die Bereitstellung des Tagungshotels bedanken, hier insbesondere bei Herrn Prof. Lackes für die Organisation sowie Herrn Wolfgramm für die umfassende Unterstützung vor Ort.

Stuttgart und Osnabrück, im November 2009

Henning Baars und Bodo Rieger

## Inhalt

Sebastian Neuhaus: Konzeption einer Referenzarchitektur für Operational Business Intelligence auf Basis einer ereignisgesteuerten, serviceorientierten Architektur	1
Christian Schultewolter: Konzeptuelle Modellierung für modellgetriebene Decision Support Systeme	15
Stephan König: Ein Wiki-basiertes Vorgehensmodell für Business Intelligence Projekte	33
Matthias Mertens, Yvette Teiken, Jürgen Appelrath: Semantische Anreicherung von strukturierten Daten und Prozessen in analytischen Informationssystemen am Beispiel von MUSTANG	53
Andreas Hilbert, Alexander E. Müller: Business Intelligence bei Entscheidungen im Human Resource Management	69
Christian Kurze, Peter Gluchowski: Computer Aided Warehouse Engineering auf Basis der Model Driven Architecture	85
Ralf Knackstedt, Łukasz Lis: Erweiterungspotenziale für die Modellierung mit „H2 for Reporting“ am Beispiel von Forschungsportalen	99
Andreas Hilbert, Andreas Schieber: Generierung von neuartigem Wissen durch die Analyse von Weblogs Vorstellung eines Forschungsvorhabens	117
Walid Mehanna: Rahmenkonzept für das Management unternehmensindividueller BI-Architekturen	131
Peter Mertens: Führungsinformationssysteme für Kontrollorgane - Paradigmenwechsel in der MIS-Lehre?	149
Andreas Hilbert, Stefan Sommer: Multi-Channel-Beziehungsmanagement für den elektronischen Kaufprozess	163

# **Konzeption einer Referenzarchitektur für Operational Business Intelligence auf Basis einer ereignisgesteuerten, serviceorientierten Architektur**

Sebastian Neuhaus

Universität Duisburg-Essen, Campus Duisburg

Mercator School of Management - Fakultät für Betriebswirtschaftslehre

Lehrstuhl für Wirtschaftsinformatik und Operations Research

## **1 Ausgangssituation und Problemstellung**

Unter dem Begriff *Business Intelligence* (BI) wird ein integrierter, IT-basierter Ansatz zur betrieblichen Entscheidungsunterstützung verstanden, der mit Hilfe von unterschiedlichen Konzepten und Technologien umgesetzt werden kann (Kemper, Mehanna, & Unger 2006). Ziel von Business Intelligence ist eine entscheidungsorientierte Sammlung und Aufbereitung von internen und externen Daten eines Unternehmens mit dem Ziel geschäftsrelevante Informationen für Analyse-, Planungs- und Steuerungszwecke zur Verfügung zu stellen. Data-Warehouse(DWH)-Systeme stellen dabei eine technische Umsetzung von Business Intelligence dar. Es handelt sich um Lösungen, die insbesondere auf die strategische Entscheidungsfindung durch eine eingeschränkte Zielgruppe von Fach- und Führungskräften ausgerichtet sind. Auf Grundlage von überwiegend retrospektiven Analysen wird Führungskräften eine Hilfestellung für strategische Unternehmensentscheidungen gegeben (Gluchowski, Gabriel, & Dittmar 2008).

Eine Unterstützung für hierarchisch niedrige, operative Ebenen ist bei den klassischen DWH-Ansätzen nicht vorgesehen (Eckerson 2007). Auf der Ebene operativer Geschäftsprozesse werden BI-Systeme und analytische Informationen bislang selten eingesetzt (Steria Mummert Consulting 2009). Eine Ausweitung des Einsatzfeldes von BI zur Steuerung operativer Geschäftsprozesse verspricht große Potenziale. Das Zusammenspiel von operativen und analytischen Informationssystemen kann bei richtigem Einsatz Prozesskosten reduzieren, die Prozessqualität und -agilität verbessern sowie Prozessdurchlaufzeiten verringern, also die Effektivität und Effizienz von operativen Geschäftsprozessen erhöhen (Bucher & Dinter 2008a, Bucher 2008). Diese Ausweitung von klassischer BI auf die operative Geschäftsebene wird aktuell unter dem Begriff Operational Business Intelligence (OpBI) in Wissenschaft und Praxis diskutiert.

Unterschiedliche wissenschaftliche Arbeiten und die betriebliche Praxis zeigen die zunehmende Notwendigkeit für Unternehmen auf, das Einsatzfeld von klassischen BI-Lösungen auf die operative Geschäftsebene im Sinne des Operational-Business-Intelligence-Ansatzes zu erweitern (vgl. Gluchowski, Kemper, & Seufert 2009, Steria Mummert Consulting 2009). Es soll somit eine Verknüpfung von operativen und analytischen Informationssystemen erreicht werden. Diese Verknüpfung bisher isoliert betrachteter Informationssysteme erfordert Anpassungen auf betriebswirtschaftlicher und informationstechnischer Ebene.

## 2 Zielsetzung der Arbeit

Derzeit existiert nur eine geringe Anzahl an wissenschaftlichen Beiträgen zu dem Thema *Operational Business Intelligence*. Eine Begriffsklärung und -abgrenzung für den Begriff *Operational Business Intelligence* findet in diesen Beiträgen nicht statt. Es existieren in diesem Themenbereich Begrifflichkeiten wie *prozessorientierte Informationslogistik*, *Process-centric Business Intelligence* oder *Process Driven Business Intelligence*, die von verschiedenen Autoren genannt werden (Bucher 2008, Bucher, Gericke, & Sigg 2009, Gassmann 2007). Im Rahmen der Dissertation soll daher zunächst eine wissenschaftliche Grundlage in Form einer Charakterisierung des Begriffs erarbeitet werden. Neben einer Analyse der wissenschaftlichen Literatur wird auch das Begriffsverständnis von Herstellern und Beratern betrachtet.

Ausgehend von den informationstechnischen Anforderungen, die aus Sicht von Wissenschaft und Praxis an integrierte Operational-Business-Intelligence-Systeme gestellt werden, soll im Rahmen der Dissertation eine Referenzarchitektur<sup>1</sup> für entsprechende Systeme konzipiert werden. Diese Referenzarchitektur beschreibt die benötigten informationstechnischen Komponenten einer Operational-Business-Intelligence-Lösung, deren Aufgaben und Funktionen sowie das Zusammenspiel der einzelnen Komponenten. Aus wissenschaftlicher Sicht liefert eine solche Referenzarchitektur die strukturierte Grundlage für weiterführende Untersuchungen. In der Praxis ermöglicht eine Referenzarchitektur die Analyse bereits im Einsatz befindlicher Informationssysteme im Hinblick auf die Einführung eines OpBI-Systems sowie die Analyse und den Vergleich entsprechender OpBI-Lösungen von verschiedenen Herstellern und Anbietern.

---

<sup>1</sup> Bei einer Referenzarchitektur handelt es sich um eine IT-Architektur, die vereinheitlichend für die IT-Architektur einer Gruppe von Informationssystemen steht. Eine Referenzarchitektur unterscheidet sich dabei insbesondere anhand ihres Abstraktionsniveaus und ihres Verwendungszwecks von einer herkömmlichen IT-Architektur (Dem 2006).



Zur Evaluation der konzipierten Referenzarchitektur wird im Rahmen eines betriebswirtschaftlichen Anwendungsszenarios eine Gestaltungsempfehlung für ein OpBI-System zur Entscheidungsunterstützung im operativen Logistikmanagement erarbeitet.

### 3 Wissenschaftstheoretische Einordnung

Die Wirtschaftsinformatik (WI), als Schnittstellendisziplin zwischen Betriebswirtschaftslehre und Informatik, bedient sich im Rahmen der Forschungsaktivitäten Instrumenten aus der Real-, Formal- als auch Ingenieurwissenschaft (WKWI 1994). Der Schwerpunkt der Wirtschaftsinformatik liegt dabei im konstruktiv-qualitativen Bereich. Das Methodenspektrum ist dabei recht gering und wird zu 91% durch sechs Kernmethoden abgedeckt: argumentativ-, konzeptionell- und formal-deduktive Analyse, Prototyping, Fallstudie und quantitative Querschnittsanalyse (Wilde & Hess 2007).

Die Dissertation beruht auf dem Paradigma des Konstruktivismus (Design Science). Das Ziel des konstruktionswissenschaftlichen Paradigmas ist es dabei, durch die Erstellung und Evaluation von Modellen, Methoden oder Systemen, nützliche IT-Systeme zu entwickeln (Wilde & Hess 2007). Als Methode liegt der Arbeit die argumentativ-deduktive Analyse zu Grunde. Basis der logisch-deduktiven Vorgehensweise sind Grundsätze und Denkmodelle, die auf logischen und existentiellen Gesetzmäßigkeiten beruhen. Durch logische Schlussfolgerungen wird das Verständnis des Besonderen und Individuellen erreicht. Dies geschieht dabei auf Grundlage bestimmter Bedingungen, die durch Erfahrungen oder gedankliches Schließen vorhanden sind (Thomas 2006). Die Schlussfolgerung erfolgt dabei argumentativ, also rein sprachlich (Wilde & Hess 2007).

### 4 Aufbau der Arbeit

Zu Beginn der Arbeit erfolgt eine eingehende Bearbeitung der Themenbereiche *Business Intelligence* und *Operational Business Intelligence* aus betriebswirtschaftlicher Sicht. Über Kategorisierungsmerkmale wird eine definitorische Abgrenzung der essentiellen Begriffe vorgenommen. Weiterhin werden verwandte Konzepte und Begrifflichkeiten, die innerhalb der wissenschaftlichen Literatur als auch in der Praxis im Kontext von OpBI genannt werden, erläutert, abgegrenzt und eingeordnet. Es handelt sich dabei insbesondere um die Konzepte *Business Activity Monitoring* (BAM) und *Process Performance Management* (PPM). Im Anschluss werden die betriebswirtschaftlichen Anforderungen an ein OpBI-System identifiziert.

In einem zweiten Schritt erfolgt eine Darstellung und Untersuchung von informationstechnischen Konzepten und Ansätzen aus dem Bereich der analytischen und operativen Informationssysteme im Hinblick auf eine mögliche Verknüpfung dieser Ansätze. Aus

dem Kontext der operativen Informationssysteme werden die Konzepte der serviceorientierten Architektur (SOA) und der ereignisgesteuerten Architektur (EDA) einer Analyse unterzogen. Diese Konzepte stellen in Kombination die Grundlage für die zu konzipierende Referenzarchitektur dar. Aus dem Bereich der klassischen Business Intelligence werden insbesondere Ansätze zur echtzeitnahen Datenversorgung (Real-time Data Warehousing<sup>2</sup>) sowie zur Automatisierung von Entscheidungen (Active Data Warehousing<sup>3</sup>) betrachtet.

Auf Grundlage der erarbeiteten betriebswirtschaftlichen Anforderungen erfolgt anschließend eine Analyse der informationstechnischen Konzepte und Ansätze für die Konzeption der Referenzarchitektur. Als Ergebnis werden die konzipierte Referenzarchitektur für Operational Business Intelligence vorgestellt und die Aufgaben und Funktionen sowie das Zusammenspiel der einzelnen Architekturkomponenten erläutert.

Eine Evaluation der entwickelten Referenzarchitektur erfolgt im Rahmen eines betriebswirtschaftlichen Anwendungsszenarios aus dem Supply Chain Management (SCM). Dazu wird ein allgemeines, branchenübergreifendes Szenario unter Berücksichtigung aktueller Trends in der Logistik erarbeitet. Neben den Grundlagen des SCM und des Supply Chain Performance Management (SCPM) werden auch die Technologien *Radio Frequency Identification* (RFID) und *EPCglobal Network* dargestellt. Auf dieser Grundlage erfolgt die Abgrenzung und Modellierung des eigentlichen Anwendungsszenarios im Kontext eines Lead Logistics Provider (LLP). Als Ergebnis der Evaluation wird eine Gestaltungsempfehlung für ein Operational-Business-Intelligence-System im Supply Chain Management aufgezeigt.

## 5 Aktueller Stand der Arbeit

Der bisherige Schwerpunkt der Forschungstätigkeiten liegt auf der Definition, Charakterisierung und Abgrenzung des Begriffs *Operational Business Intelligence*. Zur Einordnung und Systematisierung der Abgrenzung wird ein morphologischer Kasten dargestellt.

Die technische Grundlage der Referenzarchitektur soll die Kombination der Ansätze der serviceorientierten Architektur und der ereignisgesteuerten Architektur in Form einer ereignisgesteuerten, serviceorientierten Architektur (EDSOA) darstellen. Zum Verständnis einer EDSOA werden die essentiellen Inhalte strukturiert und erläutert. Des Wei-

---

<sup>2</sup> Für weitere Erläuterungen wird auf (Schelp 2006) verwiesen.

<sup>3</sup> Für weitere Erläuterungen wird auf (Thalhammer 2001) verwiesen.

teren wird herausgearbeitet, warum eine Verknüpfung von SOA und EDA einen Mehrwert verspricht.

### 5.1 Operational Business Intelligence

Im Rahmen der Dissertation wird zunächst eine Definition und Charakterisierung des Begriffs *Operational Business Intelligence* durchgeführt sowie eine Abgrenzung zu verwandten Konzepten und Ansätzen vorgenommen.

Innerhalb der derzeit verfügbaren Literatur zu dem Thema *Operational Business Intelligence* setzen die Autoren unterschiedliche Schwerpunkte in ihrer Sichtweise. Tabelle 1 zeigt verschiedene Definitionen zu den Begriffen *Operational Business Intelligence* und *Process-centric Business Intelligence*<sup>4</sup> auf.

Nach BLASUM geht es bei Operational BI um die Steuerung und Optimierung von Geschäftsprozessen unter Zuhilfenahme von Business-Intelligence-Methoden (Blasum 2006). ECKERSON liefert eine weitere praxisorientierte Definition des Begriffes *OpBI* und spiegelt damit die Sichtweise des TDWI (The Data Warehousing Institute) wider. ECKERSON stellt in seinen weiteren Ausführungen insbesondere die Verschmelzung von operativen und analytischen Prozessen zu einem Ganzen heraus und sieht den erhöhten Nutzen von OpBI unter anderem in der möglichen Ausweitung des Nutzerkreises auf das operative Management (Eckerson 2007). Zusätzlich thematisieren BLASUM und ECKERSON den deutlich engeren Zeitrahmen für Analysen im Kontext operativer Geschäftsprozesse, der differenzierte Anforderungen an die umzusetzenden Komponenten und Informationssysteme stellt (Blasum 2006, Eckerson 2007).

Quelle	Definition
Blasum 2006	Operational BI employs business intelligence methods to manage and optimize daily business operations.
Bucher, Gericke, & Sigg 2009	[...] we define “process-centric business intelligence” [PCBI] as all BI capabilities that are dedicated to the analysis as well as to the systematic purposeful transformation of business-relevant data into analytic information and that have been, at the same time, embedded into an operational process. [...] PCBI is therefore used to support operational decision-making in the context of process execution. [...] PCBI provides support for process execution by embedding BI capabilities and/or pre-processed analytic information into the process context at process runtime.

<sup>4</sup> Process-Centric Business Intelligence stellt die englische Bezeichnung für prozessorientierte Informationslogistik dar.

Chamoni 2007	Operational BI analysiert Daten von Geschäftsprozessen, verknüpft diese mit weiteren Informationen und stellt die generierten Ergebnisse und Handlungsvorschläge kurzfristig und automatisiert einer großen Zahl von Anwendern zur Verfügung, und zwar mit dem Ziel, die Geschäftsprozesse steuern und optimieren zu können.
Eckerson 2007	Operational BI delivers information and insights to a broad range of users within hours or minutes for the purpose of managing and optimizing operational or time sensitive business processes.
Gluchowski, Kemper, & Seufert 2009	Operational BI bezeichnet integrierte geschäftsprozessorientierte Systeme, die mithilfe klassischer Business-Intelligence-Methoden auf der Basis zeitnaher, prozessualer Ablaufdaten und in aller Regel auch historischer, harmonisierter Daten eine Realtime- (Near-Realtime-) Unterstützung für zeitkritische Entscheidungen während des Prozessablaufes bieten.

*Tabelle 1: Definitionen der Begriffe Operational Business Intelligence und Process-centric Business Intelligence*

Auch GLUCHOWSKI/KEMPER/SEUFERT weisen auf die Anforderung einer echtzeitnahen Entscheidungsunterstützung im Prozessablauf hin, die durch integrierte, geschäftsprozessorientierte Informationssysteme ermöglicht werden soll. Als Datenbasis dienen dabei sowohl Prozessablaufdaten als auch historische, analytische Daten (Gluchowski, Kemper, & Seufert 2009).

Insgesamt ist OpBI von klassischer BI auf strategischer und taktischer Ebene klar zu unterscheiden. Klassische Business Intelligence erlaubt in der Regel lediglich retrospektive Analysen im Kontext der strategischen und taktischen Unternehmensführung und ist auf Fach- und Führungskräfte ausgerichtet. Zumeist weisen die Analysen und Reports Latenzen von Tagen, Wochen oder Monaten auf (Gluchowski, Gabriel, & Dittmar 2008).

Im Rahmen der Dissertation wird unter dem Konzept *Operational Business Intelligence* eine Entscheidungsunterstützung auf operativer Ebene zur schnellen und flexiblen Ausführung und Steuerung von wertschöpfenden Prozessen in Echtzeit (oder echtzeitnah) durch die Verknüpfung von zeitnahen Transaktionsdaten und analytischen Informationen verstanden. Wertschöpfende Prozesse umfassen dabei sowohl primäre als auch sekundäre Geschäftsprozesse. Nutzertypen von Operational Business Intelligence stellen insbesondere Mitarbeiter auf operativer Ebene dar. OpBI umfasst neben der reinen Prozessausführung auch die Prozesssteuerung und -überwachung, so dass der Ansatz des Business

Activity Monitoring<sup>5</sup> (BAM) einen Bestandteil von OpBI darstellt, während die Aufgaben des Process Performance Management<sup>6</sup> (PPM) nicht dem OpBI-Konzept zugeordnet werden können.

Zur Umsetzung des Operational-Business-Intelligence-Konzepts können integrierte, geschäftsprozessorientierte Informationssysteme eingesetzt werden, die Funktionalitäten und Daten aus operativen als auch analytischen Informationssystemen verknüpfen. Dadurch wird eine verbesserte IT-Unterstützung der operativen Prozesse eines Unternehmens ermöglicht.

In Abbildung 1 ist ein morphologischer Kasten dargestellt, der die Einordnung und Abgrenzung des OpBI-Konzepts ermöglicht und die wesentlichen Charakteristika aufzeigt.

Mit Process-centric Business Intelligence definieren BUCHER/GERICKE/SIGG sowie BUCHER/DINTER einen zu OpBI verwandten Begriff. Unter PCBI verstehen die Autoren Funktionalitäten (Datenanalyse und Informationsbereitstellung) zur Entscheidungsunterstützung im Zusammenhang mit der Ausführung unternehmerischer, insbesondere operativer Prozesse. Dabei wird explizit die Verknüpfung von analytischen Informationen und betrieblichen Prozessen betont. Weiterhin wird herausgestellt, dass ausschließlich prozessinhärente Entscheidungen unterstützt werden, die durch einen menschlichen Akteur getroffen werden müssen (Bucher & Dinter 2008a, Bucher, Gericke, & Sigg 2009). PCBI lässt sich damit im Sinne dieser Arbeit durch den Fokus auf die reine Prozessausführung von dem Begriff *Operational Business Intelligence* abgrenzen. Business Activity Monitoring fällt demnach nicht unter den Begriff *PCBI*. Somit kann PCBI als ein Teilgebiet von Operational Business Intelligence interpretiert werden.

---

<sup>5</sup> Business Activity Monitoring bietet eine Entscheidungsunterstützung für das operative Management durch die Überwachung und Analyse kritischer Geschäftsereignisse und Prozesskennzahlen in Echtzeit. Mit Hilfe von BAM sollen Durchlaufzeiten verkürzt und die Effektivität operativer Prozesse verbessert werden (Heß 2005, Oehler 2006, Dinter & Bucher 2006).

<sup>6</sup> Beim Process Performance Management liegt der Fokus in der dauerhaften, automatisierten Analyse von Prozessstrukturen sowie der Möglichkeit eines Soll-Ist-Vergleichs von Prozesskennzahlen, um Optimierungspotenziale zu identifizieren und die Geschäftsprozesse kontinuierlich anzupassen und zu verbessern (Heß 2005, Dinter & Bucher 2006, Gluchowski, Gabriel, & Dittmar 2008).

Managementebene	Strategisch		Taktisch		Operativ	
Nutzertypen	Manager		Analysten		Abteilungsleiter	
Nutzerkreis	Klein		Mittel		Groß	
Prozesstyp	Managementprozess		Unterstützungsprozess		Geschäftsprozess	
Phasen des Prozessmanagements	Identifikation, Definition und Modellierung		Implementierung und Ausführung		Überwachung und Steuerung	
Prozessausführung	Manuell		Teilautomatisch		Vollautomatisch	
Strukturierungsgrad der Problemstellung	Unstrukturiert		Semi-strukturiert		Strukturiert	
Zeitraum für Analysen	Monat	Woche	Tag	Stunde	Minute	
Zeitbezug der Daten	Historisch			Aktuell		
Verfügbarkeit der Daten	Unkritisch			Kritisch		
Herkunft der Daten	Interne Quellen			Externe Quellen		
Strukturierungsgrad der Daten	Unstrukturierte Daten			Strukturierte Daten		
Art der Daten	Analytische Daten			Operative Daten		
Entscheidungs-latenz	Hoch		Mittel		Gering	

Abbildung 1: Charakterisierung von Operational Business Intelligence

Zur Umsetzung von Operational Business Intelligence müssen die Informations-systeminfrastruktur und Informationssysteme eines Unternehmens neuen Anforderungen gerecht werden. Neben einer erhöhten Flexibilität und Agilität der Informationssystem-infrastruktur ist auch eine Verkürzung der Latenzzeiten in der Informationsversorgung sowie eine Verknüpfung analytischer und operativer Informationen während des Prozess-ablaufes notwendig. Diese Anforderungen können klassische BI-Systeme nicht erfüllen. Im Rahmen der Literaturanalyse konnten aktuelle Technologien und Konzepte ermittelt werden, die im Kontext des Begriffs *Operational Business Intelligence* in Wissenschaft und Praxis genannt werden. Insbesondere die Ansätze der serviceorientierten Architektur und der ereignisgesteuerten Architektur konnten dabei als potenzielle Enabler für Opera-tional-Business-Intelligence-Systeme identifiziert werden. Auf Grundlage dieser Konzep-te soll eine Verschmelzung der operativen und analytischen Informationssysteme, im Sinne von Operational Business Intelligence, erreicht werden.

Nutzenpotenziale von OpBI wurden unter anderem von BUCHER/DINTER durch Exper-teninterviews und eine empirische Untersuchung erhoben. Es kann zwischen Nutzenef-fekten innerhalb und außerhalb des Unternehmens unterschieden werden. Die internen Nutzeneffekte (z. B. Verbesserung der Prozessleistung und -qualität, Beschleunigung der Prozessausführung, Steigerung der Mitarbeiterzufriedenheit mit dem Prozessablauf) wer-den dabei als wichtiger eingeschätzt als die externen Nutzeneffekte (z. B. Steigerung der

Kundenprofitabilität, -zufriedenheit und -loyalität) (Bucher & Dinter 2008a, Bucher & Dinter 2008b, Bucher, Gericke, & Sigg 2009).

## 5.2 Ereignisgesteuerte, serviceorientierte Architektur

Wie in Abschnitt 5.1 dargestellt werden aktuell die Ansätze der serviceorientierten und der ereignisgesteuerten Architektur im Kontext von Operational Business Intelligence diskutiert. Daher werden zunächst die Begriffe *SOA* und *EDA* definiert und erläutert. Im weiteren Verlauf wird dann herausgearbeitet, warum eine Verknüpfung von *SOA* und *EDA* zu einer ereignisgesteuerten, serviceorientierten Architektur einen Mehrwert verspricht und als Basis zukünftiger OpBI-Systeme genutzt werden kann.

Das Konzept der *SOA*<sup>7</sup> stellt die Grundlage für eine flexible, auf Prozesse ausgerichtete, verteilte Informationssystemarchitektur/-infrastruktur auf Basis von plattform-unabhängigen und lose miteinander gekoppelten Services (Diensten) dar. Aus technischer Sicht kapseln Services Funktionen und Daten bestehender Informationssysteme und stellen diese über standardisierte Schnittstellen in einem Netzwerk zur Verfügung. Services sind gegeneinander klar abgegrenzt, wiederverwendbar und abstrahieren von der Komplexität der zugrundeliegenden meist heterogenen Informationssysteminfrastruktur, wie z. B. Kommunikationsprotokollen, Datenformaten etc. Aus fachlicher Sicht bieten Services Funktionalitäten an, die explizit auf die Anforderungen der eigentlichen Geschäftsprozesse ausgerichtet sind (Gioldasis et al. 2003, Tilkov & Starke 2007, Heutschi 2007). Diese Services können entsprechend der geschäftlichen Anforderungen orchestriert werden, um damit komplexe Geschäftsprozesse über die Grenzen einzelner Informationssysteme hinweg IT-technisch zu unterstützen.

Eine *SOA* kann die Flexibilität und Agilität eines Unternehmens auf Grundlage einer modularen und erweiterbaren Informationssysteminfrastruktur erhöhen. Neben der Flexibilität und Agilität stellt heute das Management die Forderung effektiv und schnell auf neue Gegebenheiten und Herausforderungen zu reagieren. In diesem Kontext bezeichnen Ereignisse jegliche Art von Veränderung in der Umwelt eines Unternehmens, wie z. B. Bestandsänderungen, Absatzmengenwachse, Kursschwankungen oder die Insolvenz eines Lieferanten (Schulte 2004, Zacharias 2007). Den beschriebenen Anforderungen kann eine ereignisgesteuerte Architektur Rechnung tragen.

Bei einer *EDA* handelt es sich um ein Konzept zur Entwicklung von Informationssystemlandschaften, das auf die Ereignisverarbeitung ausgerichtet ist. Dabei stehen das

---

<sup>7</sup> Die wesentlichen Charakteristika und Designprinzipien, die einer *SOA* zu Grunde liegen, können bei (Heutschi 2007) nachgelesen werden.

Erzeugen, Entdecken und Verarbeiten einzelner Ereignisse oder Ereignisströme innerhalb und außerhalb eines Unternehmens im Vordergrund (Dunkel et al. 2008). Die Hauptanforderung an eine EDA stellt die schnelle und effiziente Bearbeitung von Ereignissen dar. Dabei kann es sich um interne und externe, erwartete und unerwartete oder atomare und komplexe (zusammengesetzte) Ereignisse handeln (Luckham 2002, Rommelspacher 2008).

Separat betrachtet stellen die serviceorientierte Architektur und die ereignisgesteuerte Architektur zwei orthogonale Ansätze dar, die unabhängig voneinander implementiert werden können. Das wichtigste Differenzierungsmerkmal ist dabei der Grad der Kopplung zwischen den Kommunikationsteilnehmern. Bei der SOA wird eine lose Kopplung der Kommunikationsteilnehmer in Form des Request-Reply-Prinzips realisiert, während der EDA eine vollständige Entkopplung nach dem Publish-Subscribe-Prinzip zu Grunde liegt, d.h. der Sender und der Empfänger besitzen keinerlei Informationen übereinander (Chappell 2004). Infolgedessen ermöglicht die EDA nicht nur eine SOA-konforme 1:1-Kommunikation zwischen den Partnern, sondern ebenfalls eine 1:n- und n:m-Kommunikation. Darüber hinaus arbeitet eine SOA die Prozesse synchronisiert und sequentiell ab, da der aufgerufene Service über den darauf folgenden logischen Schritt entscheidet und auf dessen Ausführung wartet. Die EDA fungiert dahingehend konträr. Der Sender stellt bei einer EDA ausschließlich einen Informanten dar, der über eine Statusveränderung informiert. In der EDA ist die Entscheidungsgewalt dem Empfänger übertragen, der zwischen verschiedenen Verarbeitungsmöglichkeiten unterscheiden kann, sodass die Prozesse i.d.R. asynchron und parallel ablaufen (Zacharias 2007, Dunkel et al. 2008, Rommelspacher 2008).

Die Kombination beider Ansätze zu einer ereignisgesteuerten, serviceorientierten Architektur beseitigt die Schwächen der jeweiligen Architekturen. Eine EDSOA bietet einen geeigneten Ansatz zum Aufbau einer agilen IT-Infrastruktur und zur Lösung von Integrationsproblemen (Sriraman & Radhakrishnan 2005, Zacharias 2007). Die Geschäftslogik liegt in Form von orchestrierten Services vor, wobei die initiiierenden oder resultierenden Ereignisse einer solchen Prozesskette EDA-Ereignisse darstellen. Zur Realisierung der Kommunikation und Ereignisverarbeitung liefert ein Enterprise Service Bus (ESB) die entsprechenden Dienste. Der ESB unterstützt als zentrale Kommunikationsplattform die ereignisgesteuerte, asynchrone Interaktion zwischen den einzelnen Services. In einer EDSOA auf der Grundlage eines ESB stellen Anwendungen und Services abstrakte Endpunkte dar, die auf asynchrone Ereignisse reagieren können (Chappell 2004, Zacharias 2007, Rommelspacher 2008).

Gemäß den vorangestellten Ausführungen können eine SOA und eine EDA als ein integratives, sich ergänzendes Konzept in Form einer EDSOA implementiert werden. Eine



Kombination dieser Konzepte auf Basis eines ESB steigert den Wert beider Architekturen (Woolf 2006). Der wesentliche Nutzen einer EDSOA besteht darin, dass alle Geschäftsereignisse auf dem ESB vorliegen und in Echtzeit verwendet werden können. Eine Reaktion auf eintretende Ereignisse kann unmittelbar erfolgen, z. B. in der Möglichkeit, mehrere Services parallel und asynchron anzustoßen. Dadurch wird der synchrone und sequentielle SOA-Kontrollfluss klarer und realitätsnaher abgebildet (Zacharias 2007).

## 6 Zusammenfassung und Ausblick

Der bisherige Schwerpunkt der Forschungstätigkeiten lag auf der Definition, Charakterisierung und Abgrenzung des Begriffs *Operational Business Intelligence*. Im Rahmen dieser Tätigkeiten konnte ermittelt werden, dass die Ansätze der SOA und der EDA von einem Großteil der Autoren als Grundlage für eine erfolgreiche OpBI-Implementierung genannt und diskutiert werden. Daher wurde in einem zweiten Schritt ein Begriffsverständnis der SOA und EDA sowie der Verknüpfung zu einer EDSOA erarbeitet.

Die Betrachtung von Ansätzen zur echtzeitnahen Datenversorgung sowie zur Automatisierung von Entscheidungen aus dem Bereich der klassischen Business Intelligence stellt einen nächsten Schritt dar. Es soll analysiert werden, wie diese Ansätze mit einer EDSOA kombiniert werden können.

Weiterhin soll der Praxisfall zur Evaluation der Referenzarchitektur beschrieben und das Potenzial des Einsatzes eines OpBI-System im Umfeld eines Lead Logistics Provider erarbeitet werden.

## 7 Literatur

Blasum, R. (2006). Operational BI. Whitepaper BusinessCoDe. URL: [http://www.business-code.de/cms/uploads/media/BCD\\_Operational\\_BI\\_01.pdf](http://www.business-code.de/cms/uploads/media/BCD_Operational_BI_01.pdf).

Bucher, T. (2008). Interaktionseffekte zwischen prozessorientierter Organisation und Informationslogistik. In: Dinter, B., Winter, R. (Hrsg.). Integrierte Informationslogistik. Springer, Berlin. S. 107-135.

Bucher, T., Dinter, B. (2008a). Anwendungsfälle der Nutzung analytischer Informationen im operativen Kontext. In: Bichler, M., Hess, T., Krcmar, H., Lechner, U., Mattes, F., Picot, A., Speitkamp, B., Wolf, P. (Hrsg.). Tagungsband der Multikonferenz Wirtschaftsinformatik 2008 (MKWI 2008) - München. GITO-Verlag, Berlin. S. 67-179.

Bucher, T., Dinter, B. (2008b). Process orientation of information logistics – an empirical analysis to assess benefits, design factors, and realization approaches. In: Sprague, R. (Hrsg.). Proceedings of the 41th Hawaii International Conference on System Sciences (HICSS'08). IEEE Computer Society Press, Los Alamitos, CA.

Bucher, T., Gericke, A., Sigg, S. (2009). Process-centric business intelligence. In: Business Process Management Journal Vol. 15, No. 3. S. 408-429.

Chamoni, P. (2007). Business Intelligence: von der Innovation zum Mainstream, Data Mining Cup 2007. Leipzig.

Chappell, D. (2004). Enterprise Service Bus. O'Reilly, Sebastopol.

Dern, G. (2006). Management von IT-Architekturen. Leitlinien für die Ausrichtung, Planung und Gestaltung von Informationssystemen. 2. Auflage, Vieweg, Wiesbaden.

Dinter, B.; Bucher, T. (2006). Business Performance Management, In: Chamoni P., Gluchowski, P. (Hrsg.). Analytische Informationssysteme - Business Intelligence-Technologien und -Anwendungen. 3. Auflage, Springer, Berlin. S. 23-50.

Dunkel, J., Eberhart, A., Fischer, S., Kleiner, C., Koschel, A. (2008). Systemarchitekturen für Verteilte Anwendungen. Carl Hanser Verlag, München.

Eckerson, W. (2007). Best Practices in Operational BI. Converging Analytical and Operational Processes. TDWI Best Practices Report, o. O.

Gassmann, B. (2007). Guide to Process-Centric BI Terms. Gartner Research. Number: G00151682. Stamford.

Gioldasis, N., Moumoutzis, N., Kazasis, F., Pappas, N., Christodoulakis, S. (2003). A Service Oriented Architecture for Managing Operational Strategies. In: Jeckle, M., Zhang, L.-J. (Hrsg.). Web Services - ICWS-Europe 2003. International Conference on Web Services - Europe 2003 (ICWS Europe 2003). Lecture Notes in Computer Science, Vol. 2853. Springer, Berlin. S. 11-23.

Gluchowski, P., Gabriel, R., Dittmar, C. (2008). Management Support Systeme und Business Intelligence. Computergestützte Informationssysteme für Fach- und Führungskräfte. 2. Auflage, Springer, Berlin.

Gluchowski, P., Kemper, H.-G., Seufert, A. (2009). Innovative Prozess-Steuerung. In: BI-Spektrum Ausgabe 1, 4. Jahrgang. S. 8-12.

Heß, H. (2005). Von der Unternehmensstrategie zur Prozess-Performance - Was kommt nach Business Intelligence? In: Scheer, A.-W., Jost, W., Heß, H., Kronz, A. (Hrsg.). Corporate Performance Management - ARIS in der Praxis. Springer, Berlin. S. 7-29.

Heutschi, R. (2007). Serviceorientierte Architektur - Architekturprinzipien und Umsetzung in der Praxis. Springer, Berlin.

Kemper, H.-G., Mehanna, W., Unger, C. (2006). Business Intelligence - Grundlagen und praktische Anwendungen. Eine Einführung in die IT-basierte Managementunterstützung. 2. Auflage. Vieweg, Wiesbaden.

Luckham, D. (2002). The Power of Events: An introduction to Complex Event Processing in Distributed Enterprise Systems. Addison-Wesley, Boston et al.

Oehler, K. (2006). Corporate Performance Management mit Business Intelligence Werkzeugen. Carl Hanser Verlag, München.

Rommelspacher, J. (2008). Ereignisgetriebene Architekturen. In: WIRTSCHAFTS-INFORMATIK 50 (2008) Nr. 4, S. 314-317.

Schelp, J. (2006). „Real“-Time Warehousing und EAI. In: Chamoni, P., Gluchowski, P. (Hrsg.). Analytische Informationssysteme. 3. Auflage. Springer, Berlin. S. 425-438.

Schulte, R. (2004). Using Events for Business Benefit. In: Business Integration Journal (2004) 5. S. 43-45.

Sriraman, B., Radhakrishnan, R. (2005). Event Driven Architecture Augmenting Service Oriented Architectures. URL: <http://soa.omg.org/Uploaded%20Docs/EDA/edamdaso.pdf>.

Steria Mummert Consulting AG (2009). biMA 2009. Business Intelligence-Studie 2009. Hamburg.

Thalhammer, T. (2001). Active Data Warehouses - Complementing OLAP with Analysis Rules. Dissertation. Johannes Kepler Universität Linz, Linz.

Thomas, O. (2006). Management von Referenzmodellen: Entwurf und Realisierung eines Informationssystems zur Entwicklung und Anwendung von Referenzmodellen. Logos, Berlin.

Tilkov, S., Starke, G. (2007). Einmaleins der serviceorientierten Architekturen. In: Starke, G., Tilkov, A. (Hrsg.) SOA-Expertenwissen. dpunkt-verlag, Heidelberg. S. 9-39.

Wilde, T., Hess, T. (2007). Forschungsmethoden der Wirtschaftsinformatik. Eine empirische Untersuchung. In: WIRTSCHAFTSINFORMATIK 49 (2007) Nr. 4, S. 280-287.

WKWI (1994). Profil der Wirtschaftsinformatik. Ausführungen der Wissenschaftlichen Kommission der Wirtschaftsinformatik. In: WIRTSCHAFTSINFORMATIK 36 (1994) Nr. 1, S. 80-81.

Woolf, B. (2006). Event-Driven Architecture and Service-oriented Architecture. URL: [http://www.haifa.il.ibm.com/Workshops/oopsla2006/present/w06\\_eda\\_woolf.pdf](http://www.haifa.il.ibm.com/Workshops/oopsla2006/present/w06_eda_woolf.pdf).

Zacharias, R. (2007). SOA & Event Driven Architecture (EDA) - Eine perfekte Symbiose. In: JavaMagazin (2007) 7. S. 60-69.



# Konzeptuelle Modellierung für modellgetriebene Decision Support Systeme

Christian Schultewolter

Universitaet Osnabrueck – Institut fuer Informationsmanagement und

Unternehmensfuehrung (IMU)

FB 9 - BWL/Management Support und Wirtschaftsinformatik

## 1 Motivation

### 1.1 Problemstellung

Entscheidungsunterstützungs- oder Decision Support Systeme (DSS) sind nach Alter (1996) ursprünglich computer-basierte Systeme, die den Benutzer bei der Identifikation und Auswahl geeigneter Entscheidungsalternativen in Situationen, zu denen keine Erfahrungen oder praktischen Lösungsansätze vorliegen, unterstützen. Die Auffassung der *Special Interest Group on Decision Support, Knowledge and Data Management Systems* der *Association for Information Systems* aus dem angloamerikanischen Raum DSS in die Klassen kommunikations-, daten-, dokumenten-, wissens- und modellgetriebene Systeme zu unterscheiden, findet heute im wissenschaftlichen Raum weltweit Unterstützung (Turban, Aronson, Liang & Sharda, 2007) und soll auch diesem Forschungsvorhaben zugrunde liegen. Modellgetriebene Decision Support Systeme (MDSS) werden oft auch synonym Spreadsheet-basierte DSS genannt, da Spreadsheet Pakete, wie z.B. Microsoft Excel, oft für die Implementierung solcher Systeme herangezogen werden (Turban et al., 2007; Holsapple und Whinston, 1996; Power und Sharda, 2007).

Ende der 1990er Jahre und zu Anfang des neuen Jahrtausends hatten Daten-getriebene Systeme den ursprünglich essentiellen Teil der Modelle größtenteils aus dem Fokus der Betrachtungen verdrängt. In jüngerer Vergangenheit lässt sich jedoch beobachten, dass die Komponente *Modell* wieder verstärkt Einfluss auf Entscheidungsprozesse, die durch DSS unterstützt werden, nimmt. Vor allem die Design- und Choice-Phasen nach Simon (1977) werden von solchen Systemen adressiert.

Panko und Sprague (1998) haben eine empirische Studie über Fehler in Spreadsheet-Modellen durchgeführt, die auch für modellgetriebene DSS relevant ist und später durch

Powell et al. überarbeitet wurde (Powell, Baker & Lawson, 2008). Die Ergebnisse zeigen, dass das Auftreten von Fehlern häufig auf eine logisch-falsche Modellierung zurückgeführt werden kann. Diese hat ihren Ursprung nicht ausschließlich in der Unkenntnis der speziellen Problemsituation und ihrer Entstehung, sondern wird durch eine einschränkende, oft wenig intuitive Modellierungskomponente forciert. Ist beispielsweise die Repräsentation expliziter Multidimensionalität nicht möglich, müssen in den meisten Fällen redundante Formeln für jegliche Ausprägungen spezifiziert werden, auf die der logisch repräsentierte Zusammenhang Anwendung finden soll. Dies mündet in einem erhöhten Risiko von Fehlern bei der Wartung, Pflege und Anwendung des Modells. Überdies hinaus können auch fehlendes Werkzeug-Wissen und unnatürliche Modellierungskonzepte die Fehlerrate negativ beeinflussen.

Analog zur Datenmodellierung könnte eine konzeptuelle Modellierungsebene Unabhängigkeit von Werkzeug-spezifischen Konzepten und Einschränkungen bieten sowie eine möglichst natürliche Modellierung ermöglichen. Ein positiver Nebeneffekt zeigt sich hier in der Berücksichtigung der klassischen DSS-Forderung die Distanz zwischen Entwickler und Entscheidungsträger zu minimieren (Sprague und Watson, 1996; Turban et al., 2007).

Wand und Weber (2002) beschreiben in ihrer Arbeit ein Framework für die weitere Erforschung von konzeptuellen Modellierungsansätzen. Es besteht aus einer Grammatik, einer Methode, einem Skript und dem Kontext, jeweils bezogen auf konzeptuelle Modellierung. Die Grammatik liefert Konstrukte und Regeln, welche die Vorschriften bei der Kombination von Konstrukten angeben, um reale Zusammenhänge zu modellieren (Wand und Weber, 2002). Sie wird in diesem Kontext durch die Klassifizierungskriterien und deren Kombination zu einer generischen Spezifikation für MDSS repräsentiert. Die Methode beschreibt wie reale Zusammenhänge durch Instanzen der Grammatik modelliert werden können (Modellierungssprache), das Skript stellt ein finales Dokument dar, welches die modellierten Zusammenhänge in Form der Sprache, erzeugt durch die Grammatik, repräsentiert. Der Kontext beschreibt die Begebenheiten und Situationen, in denen konzeptuelle Modellierung herangezogen wird.

Die Entwicklung einer konzeptuellen Modellierungsebene für MDSS wird des Weiteren durch eine nicht-prozedurale Modellspezifikation (Mayer, 1998), die als Charakteristikum eines DSS genannt wird, unterstützt. Ein konzeptueller Ansatz kann bei der Transformation nicht-prozedural modellierter Zusammenhänge in korrekte, prozedurale Rechenschritte helfen, indem er konsistente Kriterien für die Klassifizierung verschiedener Beziehungstypen zur Verfügung stellt. Er trägt dadurch zum Entstehen eines Frameworks bei, das während des Transformationsprozesses als Checkliste herangezogen werden kann.

Um einen Eindruck des Anwendungsbereiches zu erhalten, wird im weiteren Verlauf des Artikels eine Übersicht über typische Eigenschaften modellgetriebener DSS gegeben, für die die konzeptuelle Ebene entworfen werden soll. Nachdem die Zielsetzung konkretisiert worden ist, werden drei Klassifizierungskriterien vorgeschlagen, die eine systematische Repräsentation von Beziehungen innerhalb eines Modells ermöglichen. Im Anschluss untersucht Kapitel vier bekannte Modellierungsansätze aus verwandten Themengebieten, um eine Eignung dieser für die Umsetzung einer konzeptuellen Modellierungssprache darzulegen. Schließlich wird ein kurzer Ausblick über mögliche Szenarien und Forschungsoptionen für das weitere Vorhaben umrissen.

## 1.2 Modellgetriebene Decision Support Systeme

Die Hauptaufgabe modellgetriebener DSS besteht in der dynamischen, mehrfachen Analyse oft vektorieller Formeln oder Beziehungen, die zwischen mehreren Teilen oder Variablen in einem diskreten, multidimensionalen Modell bestehen (Power und Sharda, 2007). Im Gegensatz hierzu richten datengetriebene Systeme ihr Hauptaugenmerk auf die Konsolidierung und zeitspezifische Organisation großer Datenmengen, z.B. in Data Warehouses (March und Hevner, 2007). Das Leistungsvermögen und das Potenzial modellgetriebener DSS werden von der Mächtigkeit des Funktionsspektrums definiert, welches dem Anwender bei der Modellierung zur Verfügung steht. Demnach also den Algorithmen und mathematischen Formeln (Power und Sharda, 2007; Sprague und Watson, 1996), die für die Darstellung realer Situationen zur Verfügung stehen. Es sei jedoch erwähnt, dass MDSS nicht der einzige Typ von DSS sind, welche diese Form von Modellen nutzen. Datengetriebene Systeme beinhalten ebenfalls mathematische Modelle, z.B. um abgeleitete Kennzahlen in einem On-Line Analytical Processing (OLAP-)System zu berechnen (Koutsoukis, Mitra & Lucas, 1999). Der Unterschied zwischen den beiden Systemen wird deutlich, wenn man die Zugriffsmöglichkeiten näher betrachtet, die dem Anwender während der Nutzung des Systems auf die Formeln und Beziehungen innerhalb des Modells zur Verfügung stehen. In datengetriebenen Systemen kann deren Manipulation lediglich im Vorfeld der Erstellung einzelner Modell-Instanzen (z.B. OLAP-Würfel), beispielsweise durch Änderung der ETL-Prozesse, stattfinden. So ist es zwar möglich mehrere Szenarien innerhalb eines OLAP-Würfels abzulegen, jedoch ist es im Nachhinein nicht möglich beispielsweise die vordefinierten Dimensionen um Weitere zu ergänzen (March und Hevner, 2007) und somit zwei beliebig unterschiedliche Modell-Instanzen miteinander zu vergleichen. Modellgetriebene DSS stellen dem Anwender hingegen die gesamte Modelllogik zur Verfügung und ermöglichen dadurch unterschiedliche Analysen mit beliebigen Modellparametern, ohne die Einschränkung vordefinierter Fixwerte. Diese Unterschiede führen zu einem typischen Funktionsspektrum je DSS-Typ. Für MDSS sind hier primär Zielwertsuchen (Goal-seeking), What-if und Sensitivitätsanalysen zu erwähnen.

Bei der Zielwertsuche wird ein bestimmter, von anderen Modellkomponenten abhängiger Wert (bezogen auf Spreadsheet-basierte Systeme: ein Zellwert), derart abgeändert, dass ein gewünschtes Ergebnis erzielt wird (Coles und Rowley, 1996). Systemseitig wird nun ermittelt welche Zellen in einer direkten oder indirekten Beziehung zu dem geänderten Wert stehen. Durch Rückwärtsrechnungen, Formelumstellungen und bei diversen Werkzeugen Anwenden gezielter Verteilungsstrategien – z.B. ist denkbar, dass einige Werte unverändert bleiben sollen oder einzelne Formelbestandteile überproportional gewichtet werden (Coles und Rowley, 1996) – werden deren Werte unter den geänderten Voraussetzungen neu berechnet. In einer What-if Analyse werden hingegen in der Regel Werte von unabhängigen Zellen (Rohdaten) oder aber Formelverknüpfungen variiert, um im Weiteren die Auswirkungen der Änderungen auf das Modell und vor allem abhängige Zielwerte zu beobachten (Holsapple, 2008). Der Hauptunterschied einer Sensitivitätsanalyse findet sich in den ermittelten Kennzahlen. Diese beschreiben keine resultierenden Werte oder Auswirkungen leichter Daten- oder Strukturvariationen für ein Modell, sondern ermitteln für zuvor bestimmte Parameter diejenigen Bereiche, in denen eine Wertänderung keine Auswirkung auf abhängige Teile des Modells hat (Werners, 2008; Insua, 1990). Die Bestimmung erfolgt in der Regel mittels komplexer Algorithmen (z.B. lineare Programmierung), mit deren Hilfe diese Einfluss-Elastizitäten bestimmt werden können (Turban et al., 2007).

### **1.3 Zielsetzung**

Das Gestaltungsziel der Arbeit ist das Entwickeln einer konzeptuellen Modellierungsebene einschließlich einer Modellierungssprache, die den Anwender beim Abbilden logischer Zusammenhänge in einem modellgetriebenen DSS unterstützen kann. Es soll ein Vorschlag erarbeitet werden, der das Einfügen einer solchen Ebene in die Architektur modellgetriebener DSS thematisiert.

Um dieses Ziel zu erreichen, muss sichergestellt werden, dass die Modellierungssprache sämtliche zu modellierende Fälle abbilden kann. Des Weiteren sollte eine möglichst standardisierte technische Implementierung die Interaktion mit bestehenden DSS-Werkzeugen unterstützen. Hierfür sind vorhandene Ansätze multidimensionaler Modellierung auf ihre Eignung zu überprüfen.

Aufbauend auf diesen Gegebenheiten wird im nächsten Kapitel ein Vorschlag von Kriterien zur Klassifizierung unterschiedlicher Beziehungen innerhalb dieser Systeme ausgearbeitet. Für ein leichteres Verständnis wird jede Beziehung durch ein vereinfachtes Beispiel veranschaulicht und durch eine kleine Skizze visuell unterstützt.



## 2 Klassifizierungskriterien von Modellbeziehungen

### 2.1 Basisannahmen

Zentrales Element der vorgeschlagenen generischen Spezifikation ist die (formelmäßige) Beziehung zwischen Modellzellen innerhalb eines diskreten, multidimensionalen Raums. Jede Dimension eines Raums soll generell die gleichen Typen arithmetischer Verknüpfungen anbieten können. Demzufolge sollen Dimensionen, die eine spezielle Aufgabe wahrnehmen oder spezielle Funktionen anbieten, vermieden werden. Dadurch ist es möglich auch große Modelle schnell und leicht verstehen und mit angemessenem Aufwand pflegen zu können. Letztlich kommt diese Forderung der flexiblen Anwendungen von DSS-Funktionen zu gute. So muss beispielsweise sichergestellt werden, dass für eine konsolidierende Ausprägung Jahressumme Werte anderer Dimensionen summiert (z.B. Mengen oder Umsätze) oder als Durchschnitt (z.B. Quoten oder Preise) berechnet werden. Eine ähnliche Forderung findet sich bereits in den zwölf Regeln für OLAP-Systeme von Edgar F. Codd (Regel 6 - Generische Dimensionalität) (E. Codd, S. Codd & Salley, 1993). Diesem Ansatz folgend ist der Ausdruck Dimension im Weiteren als generischer Begriff zu verstehen. Das bedeutet, dass Kennzahlen und Dimensionen, im Gegensatz zu anderen mehrdimensionalen Modellierungsansätzen verwandter Themengebiete, keine streng getrennten Bestandteile eines Modells mit unterschiedlichen Bedeutungen darstellen (wie z.B. im SDWM, vgl. Kapitel 3.1), sondern Kennzahlen auch qualifizierende Dimension repräsentieren können. Des Weiteren ist zu erwähnen, dass eine Verknüpfung, die eine bestimmte Logik ausdrückt, lediglich einmal definiert und der Modellierungsprozess nicht-prozedural sein sollte, um eine konsistente und nicht redundante Modellierung gewährleisten zu können.

### 2.2 Kriterien zur Klassifizierung

Die Konstruktion einer generischen Spezifikation möglicher Modellbeziehungen bildet die Basis für eine konzeptuelle Modellierungsebene. Dem Framework für die weitere Erforschung von konzeptuellen Modellierungsansätzen von Wand und Weber (2002) folgend, müssen die Kriterien für eine generische Spezifikation von Modellbeziehungen in MDSS die Modellierung sämtlicher realer Situationen und Zusammenhänge ermöglichen. Die Arbeit von Green (1997) erarbeitet für dieses Framework die beiden Kriterien minimale ontologische Überschneidung (MOO - minimal ontological overlap) und maximale ontologische Vollständigkeit (MOC - maximum ontological completeness), welche durch adäquaten Einsatz der Grammatik erreicht werden sollen. MOO drückt aus, dass ein Konstrukt nicht durch andere grammatikalische Konstrukte oder deren Kombination ausgedrückt werden kann. Hierdurch verringert sich die Wahrscheinlichkeit, dass konfliktäre ontologische Repräsentationen eines realen Zusammenhangs auftreten kön-

nen. Bezogen auf dieses Paper bedeutet dies also, dass es für die Modellierung realer Situationen lediglich eine korrekte Möglichkeit (Kriterien-Kombination) gibt. Maximale ontologische Vollständigkeit wird erreicht, wenn durch die Kombination der vorhandenen grammatikalischen Konstrukte sämtliche reale Phänomene mittels der Grammatik repräsentiert werden können. Gleichzeitig erhöht sich die Wahrscheinlichkeit, dass eine Domäne komplett dargestellt ist.

Jede mögliche Situation, die modelliert werden soll, wird durch ein betriebswirtschaftliches Beispiel veranschaulicht. Diese fokussieren vor allem die Darstellung der dimensionalen Differenzierung von Modellbeziehungen sowie die Herausforderung der Transformation von nicht-prozedural modellierten Zusammenhängen in automatisiert ausführbare Berechnungsschritte (Rieger, 1993). Aufgrund dieser Zusammenhänge und der Möglichkeit sämtliche reale Zusammenhänge modellieren zu können, müssen die Klassifizierungskriterien in der Lage sein die charakteristischen, multidimensionalen Strukturen und sich daraus ergebenden Konsequenzen für die Transformation zu unterscheiden. Multidimensionale Räume in Kombination mit vektoriellen Formeln führen innerhalb Spreadsheets-basierter Systeme in der Regel zu Überschneidungen von Formeln für gewisse Zellbereiche, die bestimmte Formelreihenfolgen bedingen, und speziellen Berechnungsreihenfolgen beteiligter Zellen. Diese beiden Reihenfolgen und deren Unterscheidung sind für das weitere Verständnis eminent wichtig, da sie aus unterschiedlichen Zusammenhängen resultieren. Sie haben ihren Ursprung in speziellen Eigenschaften, die von den Klassifizierungskriterien auf angemessene Art und Weise dargestellt werden müssen. Werden diese Forderungen berücksichtigt, erfüllt eine derartige Spezifikation die Kriterien nach Green sowie Wand und Weber zwangsläufig ebenso. Folgende binäre (zwei mögliche Ausprägungen) Kriterien, welche in ihrer Grundstruktur auf Rieger (1993) zurückgehen, werden vorgeschlagen: *Grad*, *Homogenität* und *Simultaneität* von Dimensionen, die an einer Beziehung beteiligt sind (vgl. Abbildung 1).

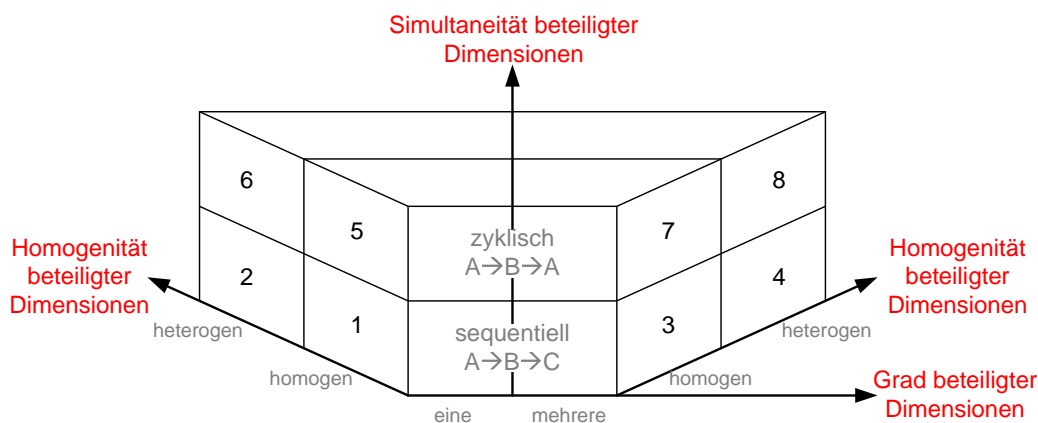


Abbildung 1: Fallklassifizierung modellgetriebener DSS

Folgendes Beispiel erläutert kurz die Wirkung der Ausprägung *Grad*: In einem Modell sind die beiden Dimensionen Kennzahlen und Zeit enthalten. Auf einer Kennzahlen Dimension befinden sich die beiden Formeln „Umsatz = Menge \* Preis“ und „Gewinn = Umsatz – Kosten“. Auf der Dimension Zeit ist lediglich eine Summenformel, die die zwei Ausprägungen „2009“ und „2010“ in einer Ausprägung „Summe“ addiert, vertreten. Bei intra-dimensionalen Formeln ist die Reihenfolge der Dimensionen hier irrelevant, stattdessen spielt nur die Sortierung der Zellen innerhalb der Dimension eine Rolle. Für inter-dimensionale Referenzen muss die Reihenfolge der Dimensionen für alle betroffenen Zellen berücksichtigt werden. Dies ist der Fall, da die Werte referenzierter Zellen, die mit Formeln einer anderen Dimension belegt sind, als der an einer bestimmten Stelle des Berechnungsprozesses aktuellen, bereits vorliegen müssen. Diejenigen Zellen, die nicht aus der Kreuzung mit den Dimensionsausprägungen „2009“ oder „2010“ der Zeit Dimension entstehen, sind zugleich mit der Summenformel aus der Zeit Dimension belegt. Für diese muss ebenfalls die Reihenfolge der Dimensionen, d.h. der Berechnungsformeln aus unterschiedlichen Dimensionen, berücksichtigt werden, um ein korrektes Modell zu erhalten. Generell werden konsolidierende Formeln, wie im vorliegenden Beispiel die Summenformel der Zeit Dimension, eher am Ende eines Berechnungsprozesses platziert, es gibt jedoch Ausnahmen.

Das Kriterium *Homogenität* gibt an, ob die logische Struktur eines Modellierungsproblems die Bildung von Intervallen erfordert. Mögliche Ausprägungen sind homogen, falls keine Fallunterscheidungen nötig sind und heterogen für Fälle, in denen mehrere Formeln unterschieden werden müssen (Intervalle, vgl. Abbildung 1). Liegen verschiedene Intervalle für eine Dimensionsausprägung vor, so müssen intra- und inter-dimensionale Strukturen unterschieden werden. Bei intra-dimensionalen Formeln kann lediglich ein Wert innerhalb derselben Dimension für die Unterscheidung der Formeln verantwortlich zeichnen. Im inter-dimensionalen Fall ist es möglich, dass Ausprägungen weiterer Dimensionen die Anwendung intervallspezifischer Formeln bedingen. Bei mehreren, unterschiedlichen Formeln für eine Dimensionsausprägung kann man sich eine Intervallbildung auch als Partitionierung des Modellraumes vorstellen.

Das dritte und letzte Kriterium *Simultanität* gibt an, ob simultane Strukturen innerhalb von Formeln oder Referenzen vorliegen. Die möglichen Werte werden hierbei durch zyklisch und sequentiell dargestellt. Simultanität besteht, wenn zwei oder mehr Dimensionsausprägungen miteinander zur selben Zeit interagieren, so dass ein Kreisschluss entsteht (Zeit-diskret). In betriebswirtschaftlichen Problemen spielt Simultanität oft eine wichtige Rolle. Je mehr Details bei der Modellierung berücksichtigt werden, desto realistischer ist das Modell der realen Zusammenhänge, da weniger Annahmen getroffen werden müssen. Die Wahrscheinlichkeit eines Auftretens simultaner Strukturen erhöht sich mit dem De-

taillierungsgrad der modellierten Zusammenhänge, da bei mehreren Größen in der Regel auch mehrere Interaktionen innerhalb eines Modells stattfinden (ansonsten würde das Berücksichtigen einer Größe wenig Sinn machen). Simultaneität kann sich auch hinter einer Kette formelmäßiger Zusammenhänge verbergen. Auch in solchen Fällen muss jedoch eine gesonderte Rechnung durchgeführt werden, um ein valides Modell zu erhalten.

### 2.3 Szenarien

Der folgende Abschnitt erläutert acht Modellierungsfälle anhand von typischen betriebswirtschaftlichen Beispielen. Die acht Fälle ergeben sich aus der kartesischen Kombination der Ausprägungen der Klassifizierungskriterien untereinander, die im vorangegangenen Kapitel vorgestellt wurden (drei Kriterien á zwei Ausprägungen). Die Beispiele sind stark vereinfacht dargestellt und auf zwei Dimensionen limitiert, da dies eine einfache Darstellung der wesentlichen, bedeutsamen Zusammenhänge ermöglicht. Des Weiteren würde das Einführen weiterer Dimensionen in dem hier präsentierten Kontext keinen Erkenntnisgewinn mit sich bringen. Die charakteristischen Wirkungen werden für jeden Fall mittels eines Ursache-Wirkungs-Diagramms skizziert. Ein roter Pfeil gibt hierbei die Wirkungsrichtung der Formeln für jeden spezifischen Fall an. Ist dieser diagonal angeordnet, liegt eine inter-dimensionale Beziehung und wenn er vertikal verläuft, liegt eine intra-dimensionale Beziehung vor. Zwei gegenläufige Pfeile stellen wiederum simultane Strukturen dar und eine gestrichelte Linie repräsentiert heterogene Beziehungen, für die eine Fallunterscheidung hinsichtlich der Formeln vorgenommen werden muss. Zweitrangige oder analoge Wirkungen werden innerhalb der Diagramme nicht berücksichtigt.

Im Folgenden werden zunächst die Fälle 1 bis 4 (vgl. Abbildung 2) behandelt, die sequentielle Strukturen kennzeichnen. Danach werden die Fälle 5 bis 8 (vgl. Abbildung 3), die simultane Strukturen aufweisen, näher erläutert.

#### Fall 1:

Fall 1 kennzeichnet intra-dimensionale Beziehungen, welche ohne Fallunterscheidung und ausschließlich sequentielle Strukturen aufweisen. Das bedeutet, dass eine intra-dimensionale Formel „Umsatz = Menge \* Preis“ auf jede Zelle angewandt wird, die durch Kreuzung mit der Dimensionsausprägung „Umsatz“ entsteht.

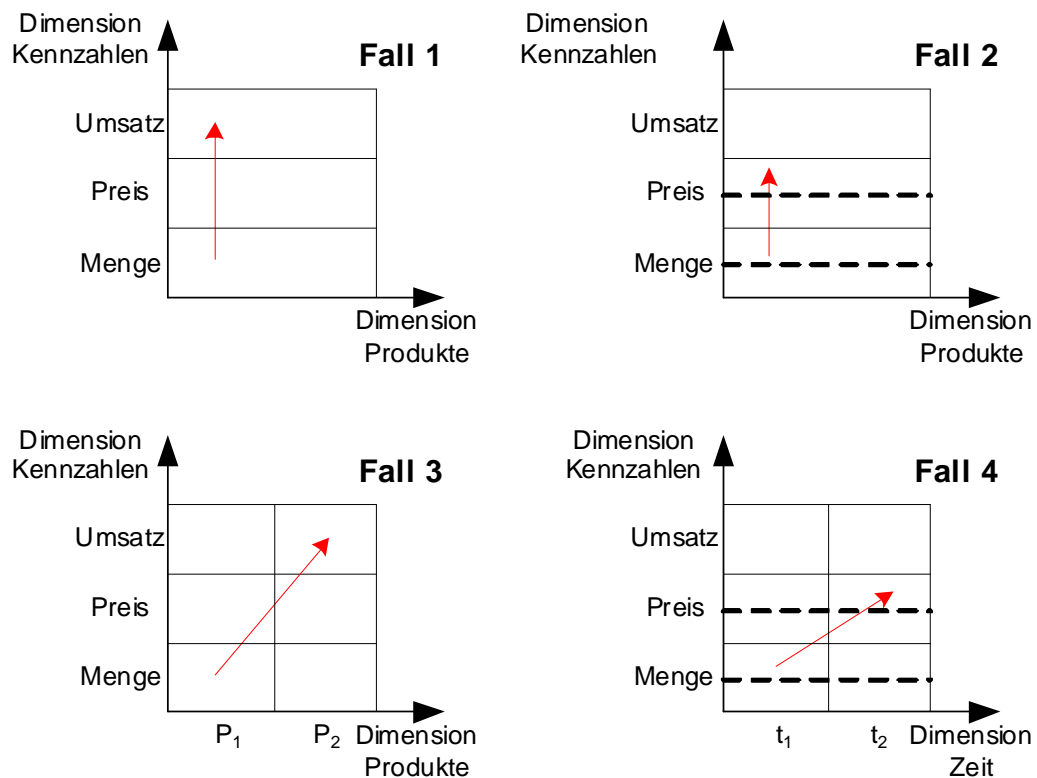


Abbildung 2: Fälle 1 bis 4 mit sequentieller Struktur

Die korrekte Reihenfolge der Berechnungsschritte kann durch einfaches Umsortieren der Zellen bzw. ein Sortieren der Formeln für Zellen, die mit mehr als einer Formel belegt sind, erreicht werden. Für alle Zellen, die lediglich mit intra-dimensionalen Formeln belegt sind, ist die Reihenfolge der Dimensionen irrelevant (vgl. Beispiel für die Wirkung des Klassifizierungskriteriums Grad aus Kapitel 3.2).

#### Fall 2:

Zusätzlich zu Fall 1 werden in diesem Fall Intervalle, das heißt heterogene Formeln eingeführt. Spezifische Werte einer beliebigen Dimensionsausprägung sorgen bei Beziehungen dieser Art dafür, dass unterschiedlich definierte Formeln für die Berechnung ein und derselben Dimensionsausprägung verantwortlich zeichnen. In Bezug auf das Beispiel aus Fall 1 könnte ein Rabattsatz, der ab einer bestimmten Menge gewährt wird, diesen Fall darstellen. Die Formeln könnten wie folgt gestaltet sein: „Umsatz = Menge \* Preis \* (1 – Rabattsatz)“. Für eine Menge, die kleiner als 100 ist, würde der Rabattsatz automatisch „0“ sein und für Mengen größer 100 durch eine Formel „Menge / 10.000“ berechnet werden.

Für die korrekte Berechnungsreihenfolge muss der Umsatz bei Mengen größer „100“ immer nach dem Rabatt berechnet werden. Für Mengen kleiner „100“ fällt diese Sperre weg, da die Formel für den Umsatz an beiden Stellen des Berechnungsprozesses das gleiche Ergebnis zur Folge hätte.

**Fall 3:**

Während in Fall 1 und 2 ausschließlich intra-dimensionale Referenzen vorlagen beschreibt Fall 3 einen „echten“ multidimensionalen Raum, in dem eine Formel mindestens zwei Dimensionen miteinander in Relation setzt. Eine eingeschränkte Referenz bezüglich bestimmter Werte (Intervalle) liegt jedoch nicht vor.

Bei der Transformation nicht-prozedural modellierter Strukturen in prozedurale Berechnungsschritte muss die Reihenfolge, nach der die Dimensionen abgearbeitet werden, beachtet werden. Wird eine Dimensionsausprägung innerhalb einer Formel, die einer weiteren Dimension entstammt, referenziert, muss die betreffende Zelle zuvor berechnet worden sein. Es wird schnell klar, dass es nicht machbar ist, Zellen einer ganzen Dimension nacheinander zu berechnen, sondern eine bestimmte Reihenfolge unabhängig von Dimensionen bei der Berechnung der Zellen hergestellt werden muss.

Ein Beispiel für diesen Typ von Beziehung stellen Formeln für die Abbildung eines mehrstufigen Produktionssystems dar. Es muss sichergestellt werden, dass nicht mehr (Zwischen-)Produkte zur Herstellung eingeplant werden, als es die jeweils vorhergehende Stufe zulässt. So kann zum Beispiel in einer Tischlerei nicht mit 100 Tischen (vgl. Abbildung 2 Fall 3 Größe  $P_2$ ) geplant werden, sofern lediglich 100 Tischbeine (vgl. Abbildung 2 Fall 3 Größe  $P_1$ ) für den nächsten Produktionslauf vorliegen.

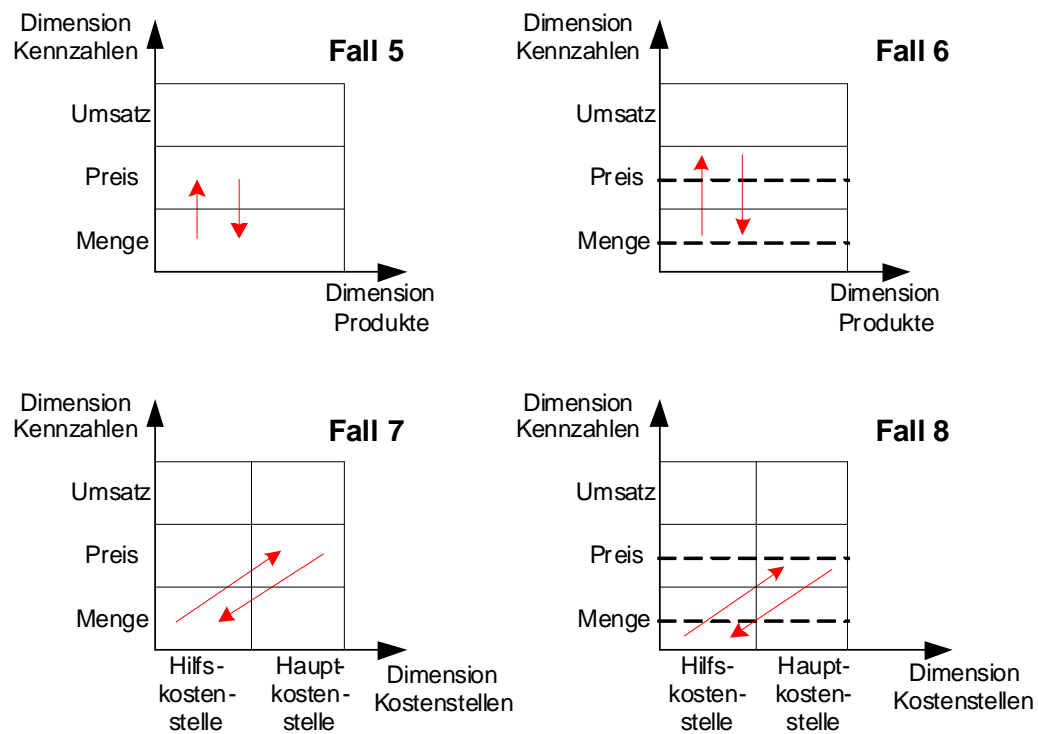


Abbildung 3: Fälle 5 bis 8 mit simultaner Struktur

#### Fall 4:

Fall 4 als letzte sequentielle Beziehung ermöglicht zusätzlich zu Fall 3 eine Fallunterscheidung (Intervalle) für bestimmte Werte von Dimensionsausprägungen. Im Vergleich zum Beispiel aus Fall 2 würde die Absatzmenge einer vorangegangenen Periode über die Möglichkeit der Gewährung eines Rabattes entscheiden. Dies hätte zur Folge, dass in der Formel für den Rabattsatz eine Zeitdimension zugegriffen werden müsste.

Die folgenden Fälle 5 bis 8 (vgl. Abbildung 3) sind durch simultane Berechnungsstrukturen charakterisiert. Eine simultane Formel muss in einer gesonderten Nebenrechnung, für die es mehrere Strategien gibt, aufgelöst werden. Häufig wird auf Modelle Linearer Programmierung zurückgegriffen, da diese in der Regel auf einfache Art und Weise sehr nah an die optimale Lösung herankommen.

#### Fall 5:

Der erste Fall mit simultanen Strukturen berücksichtigt lediglich eine Dimension und keine Fallunterscheidungen für die Formel einer Dimensionsausprägung.

Nimmt man das Beispiel aus Fall 1 bezüglich der einfachen Umsatzberechnung, werden in diesem Fall nun eine Angebots- und eine Nachfragefunktion abgebildet. Diese beeinflussen sich jeweils gegenseitig, indem die Menge vom Preis (Angebotsfunktion) und der Preis von der Menge (Nachfragefunktion) abhängen.

Beim Etablieren einer korrekten Berechnungsreihenfolge bezüglich Zellen und Formeln muss jede Zelle, die direkt von der simultanen Formel abhängt, in einer Nebenrechnung berechnet werden. In dem Beispiel wären dies die beiden Zellen „Preis“ und „Menge“ des abgebildeten Produkts. Die Nebenrechnung müsste nun an einer bestimmten Stelle des gesamten Berechnungsprozesses platziert werden. Diese hängt wiederum von der Sortierung der restlichen Zellen und dem Zusammenspiel dieser mit den simultanen Größen ab. Für Zellen, die mit mehr als einer Formel belegt sind, die aus unterschiedlichen Dimensionen stammen, ist zudem die Formelsortierung (Reihenfolge der Dimensionen) relevant.

**Fall 6:**

Fall 6 führt die Berücksichtigung von unterschiedlichen Formelalternativen für eine Dimensionsausprägung im simultanen Kontext ein. Zusätzlich zu Fall 5 wären die Angebots- bzw. Nachfragefunktion für ein bestimmtes Mengenintervall (Nachfragefunktion) bzw. ein bestimmtes Preisintervall (Angebotsfunktion) völlig unelastisch. Ab einer bestimmten Menge bzw. einem bestimmten Preis würden sich die Funktionen jedoch auf ähnliche Weise, wie in Fall 5 beschrieben, beeinflussen. Dies hat zur Folge, dass die unterschiedlichen Intervallformeln, sowohl für die Menge als auch für den Preis, an unterschiedlichen Stellen des Berechnungsprozesses platziert werden würden, da die simultanen Zusammenhänge wiederum in einer Nebenrechnung gelöst werden müssen.

**Fall 7:**

Dieser simultane Zusammenhang verbindet Ausprägungen unterschiedlicher Dimensionen miteinander, unterscheidet hierbei jedoch keine Formelalternativen (homogen).

Die innerbetriebliche Leistungsverrechnung ist ein klassisches Beispiel hierfür. Eine Hilfskostenstelle ist einer Hauptkostenstelle zugeordnet. Auf der einen Seite konsumiert die Hilfskostenstelle Leistungen der Hauptkostenstelle, auf der anderen Seite stellt sie der Hauptkostenstelle Leistung zur Verfügung. Die benötigte Leistung der Hilfskostenstelle beeinflusst den Preis der Hauptkostenstelle und umgekehrt.

Um in diesem Fall eine korrekte Berechnungsreihenfolge der Zellen und Formeln herzustellen müssen wiederum die Reihenfolge der Dimensionen beachtet und die Nebenrechnung zur Auflösung der simultanen Größen an eine spezifische Stelle des Gesamtprozesses eingegliedert werden.



**Fall 8:**

Der achte und letzte Fall deckt die Möglichkeit von simultanen Formeln in einem multi-dimensionalen Raum ab, die – im Gegensatz zu Fall 7 – nicht in gleicher Weise für alle Werte der betroffenen Dimensionsausprägung gelten. Erweitert man das Beispiel der innerbetrieblichen Leistungsverrechnung aus Fall 7 dahingehend, dass Kapazitätsgrenzen die Menge an Leistung beschränken, die eine Hilfskostenstelle einer Hauptkostenstelle zur Verfügung stellt, ergibt sich als Konsequenz, dass die Hauptkostenstelle selbige Leistung von außen zukaufen muss. Für einige Werteausprägungen von Elementen der Dimension Kennzahlen gilt also eine unterschiedliche Berechnungsvorschrift (einmal „intern“ und einmal „extern“).

Die unterschiedlichen Intervallformeln müssen an unterschiedlichen Stellen des Berechnungsprozesses platziert werden da die simultanen Zusammenhänge sich auf ein Intervall beziehen und wiederum in einer Nebenrechnung gelöst werden müssen. Diese muss in den gesamten Berechnungsprozess eingegliedert werden und zudem ist die Reihenfolge der Dimensionen aufgrund der inter-dimensionalen Beziehungen relevant.

Im Folgenden werden nun bekannte Ansätze konzeptueller Modellierung kurz vorgestellt und auf ihre Potenziale bei der Realisierung einer konzeptuellen Modellierungsebene in MDSS vor dem Hintergrund der gerade vorgestellten Fallklassifizierung untersucht.

### **3 Bekannte Ansätze konzeptueller Modellierung**

#### **3.1 Semantisches Data-Warehouse Modell (SDWM)**

Das semantische Data-Warehouse Modell (Böhnlein und Ende, 2001) ist ein an der Universität Bamberg entwickelter Ansatz, dessen Intention die Bereitstellung einer rein konzeptuellen Betrachtungsweise multidimensionaler Datenstrukturen im OLAP- bzw. Data-Warehouse Umfeld ist. Grundsätzlich liegt dem Ansatz das Konzept der Bildung von Sichten zu Grunde. Die Autoren bezeichnen quantitative Daten als Kennzahlen und qualitative Aspekte als Dimensionen.

Die Dimensionssicht wird durch die Bestandteile Dimension, Dimensionshierarchiestufe, Aggregationsbeziehung und Dimensionsschnittstelle charakterisiert. In einer Dimension sind zumindest zwei Dimensionshierarchiestufen und eine Aggregationsbeziehung vorhanden. Durch die Aggregationsbeziehung werden Dimensionshierarchiestufen miteinander verbunden. Die Dimensionsausprägungen sind einzelnen Hierarchiestufen zugeordnet und bilden von der niedrigsten zur höchsten Stufe eine Verdichtung der Daten über eine Aggregation ab. Eine Dimensionsschnittstelle bildet eine Verknüpfung von Dimensionen

mit einer einzelnen Basiskennzahl ab. Dies stellt die Wiederverwendbarkeit von Dimensionen sicher.

Hauptbestandteile der Kennzahlensicht sind die Elemente Basiskennzahl, abgeleitete Kennzahl, Dimensionsschnittstelle und Dimension. Das Kombinieren von Dimensionen und Kennzahlen als zentrale Eigenschaft multidimensionaler Datenstrukturen steht hier im Fokus. Basiskennzahlen, welche über eine eindeutige Dimensionsschnittstelle mit einer Dimension verknüpft werden, können über gerichtete Kennzahlbeziehungen zu abgeleiteten Kennzahlen kombiniert werden. Eine abgeleitete Kennzahl besitzt zur Spezifikation ihrer Zusammensetzung ein Attribut Berechnungsvorschrift. Die Möglichkeit nicht-additive, semi-additive und uneingeschränkt additive Verdichtungen innerhalb von Dimensionen abzubilden, wird durch ein Attribut Aggregierbarkeit, welches der Dimensionsschnittstelle zugeordnet ist, realisiert. Ebenfalls findet sich dort ein Attribut Dimensionslinie, welches die Granularität einer Basiskennzahl angibt, d.h. bis zu welcher Hierarchiestufe einer Dimension eine Verfeinerung möglich ist.

Bezogen auf den Kontext modellgetriebener DSS ergeben sich aus diesen Strukturen Einschränkungen der Modellierungsmächtigkeit. So ist es z.B. nicht möglich, Teilmengendifferenzierte Kennzahlenspezifikationen innerhalb einer Dimension zu definieren (keine Intervalle bzw. heterogene Dimensionen in Bezug auf Referenzen möglich). Des Weiteren ergibt sich, dass andere Beziehungen als Aggregationen zwischen zwei Dimensionsausprägungen (auf derselben Hierarchiestufe) nicht mittels SDWM abzubilden sind. Simultane Referenzen, welche für die Mächtigkeit modellgetriebener DSS oft relevant sind, werden in diesem System nicht gesondert berücksichtigt. Eine prinzipiell denkbare SDWM-Erweiterung soll hier nicht weiter verfolgt werden, da dies angesichts des Sichtenkonzepts zu redundanten Modellierungskonzepten führt.

### 3.2 Multidimensional Modeling Language (MML)

Die Multidimensional Modeling Language (Herden und Harren, 2004) ist eine an der Universität Oldenburg entwickelte Sprache, zur konzeptionellen Modellierung multidimensionaler Data-Warehouse-Schemata. Grundsatz ist die Differenzierung von Daten, dem zugehörigen multidimensionalen Kontext und Elementen für die Beschreibung von Struktureigenschaften. Die MML wurde mittels einer multidimensionalen Erweiterung der Unified Modeling Language (UML) realisiert und nutzt typische Konstrukte aus der objektorientierten Programmierung, wie z.B. Vererbung oder Generalisierung.

MML besteht aus verschiedenen Konstrukten, die in die fünf Bereiche Daten-Elemente, Hilfsmetaklassen, Multidimensionaler Kontext, Allgemeine Verbindungen und Properties unterteilt sind. Ursprung eines jeden MML-Konstrukts ist die abstrakte Metaklasse MMLElement, in der jeder Klasse eine eindeutige Bezeichnung zugeordnet wird. Der

Bereich Daten-Elemente entspricht einem Datentyp und beschreibt, ob es sich bei einem Daten aufnehmenden Element um einen elementaren oder komplexen Typ handelt. Im Bereich Hilfsmetaklassen werden die Eigenschaften der Vererbung einer Klasse gesteuert und der Bereich Multidimensionaler Kontext führt die bekannte Unterscheidung in Fakten und Dimensionsdaten ein. Mögliche Verknüpfungen zwischen Schemaelementen werden in den Bereichen Allgemeine Verbindungen und Properties geregelt. Während erstgenannter typische objektorientierte Konstrukte, wie z.B. Assoziation und Komposition zur Verfügung stellt, dient der Bereich Properties der Darstellung multidimensionaler Sachverhalte, wie z.B. dem Aufbau von Hierarchien.

Bezogen auf ein modellgetriebenes DSS folgen hieraus eingeschränkte Möglichkeiten bei der Modellierung bestimmter Sachverhalte. So ist es zwar für eine bestimmte Grundgesamtheit von Elementen möglich mittels eines „SharedRollUp“ aus dem Bereich Properties intervallspezifische Berechnungsvorschriften zu definieren, eine Betrachtung von Dimension als generischer Begriff, der auch Faktenelemente als qualifizierende Dimension zulässt, ist jedoch nicht möglich. Weitere Einschränkungen liegen bei der Berücksichtigung von Besonderheiten simultaner Referenzen vor. Diese bleiben gänzlich unberücksichtigt, was eine gesonderte Behandlung im Weiteren erschwert.

## 4 Weiteres Vorgehen

Aus wissenschaftstheoretischer Sicht sollen die bislang vorgestellten Ergebnisse im Sinne des konstruktionsorientierten Ansatzes in Verbindung mit deduktiven Elementen zur Erlangung neuer Erkenntnisse fortgeführt werden. Dieser Ansatz bietet sich hier vor allem deshalb an, da eine Situation gegeben ist, zu der bisher keine oder nur weniger zufrieden stellende Lösungsansätze vorhanden sind, sich gleichwohl aber eindeutige Anforderungen spezifizieren lassen (Frank, 2007). Diese Anforderungen konnten durch die, auf Basis von aus der Literatur gewonnenen Kriterien, hergeleitete Fallklassifizierung für modellgetriebene DSS definiert werden. Auf Basis dieser Vorarbeiten werden folgende Forschungsaspekte für konkrete Schritte im Rahmen des Dissertationsprojekts gesehen:

Die anhand von singulären Beispielen veranschaulichten Modellierungsfälle sollen in einem nächsten Schritt zu einem Referenzmodell weiterentwickelt werden. Im Folgenden soll diese Referenz als Kriterium zur Beurteilung der Mächtigkeit aktuell am Markt befindlicher DSS-Werkzeuge herangezogen und deren Fehleranfälligkeit anhand einer zu entwickelnden Metrik gemessen werden. Hierzu werden ausgewählte Personen, die bereits über Erfahrung in der Modellierung betriebswirtschaftlicher Problemstellungen mittels DSS-Werkzeugen verfügen, die einzelnen Fälle mit einem Werkzeug implementieren. Danach wird die Güte der Modellierung anhand der Metrik überprüft.

Anschließend wird die Hypothese, dass eine konzeptuelle Modellierung die Fehleranfälligkeit eines MDSS reduzieren kann, untersucht. In diesem Zusammenhang werden aktuelle Konzepte und Ansätze auf die Möglichkeit untersucht, ob und inwiefern sie einen Beitrag bei der Umsetzung einer konzeptuellen Modellierungssprache für MDSS leisten können. Hierbei handelt es sich hauptsächlich um Konzepte multidimensionaler Modellierung, die ihren Ursprung im On-Line Analytical Processing haben, wie z.B. die in Kapitel 3 erwähnten Ansätze SDWM und MML.

Nach der Evaluation bestehender Ansätze wird aufgrund der gewonnenen Erkenntnisse ein Vorschlag einer Modellierungssprache für modellgetriebene DSS erarbeitet. Die einzelnen Differenzierungsmerkmale der Fallklassifizierung aus Teil 1 werden durch Sprachkonstrukte und -attribute repräsentiert. Für die technische Realisierung ist die eXtensible Markup Language (XML) vorgesehen.

Abschließend könnte eine prototypische Implementierung der konstruierten Modellierungssprache entwickelt und dokumentiert werden. Neben der technischen Realisierung sollen in einer abschließenden Evaluation Relevanz, Möglichkeiten und Grenzen der konzeptuellen Modellierungssprache ausgearbeitet werden, indem ein Test anhand der erarbeiteten Referenz durchgeführt wird. Wiederum sollen die ausgewählten Personen die einzelnen Fälle implementieren und anschließend die Güte der Modellierung überprüfen, jedoch ausgehend von Spezifikationen in der konzeptuellen Modellierungssprache. Für den Vergleich und die Darstellung der Ergebnisse wird wiederum die bereits ausgearbeitete Metrik herangezogen.

## 5 Literatur

Alter, S. (1996). *Information Systems: A management perspective*. Menlo Park, CA: Benjamin / Cummings Publishing.

Böhnlein, A., Ende, U.-v. (2001). Ein konzeptuelles Data Warehouse-Modell für die Erstellung multidimensionaler Datenstrukturen. In Gesellschaft für Informatik, Fachgruppe 5.10 Informationssystem-Architekturen (Hrsg.), *Rundbrief 8. Jahrgang* (S. 25-57).

Codd, E., Codd, S., & Salley, C. (1993). *Providing OLAP to User-Analysts: An IT Mandate*. Abgerufen von: [http://www.minet.uni-jena.de/dbis/lehre/ss2005/sem\\_dwh/lit/Cod93.pdf](http://www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf) (am 15.06.2009).

Coles, S., & Rowley, J. (1996). Spreadsheet modelling for management decision making. *Industrial Management & Data Systems*, 96(7), 17-23.

Frank, U. (2007). Ein Vorschlag zur Konfiguration von Forschungsmethoden in der Wirtschaftsinformatik, in: F. Lehner; S. Zelewski (Hrsg.), *Wissenschaftstheoretische Fundierung und wissenschaftliche Orientierung der Wirtschaftsinformatik* (S.155-184). Berlin, Deutschland: Gito-Verlag.

Green, P. (1997). Use of Information Systems Analysis and Design (ISAD) Grammars in Combination in Upper CASE Tools – An Ontological Evaluation. In K. Siau, Y. Wand, & J. Parsons (Hrsg.), *Proceedings of the 2nd CAI-SE/IFIP8.1 International Workshop on the Evaluation of Modeling Methods in Systems Analysis and Design (EMM-SAD'97)* (S. 1-12). Barcelona, Spanien.

Herden, O., Harren, A. (2004). Die ODAWA-Methodik für den Entwurf von Data-Warehouse-Datenbanken. *Informatik - Forschung und Entwicklung*, 19(2), 87-96.

Holsapple, C. W., & Whinston, A. B. (1996). *Decision Support Systems: a knowledge-based approach*. Minneapolis, MN: West Publishing.

Holsapple, C. W. (2008). DSS Architecture and Types. In F. Burstein, & C. W. Holsapple (Hrsg.), *Handbook on Decision Support Systems 1* (S. 163-189). Berlin, Deutschland: Springer-Verlag.

Insua, D. R. (1990). *Sensitivity analysis in multi-objective decision making* (Lecture Notes in Economic and Mathematical Systems). Berlin, Deutschland: Springer-Verlag.

Koutsoukis, N. S., Mitra, G., & Lucas, C. (1999). Adapting on-line analytical processing for decision modelling: the interaction of information and decision technologies. *Decision Support Systems*, 26(1), 1-30.

March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, 43(3), 1031-1043

Mayer, M. K. (1998). Future trends in model management systems: parallel and distributed extensions. *Decision Support Systems*, 22(4), 325-335.

Panko, R., & Sprague, R. Jr. (1998). Hitting the wall: errors in developing and code inspecting a 'simple' spreadsheet model. *Decision Support Systems*, 22(4), 337-353.

Powell, S. G., Baker, K. R., & Lawson, B. (2008). A critical review of the literature on spreadsheet errors. *Decision Support Systems*, 46(1), 128-138.

Power, D., & Sharda, R. (2007). Model-driven decision support systems: Concepts and research directions. *Decision Support Systems*, 43, 1044-1061.

Rieger, B. (1993). *Der rechnergestützte Arbeitsplatz für Führungskräfte*. Habilitationsschrift, Technische Universität Berlin.

Simon, H. A. (1977). *The new science of management decision*. Englewood Cliffs, NJ: Prentice-Hall.

Sprague, R. Jr., & Watson, H. (1996). *Decision Support for Management*. Upper Saddle River, NJ: Prentice-Hall.

Turban, E., Aronson, J., Liang, T.-P., & Sharda, R. (2007). *Decision Support and Business Intelligence Systems*. Upper Saddle River, NJ: Pearson.

Wand, Y., & Weber, R. (2002). Research Commentary: Information Systems and Conceptual Modeling – A Research Agenda. *Information Systems Research*, 13(4): 363-376.

Werners, B. (2008). *Grundlagen des Operations Research*. Berlin, Deutschland: Springer-Verlag.



# **Ein Wiki-basiertes Vorgehensmodell für Business Intelligence Projekte**

Stephan König

Fachhochschule Hannover – Fakultät IV (Wirtschaft und Informatik)

## **1 Einleitung**

Business Intelligence (BI) umfasst die Integration von Strategien, Prozessen und Technologien, um im Umfeld der entscheidungsunterstützenden Systeme aus fragmentierten, inhomogenen Unternehmens-, Markt- und Wettbewerberdaten erfolgskritisches Wissen über Status, Potenziale und Perspektiven zu generieren und dies für Analyse-, Planungs- und Steuerungszwecke geeignet darzustellen (Gluchowski et al., 2008; Kemper et al., 2006).

Projekte zur Entwicklung von BI Anwendungen unterscheiden sich in wichtigen Punkten vom Software Engineering im Rahmen klassischer Anwendungsentwicklungsprojekte (Anahory et al., 1997; Becker et al., 2006; Fütting, 2000; Giorgini, 2005; Gluchowski et al., 2008; Heck-Weinhart et al., 2003; Keppel et al., 2001; Konzelmann, 2008; Oehler, 2006; Strauch et al., 2002):

- Die Anforderungsanalyse ist schwierig, da sich zukünftiges exploratives Anwenderverhalten nur schwer in vorab klar definierten Use Cases abbilden lässt.
- Es handelt sich bei BI Anwendungen um komplexe zentrale Systeme, die eng mit bestehenden operativen Systemen vernetzt sind und eine enge Zusammenarbeit zwischen der IT Abteilung und den Fachabteilungen erfordern.
- Die abteilungsübergreifende Integration der Daten erfordert neben politischem Geschick insbesondere auch eine einheitliche Definition von fachlichen Begrifflichkeiten und Metriken.

Die Verwendung einer geeigneten Vorgehensweise gehört zu den wesentlichen Erfolgsfaktoren von BI Projekten (Bauer et al., 2008; Oehler, 2006). In der Praxis haben sich

iterative Vorgehensweisen bewährt (Bauer et al., 2008; Becker et al., 2006; Breslin, 2004; Gardner, 1998; Gluchowski et al., 2008; Heck-Weinhart et al., 2003; Kemper et al., 2006; Sen et al., 2005). Sie stellen einen Kompromiss zwischen möglichst kurzen Releasezyklen (Risikoreduktion, kurzfristige Umsetzung fachlicher Anforderungen) und den oben genannten Spezifika von BI Projekten dar.<sup>8</sup>

In diesem Zusammenhang stehen die folgenden beiden Fragen im Vordergrund dieser Untersuchung:

1. Wie sieht ein anwendungsorientiertes Vorgehensmodell für den praktischen Einsatz in BI Projekten aus?
2. Welche Form der Darstellung des Vorgehensmodells ermöglicht ein Wissensmanagement über mehrere Iterationen? Kann gleichzeitig die Weiterentwicklung des Vorgehensmodells und die Dokumentation der Projektergebnisse unterstützt werden?

## **2 Anforderungen an ein anwendungsorientiertes Vorgehensmodell für BI Projekte**

Unter einem Vorgehensmodell wird eine Folge von Aktivitäten und Ergebnissen verstanden, die zur erfolgreichen Durchführung eines Projekts erforderlich sind. Ergänzt wird ein Vorgehensmodell durch Prinzipien, Methoden, Techniken und Werkzeuge. Hauptziel eines Vorgehensmodells ist es, eine komplexe Problemstellung in leichter lösbare Teilaufgaben zu reduzieren und diese phasenweise zu strukturieren, um so schneller und verlässlicher die angestrebten Projektziele zu erreichen.

Vorgehensmodelle erlauben es, Erfahrungswissen zu explizieren und so eine weitgehende Personenunabhängigkeit zu erreichen. Vorgehensmodelle dürfen aber nicht zum Selbstzweck werden und sollten nicht im Widerspruch zu Intuition und Kreativität stehen. Vielmehr sind sie als Leitfaden für die Lösung praktischer Probleme zu verstehen.

Das im folgenden zu entwickelnde Vorgehensmodell für BI Projekte soll sich darüber hinaus durch folgende Charakteristika auszeichnen:

---

<sup>8</sup> Für die Entwicklung kleinerer isolierter Software-lastiger Anwendungen mit erfahrenen Entwicklern und aktiver Beteiligung zukünftiger Anwender werden zunehmend agile Vorgehensweisen eingesetzt. Deren Ansätze können insbesondere bei der Festlegung eines möglichst begrenzten Umfanges der Iterationen eines BI Projektes hilfreich sein (siehe z.B. Hughes, 2008; Brobst et al., 2009).



- **Detailgrad:** Hauptzielgruppe sind Projektmanager. Für diese Gruppe ist es entscheidend, alle wesentlichen Aktivitäten im Sinne einer Checkliste im Blickfeld zu haben. Das BI Vorgehensmodell ist weniger für Spezialisten gedacht, die Antworten auf fachliche oder technische Detailfragen suchen.
- **Operationalisierbarkeit:** Akzeptanzprobleme vieler Vorgehensmodelle in der Praxis lassen es sinnvoll erscheinen, auf eine einfache Operationalisierbarkeit der Aktivitäten zu achten, um so eine tatsächliche Hilfestellung in der Projektarbeit zu erreichen. Dies muss durch die Wahl eines geeigneten Detaillierungsgrades unterstützt werden.
- **Aufteilung in Phasen und Disziplinen:** Das BI Vorgehensmodell gliedert die Aktivitäten in Projektphasen. Die Aktivitäten der jeweiligen Phase werden darüber hinaus in Disziplinen gruppiert.
- **Vollständigkeit:** Das BI Vorgehensmodell ist für ein umfassendes BI Projekt im Sinne der oben gegebenen BI Definition gedacht. Für spezifischere Projekte werden einige Aktivitäten entfallen bzw. müssen modifiziert werden.
- **Iterativ:** Der Fokus liegt auf der ersten Iteration. Folgende Iterationen sind in aller Regel weniger komplex, da viele Grundlagen bereits vorhanden sind. Spezielle Aufgaben des BI Programmmanagements stehen nicht im Vordergrund.
- **Unabhängigkeit:** Das BI Vorgehensmodell ist Produkt- und Architekturunabhängig. Es vermittelt keine Kenntnisse über BI Standardarchitekturen sondern setzt diese voraus. Grundsätzlich folgt die Vorgehensweise aber dem angebotsorientierten Ansatz von Kimball (Business Dimensional Lifecycle Model) (Breslin, 2004; Kimball, 2008).

Angedachte Einsatzszenarien für das BI Vorgehensmodell liegen in (BI) Beratungsunternehmen, die z.B. bei wachsender Mitarbeiterzahl eine einheitliche Vorgehensweise erreichen möchten. Darüber hinaus ist es aber auch für alle Unternehmen gedacht, die vor der Herausforderung komplexer BI Projekte stehen und ein praxisnahes Vorgehensmodell suchen.

### **3 Stand der Forschung**

In der Literatur finden sich zahlreiche Ansätze für BI Vorgehensmodelle. Diese lassen sich wie folgt kategorisieren:

- Spezielle Lehrbücher, die sich ausschließlich mit dem Thema BI Vorgehensmodelle beschäftigen und ein Vorgehensmodell im Detail beschreiben (Kimball, 2008; Moss et al., 2006; Hughes, 2008).
- Allgemeine Lehrbücher und Veröffentlichungen im Themenumfeld BI und Data Warehousing, die sich (auch) mit BI Projektmanagement beschäftigen (Anahory et al., 1997; Chenoweth, 2006; Heck-Weinhart et al., 2003; Inmon, 2005; Reinschmidt et al., 2000; Sammon et al., 2000; Schirp, 2001; Westermann, 2000) und z.T. auch konkrete Vorgehensmodelle skizzieren (Gardner, 1998; Gluchowski et al., 2008; Höhn, 2000; Kemper et al. 2006; Turban et al. 2006).
- Veröffentlichungen (Becker et al., 2006; Keppel et al., 2001; Konzelmann, 2008; Sen et al., 2005) und Lehrbücher (Schrödl, 2009; Staade et al., 2006), die produktspezifische BI Vorgehensmodelle vorstellen.
- Umfangreiche wissenschaftliche Arbeiten, die z.T. sehr komplexe und methodisch fundierte BI Vorgehensmodelle entwerfen (Hecht, 2001; Holten, 1999).
- Veröffentlichungen, die sich auf bestimmte Phasen oder Aspekte eines BI Vorgehensmodells fokussieren (Becker et al., 2004; Giorgini, 2005; Goeken, 2004; Goeken, 2007; Golfarelli et al., 1998; Herder, 2001; Strauch et al., 2002).

Insgesamt unterstreichen die große Anzahl und das breite Spektrum der verfügbaren Literatur die Relevanz der Thematik. Ein BI Vorgehensmodell mit den oben aufgeführten Charakteristika ist aber bisher nicht veröffentlicht worden.

Im Folgenden werden die vorhandenen Gestaltungsansätze und beobachteten Praktiken konzeptionell homogenisiert und im Sinne der oben aufgeführten Charakteristika in ein konsistentes BI Vorgehensmodell zusammengeführt.

## 4 Grundstruktur des BI Vorgehensmodells

Das BI Vorgehensmodell gliedert sich in sieben Projektphasen (siehe Tabelle 1). Die Phasen weisen die in Tabelle 1 genannten Ziele und Meilensteine auf. Der Durchlauf aller sieben Phasen entspricht dabei einer Iteration im Sinne der iterativen Vorgehensweise. Die Phasen Deployment und Produktion wurden aufgrund ihrer Kritikalität für den Gesamterfolg des Projekts explizit mit aufgenommen.

Projektphasen	Ziel(e)	Meilenstein(e)
<b>Initialisierung</b>	Definition des BI Projekts	Abnahme Projektdefinition Projektfreigabe
<b>Analyse</b>	Detailanalyse fachliche und technische Anforderungen. Erstellung Projektplan	Abnahme Pflichtenheft. Abnahme Projektplan
<b>Design</b>	Detailentwurf BI Anwendungen, Datenbanken und ETL Prozesse	Abnahme fachliches und technisches Feinkonzept. Abnahme A&I Plattform
<b>Realisierung</b>	Realisierung der fachlichen und technischen Feinkonzepte bis zur Implementierungsreife	Abnahme Unit Test BI Anwendungen, Datenbanken und ETL Prozesse
<b>Test</b>	Durchführung der verschiedenen Testarten	Fachliche und technische Abnahme. Produktionsfreigabe
<b>Deployment</b>	Vorbereitung Produktion. Deployment der fertigen BI Lösung in die bestehende IT Umgebung	Start Produktion
<b>Produktion</b>	Stabile Produktion. Wachstum. Erweiterungen	Definition Folgeprojekt

*Tabelle 1: Grundstruktur des BI Vorgehensmodells - Projektphasen*

Die Aktivitäten der jeweiligen Phase werden in sieben Disziplinen gruppiert (siehe Tabelle 2). Dies ermöglicht eine bessere Strukturierung der zahlreichen Aktivitäten einer Phase.

Disziplinen	Beschreibung
allgemeines Projektmanagement	Aktivitäten des allgemeinen Projektmanagements. Meist wenig BI spezifisch. Beispiel: Projektstatusüberwachung
Fachlichkeit	Aktivitäten im direkten Zusammenhang mit den fachlichen Zielen des BI Projektes. Beispiel: Definition und Detailanalyse der analytischen Anforderungen
BI Anwendungen	Aktivitäten zur Realisierung der verschiedenen Benutzerschnittstellen (Front End). Beispiel: Installation und Konfiguration BI Anwendungen
Daten	Aktivitäten zur Modellierung und Implementierung der verschiedenen logischen und physikalischen Datenmodelle. Beispiel: Detaildesign Faktentabelle
ETL	Aktivitäten im Umfeld der ETL Prozesse. Beispiel: Erstellen Source-to-Target Map

Disziplinen	Beschreibung
Metadaten	Aktivitäten zu Entwurf und Umsetzung eines Metadatenmodells. Beispiel: Inventur Metadaten
Architektur & Infrastruktur	Aktivitäten im Zusammenhang mit der Bereitstellung einer geeigneten Architektur und Infrastruktur. Beispiel: Design Infrastruktur

Tabelle 2: Grundstruktur des BI Vorgehensmodells - Disziplinen

Somit lassen sich die zu definierenden Aktivitäten eindeutig einer Iteration, Phase und Disziplin zuordnen (siehe Abbildung 1).

		Phasen							Iteration N
		Phasen							Iteration 1
		Initialisierung	Analyse	Design	Realisierung	Test	Deployment	Produktion	
Disziplinen	Allg. PM								
	Fachlichkeit								
	BI Anwendungen								
	Daten			Aktivität					
	ETL								
	Metadaten								
	A&I								

Abbildung 1: Zuordnung einer Aktivität zu Iteration, Phase und Disziplin

Die einzelnen Aktivitäten werden durch die in Tabelle 3 dargestellten Attribute spezifiziert. Dabei gibt es einen iterationsunabhängigen ersten und einen iterationsabhängigen zweiten Teil (die letzten drei Attribute). Der iterationsabhängige zweite Teil erlaubt es, im Sinne der zweiten Ausgangsfrage (Was ist eine zur Dokumentation der Projektergebnisse geeignete Darstellungsform?) auch Ergebnisse der Iteration zu dokumentieren (Details siehe folgende Abschnitte).

Im Rahmen dieser Untersuchung werden die Grundstruktur des BI Vorgehensmodells, alle Aktivitäten und eine beispielhafte Spezifizierung aller Attribute einer Aktivität (Erstellen Source-to-Target Map) definiert. Eine vollständige Spezifizierung aller Aktivitäten ist erst im späteren unternehmensspezifischen Praxiseinsatz unter Berücksichtigung kultureller und organisatorischer Gegebenheiten vorgesehen.

Aktivität
Name
ID
Phase
Disziplin
Beschreibung
Beteiligte Rollen (RACI) <sup>9</sup>
Eingangsdokumente
Methoden
Templates
Werkzeuge
Tipps und Hinweise
Referenzen
Iteration
Ergebnisse
Ergebnisdokumente

Tabelle 3: Attribute der Aktivitäten

## 5 Aktivitäten des BI Vorgehensmodells

In den Tabellen 4 bis 7 sind die Aktivitäten der jeweiligen Phasen und Disziplinen aufgeführt. Bei der Auswahl der Aktivitäten auf Basis der veröffentlichten Literatur und beobachteten Praktiken wurde insbesondere auf einen sinnvollen Detailgrad und Vollständigkeit geachtet. So soll der Projektmanager in dem Vorgehensmodell auf einen Blick alle wesentlichen Aktivitäten erkennen können. Damit soll verhindert werden, dass wesentliche Aktivitäten versehentlich vergessen oder erst zu spät erkannt werden. Andererseits sollte der Detailgrad nicht zu gering sein, damit Spezifika von BI Projekten klar erkennbar werden. Wie bereits erwähnt, können je nach Projektausrichtung, Projektgröße und Iteration einzelne Aktivitäten entfallen bzw. müssen modifiziert werden. Komplexität und Aufwände der verschiedenen Aktivitäten sind unterschiedlich. Dies muss bei einer

---

<sup>9</sup> RACI: Responsible, Accountable, Consulted, Informed

späteren vollständigen Spezifizierung aller Attribute der einzelnen Aktivitäten berücksichtigt werden.

	<b>Initialisierung</b>	<b>Analyse</b>
<b>Ziele</b>	<ul style="list-style-type: none"> <li>• Definition des BI Projektes</li> </ul>	<ul style="list-style-type: none"> <li>• Detailanalyse fachliche und technische Anforderungen</li> <li>• Erstellung Projektplan</li> </ul>
<b>Meilensteine</b>	<ul style="list-style-type: none"> <li>• Abnahme Projektdefinition</li> <li>• Projektfreigabe</li> </ul>	<ul style="list-style-type: none"> <li>• Abnahme Pflichtenheft</li> <li>• Abnahme Projektplan</li> </ul>
<b>Allgemeines Projektmanagement</b>	<ul style="list-style-type: none"> <li>• Identifikation Projektziele</li> <li>• Auswahl Projektmethodik</li> <li>• Analyse Stakeholder</li> <li>• Erstellung Business Case</li> <li>• Entwurf Projektorganisation</li> <li>• Identifikation Erfolgskriterien</li> <li>• Skizze Projektplan</li> </ul>	<ul style="list-style-type: none"> <li>• Erstellung Projektplan</li> <li>• Auswahl Projektmanagementwerkzeug</li> <li>• Implementierung Projektorganisation und -prozesse</li> <li>• Projektdokumentation</li> <li>• Projektmarketing</li> <li>• Projektstatusüberwachung</li> <li>• Projektcontrolling</li> <li>• Change Management</li> <li>• Risiko Management</li> <li>• Qualitätsmanagement</li> </ul>
<b>Fachlichkeit</b>	<ul style="list-style-type: none"> <li>• Definition BI Strategie</li> <li>• Analyse BI Reifegrad</li> <li>• Definition analytische Anforderungen</li> <li>• Identifikation einzubeziehende Schlüsselprozesse</li> <li>• Identifikation Auswirkungen auf Organisation</li> <li>• Machbarkeitsanalyse</li> </ul>	<ul style="list-style-type: none"> <li>• Informationsbedarfsanalyse</li> <li>• Detailanalyse analytische Anforderungen</li> <li>• Homogenisierung Begrifflichkeiten</li> <li>• Definition Service Level Anforderungen</li> <li>• Validierung und Priorisierung der Anforderungen</li> <li>• Entwurf phasenweise Einführungsstrategie</li> <li>• Entwurf Organisationskonzept</li> </ul>
<b>BI Anwendungen</b>		<ul style="list-style-type: none"> <li>• Anforderungsanalyse Informationsservices</li> <li>• Analyse Benutzerverhalten</li> <li>• Entwurf Prototyp</li> <li>• Implementierung Prototyp</li> </ul>
<b>Daten</b>	<ul style="list-style-type: none"> <li>• Grobes Data Profiling</li> </ul>	<ul style="list-style-type: none"> <li>• Analyse Datenanforderungen</li> <li>• Data Profiling</li> <li>• Semantische Datenmodellierung</li> <li>• Skizze logische Datenmodellierung</li> <li>• Entwurf Datenarchitektur</li> <li>• Definition Datenbank Standards</li> </ul>
<b>ETL</b>		<ul style="list-style-type: none"> <li>• Anforderungsanalyse ETL Prozess</li> <li>• Erstellen Source-to-Target Map</li> <li>• Entwurf ETL Data Stores</li> <li>• Auswahl ETL Tool</li> </ul>

	Initialisierung	Analyse
<b>Metadaten</b>		<ul style="list-style-type: none"> <li>• Anforderungsanalyse Metadaten</li> <li>• Inventur Metadaten</li> <li>• Auswahl Metadaten Management Tool</li> </ul>
<b>Architektur &amp; Infrastruktur</b>	<ul style="list-style-type: none"> <li>• A&amp;I Vision</li> <li>• Übersicht A&amp;I (Ist)</li> <li>• Machbarkeitsanalyse A&amp;I</li> <li>• Einrichtung A&amp;I Task Force</li> </ul>	<ul style="list-style-type: none"> <li>• Anforderungsanalyse BI Anwendungsarchitektur</li> <li>• Design BI Anwendungsarchitektur</li> <li>• Design Infrastruktur</li> <li>• Definition Standards</li> <li>• A&amp;I Produktauswahl</li> </ul>

Tabelle 4: Aktivitäten der Phasen Initialisierung und Analyse

	Design	Realisierung
<b>Ziele</b>	<ul style="list-style-type: none"> <li>• Detailentwurf BI Anwendungen, Datenbanken und ETL Prozesse</li> </ul>	<ul style="list-style-type: none"> <li>• Realisierung der fachlichen und technischen Feinkonzepte bis zur Implementierungsreife</li> </ul>
<b>Meilensteine</b>	<ul style="list-style-type: none"> <li>• Abnahme fachliches und technisches Feinkonzept</li> <li>• Abnahme A&amp;I Plattform</li> </ul>	<ul style="list-style-type: none"> <li>• Abnahme Unit Test BI Anwendungen, Datenbanken und ETL Prozesse</li> </ul>
<b>Allgemeines Projektmanagement</b>	<ul style="list-style-type: none"> <li>• Aktualisierung Projektplan</li> <li>• Projektstatusüberwachung</li> <li>• Validierung Projektmanagementprozesse</li> <li>• Projektcontrolling</li> <li>• Projektdokumentation</li> <li>• Beschaffung</li> <li>• Projektmarketing</li> <li>• Change Management</li> <li>• Risiko Management</li> <li>• Qualitätsmanagement</li> </ul>	<ul style="list-style-type: none"> <li>• Aktualisierung Projektplan</li> <li>• Projektstatusüberwachung</li> <li>• Validierung Projektmanagementprozesse</li> <li>• Projektcontrolling</li> <li>• Projektdokumentation</li> <li>• Beschaffung</li> <li>• Projektmarketing</li> <li>• Change Management</li> <li>• Risiko Management</li> <li>• Qualitätsmanagement</li> <li>• Entwurf Testkonzept</li> <li>• Entwurf Schulungskonzept</li> </ul>
<b>Fachlichkeit</b>		
<b>BI Anwendungen</b>	<ul style="list-style-type: none"> <li>• Feinkonzept BI Anwendungen</li> <li>• Auswahl BI Produkte</li> </ul>	<ul style="list-style-type: none"> <li>• Installation und Konfiguration BI Anwendungen</li> </ul>

<b>Daten</b>	<ul style="list-style-type: none"> <li>• Logische Datenmodellierung</li> <li>• Detaildesign Faktentabellen</li> <li>• Detaildesign Dimensionstabellen</li> <li>• Design OLAP Datenbank</li> <li>• Design physikalische Datenbanken</li> <li>• Auswahl RDBMS und OLAP Produkte</li> <li>• Design ODS<sup>10</sup> und Data Marts</li> <li>• Design Processing Data Stores</li> <li>• Konzept Stammdaten</li> <li>• Data Aging Konzept</li> </ul>	<ul style="list-style-type: none"> <li>• Realisierung logische und physikalische Datenmodelle</li> <li>• Erstellen Build Skripte</li> <li>• Implementierung Testdatenbank</li> </ul>
<b>ETL</b>	<ul style="list-style-type: none"> <li>• Finalisierung Source-to-Target Map</li> <li>• Design historisches und inkrementelles Laden</li> <li>• Detaildesign Extraktion</li> <li>• Detaildesign Transformation</li> <li>• Detaildesign Laden</li> <li>• Design ETL Data Stores</li> <li>• Design ETL Management Prozesse</li> </ul>	<ul style="list-style-type: none"> <li>• Realisierung ETL Prozesse</li> </ul>
<b>Metadaten</b>	<ul style="list-style-type: none"> <li>• Feinkonzept Metadaten</li> <li>• Installation Metadaten Management Tool</li> <li>• Test Metadaten Management Tool</li> </ul>	<ul style="list-style-type: none"> <li>• Konfiguration Metadaten Repository</li> <li>• Realisierung Ladeprozesse fachliche und technische Metadaten</li> <li>• Laden fachliche und technische Metadaten</li> </ul>
<b>Architektur &amp; Infrastruktur</b>	<ul style="list-style-type: none"> <li>• Beschaffung und Installation A&amp;I Produkte</li> <li>• Realisierung und Test A&amp;I</li> </ul>	<ul style="list-style-type: none"> <li>• Bereitstellung Entwicklungsumgebungen</li> <li>• Bereitstellung Versionskontrollsystem</li> <li>• Entwurf technisches Migrationskonzept</li> </ul>

*Tabelle 5: Aktivitäten der Phasen Design und Realisierung*

---

<sup>10</sup> Operational Data Store



	Test	Deployment
<b>Ziele</b>	<ul style="list-style-type: none"> <li>Durchführung der verschiedenen Testarten</li> </ul>	<ul style="list-style-type: none"> <li>Vorbereitung Produktion</li> <li>Deployment der fertigen BI Lösung in die bestehende IT Umgebung</li> </ul>
<b>Meilensteine</b>	<ul style="list-style-type: none"> <li>Fachliche und technische Abnahme</li> <li>Produktionsfreigabe</li> </ul>	<ul style="list-style-type: none"> <li>Start Produktion</li> </ul>
<b>Allgemeines Projektmanagement</b>	<ul style="list-style-type: none"> <li>Aktualisierung Projektplan</li> <li>Projektstatusüberwachung</li> <li>Validierung Projektmanagementprozesse</li> <li>Projektcontrolling</li> <li>Projektdokumentation</li> <li>Beschaffung</li> <li>Projektmarketing</li> <li>Change Management</li> <li>Risiko Management</li> <li>Qualitätsmanagement</li> <li>Definition Testansatz</li> <li>Planung, Vorbereitung und Durchführung Testarten</li> <li>Auswahl und Implementierung Testmanagement Tool</li> <li>Priorisierung Fehlerbehebung</li> <li>Dokumentation Testergebnisse</li> </ul>	<ul style="list-style-type: none"> <li>Aktualisierung Projektplan</li> <li>Projektstatusüberwachung</li> <li>Validierung Projektmanagementprozesse</li> <li>Projektcontrolling</li> <li>Projektdokumentation</li> <li>Beschaffung</li> <li>Projektmarketing</li> <li>Change Management</li> <li>Risiko Management</li> <li>Qualitätsmanagement</li> <li>Planung Betrieb und Wartung</li> <li>Fertigstellung Betriebsführungshandbuch</li> <li>Qualifizierung Personal</li> </ul>
<b>Fachlichkeit</b>	<ul style="list-style-type: none"> <li>Erstellen fachliche Testfälle</li> <li>Definition Testdaten</li> <li>Definition Abnahmekriterien</li> </ul>	<ul style="list-style-type: none"> <li>Umstellung Prozesse und Organisation</li> <li>Einrichtung BI Governance Strukturen (BICC<sup>11</sup>)</li> </ul>
<b>BI Anwendungen</b>	<ul style="list-style-type: none"> <li>Test Benutzerschnittstellen</li> <li>Vorbereitung und Durchführung Training Anwender</li> <li>Erstellen Anwenderdokumentation</li> </ul>	<ul style="list-style-type: none"> <li>Fertigstellung Drehbuch Deployment BI Anwendungen</li> <li>Deployment BI Anwendungen</li> <li>Anwendersupport</li> </ul>
<b>Daten</b>	<ul style="list-style-type: none"> <li>Bereitstellung Testdaten</li> <li>Fachliche und technische Tests Datenbanken</li> <li>Test Datenqualität</li> </ul>	<ul style="list-style-type: none"> <li>Fertigstellung Drehbuch Deployment RDBMS und OLAP</li> <li>Deployment RDBMS und OLAP</li> </ul>
<b>ETL</b>	<ul style="list-style-type: none"> <li>Fachliche und technische Tests der ETL Prozesse</li> </ul>	<ul style="list-style-type: none"> <li>Fertigstellung Drehbuch Deployment ETL Prozess</li> <li>Deployment ETL Prozess</li> <li>Historisches Laden</li> <li>Einrichtung Regelbetrieb</li> </ul>
<b>Metadaten</b>	<ul style="list-style-type: none"> <li>Test Metadaten Repository</li> </ul>	<ul style="list-style-type: none"> <li>Fertigstellung Drehbuch Deployment Metadaten</li> <li>Deployment Metadaten Repository</li> </ul>

---

<sup>11</sup> BI Competence Center

	Test	Deployment
<b>Architektur &amp; Infrastruktur</b>	<ul style="list-style-type: none"> <li>• Bereitstellung Testumgebungen</li> <li>• Technische Tests</li> <li>• Test PC Anforderungen und Konfigurationen</li> </ul>	<ul style="list-style-type: none"> <li>• Fertigstellung Drehbuch Deployment A&amp;I</li> <li>• Deployment A&amp;I</li> <li>• Bereitstellung Produktions- und Wartungsumgebung</li> <li>• Einrichtung Regelbetrieb</li> </ul>

*Tabelle 6: Aktivitäten der Phasen Test und Deployment*

	Produktion
<b>Ziele</b>	<ul style="list-style-type: none"> <li>• Stabile Produktion, Wachstum, Erweiterungen</li> </ul>
<b>Meilensteine</b>	<ul style="list-style-type: none"> <li>• Definition Folgeprojekt</li> </ul>
<b>Allgemeines Projektmanagement</b>	<ul style="list-style-type: none"> <li>• Stabilisierung Produktion</li> <li>• Erfolgskontrolle</li> <li>• Dokumentation Lessons Learned</li> <li>• Erstellung Projektabschlussbericht</li> <li>• Change Management</li> <li>• Fortführung Projektmarketing</li> <li>• Definition Folgeprojekt</li> </ul>
<b>Fachlichkeit</b>	<ul style="list-style-type: none"> <li>• Priorisierung Erweiterungsmöglichkeiten</li> </ul>
<b>BI Anwendungen</b>	<ul style="list-style-type: none"> <li>• BI IT Service Management (ITIL)</li> <li>• (Exception) Monitoring</li> <li>• Anwenderunterstützung</li> <li>• Identifikation Verbesserungs- und Erweiterungsmöglichkeiten</li> <li>• Fortführung Anwendertraining</li> <li>• Anwendungswartung</li> </ul>
<b>Daten</b>	<ul style="list-style-type: none"> <li>• Wartung Datenbanken</li> </ul>
<b>ETL</b>	<ul style="list-style-type: none"> <li>• Überwachung und Stabilität ETL Prozess</li> </ul>
<b>Metadaten</b>	<ul style="list-style-type: none"> <li>• Überwachung Metadaten</li> </ul>
<b>Architektur &amp; Infrastruktur</b>	<ul style="list-style-type: none"> <li>• Überwachung Security und Compliance</li> <li>• Performancemonitoring und -tuning</li> <li>• Upgrades</li> <li>• Regelmäßige Backups</li> <li>• Archivierung</li> <li>• Identifikation Erweiterungen Infrastruktur</li> <li>• Pflege Infrastrukturkarte</li> </ul>

*Tabelle 7: Aktivitäten der Phase Produktion*

Exemplarisch ist in Tabelle 8 die Aktivität *Erstellen Source-to-Target Map* der Phase *Analyse* und Disziplin *ETL* gemäß den Attributen aus Tabelle 3 spezifiziert.

<b>Name</b>	<b>Erstellen Source-to-Target Map</b>
<b>ID<sup>12</sup></b>	AN.ET.2
<b>Phase</b>	Analyse
<b>Disziplin</b>	ETL
<b>Beschreibung</b>	<p>Die Source-to-Target Map gibt einen Gesamtüberblick über die ETL Prozesse und den daran beteiligten operativen Datenquellen und BI Zieldatenbanken. Als wesentliche Grundlage für das folgende ETL Design vermindert sie das Risiko nicht erkannter gravierender Quelldatenprobleme.</p> <p>Die Source-to-Target Map enthält alle im Umfeld der ETL Prozesse erforderlichen wesentlichen Angaben über technische Plattformen, Tabellen, Datenformate, Extraktions- und Transformationsregeln, Abhängigkeiten, Mengengerüste und offene Fragen. Sie eignet sich insbesondere auch, um die Komplexität der ETL Prozesse dem Management zu verdeutlichen.</p>
<b>Beteiligte Rollen (RA-CI)</b>	Projektmanager (Informed, Accountable) ETL Architekt (Responsible) Fachlicher Ansprechpartner (Consulted) Datenarchitekt (Consulted) Datenbankadministrator (Consulted) Technischer Architekt (Consulted) Datenqualitätsmanager (Consulted)
<b>Eingangsdokumente<sup>5</sup></b>	IN.DA.1 (Ergebnis Grobes Data Profiling) IN.FA.4 (Dokumentation einzubeziehender Schlüsselprozesse)
<b>Methoden</b>	<ul style="list-style-type: none"> <li>• Identifikation aller wesentlichen Angaben über technische Plattformen, Tabellen, Datenformate, Extraktions- und Transformationsregeln, Abhängigkeiten, Mengengerüste und offenen Fragen.</li> <li>• Grafische und/oder tabellarische Aufbereitung der Ergebnisse in der Source-to-Target Map.</li> </ul>
<b>Templates<sup>5</sup></b>	AN.ETL.2_Sample_Source_to_Target_Map.ppt AN.ETL.2_Sample_Source_to_Target_Map.xls
<b>Werkzeuge</b>	Microsoft Office SQL Data Profiling Tool
<b>Tipps und Hinweise</b>	<ul style="list-style-type: none"> <li>• Bei der Quelldatenanalyse ggf. auch auf den Quellcode der Anwendungen zurückgreifen.</li> <li>• Archivierte Produktionsdaten können ältere Datenstrukturen aufweisen.</li> <li>• Der Fokus sollte auf den Daten liegen, die im direkten Zusammenhang mit den fachlichen Anforderungen an die BI Lösung benötigt werden.</li> <li>• SQL Kenntnisse helfen, um sich einen schnellen Überblick über die Quellsysteme zu verschaffen.</li> <li>• Auch auf Ausnahmen und fehlende Daten achten (insbesondere in Bezug auf einen unzureichenden Detailgrad).</li> <li>• Offene Fragen nicht in das Data Warehouse Umfeld verlagern. Sie müssen vor dem ETL Design geklärt werden.</li> <li>• Die Source-to-Target Map ist unabhängig vom ETL Tool.</li> <li>• Es kann sinnvoll sein, zwei Versionen der Source-to-Target Map vorzuhalten: Eine einfache für die externe Kommunikation und eine detaillierte für das Projektteam.</li> <li>• Die Unterstützung durch die Eigentümer der Anwendungen, die die Daten bereitstellen sollen, ist erfolgskritisch.</li> </ul>

<sup>12</sup> Die Kürzel ergeben sich aus den Anfangsbuchstaben der Phase und Disziplin und dem Listenplatz der Aktivität in den Tabellen 4 bis 7 (und bei den Ergebnisdokumenten zusätzlich der Iterationsangabe).

Name	Erstellen Source-to-Target Map
Referenzen	Kimball, R. (2008). The Data Warehouse Lifecycle Toolkit. Indianapolis, Ind.: Wiley. (Chapter 10, Seite 428 - 429) Moss, L.T., & Atre, S. (2006). Business intelligence roadmap: The complete project lifecycle for decision-support applications. Boston, Mass.: Addison-Wesley. (Chapter 9, Seite 225 - 229)
Iteration	1
Ergebnisse	Source-to-Target Map, die als zentrale Basis für das folgende Design der ETL Prozesse (DE.ET) dient und in diesem Zusammenhang zu Beginn der Designphase noch einmal finalisiert wird.
Ergebnisdokumente <sup>5</sup>	AN.ETL.2.1_Source_to_Target_Map.xls

Tabelle 8: Spezifikation der Aktivität Erstellen Source-to-Target Map

## 6 Einsatz von Wiki's im Projektmanagement

Die Darstellungsform des anwendungsorientierten Vorgehensmodells für den praktischen Einsatz in BI Projekten sollte ein möglichst einfaches und intuitives Wissensmanagement über mehrere Iterationen ermöglichen. Dies gilt insbesondere für die Weiterentwicklung des Vorgehensmodells selber. Es bietet sich aber aufgrund der engen Verzahnung an, auch die Dokumentation der Projektergebnisse vorzusehen. Dann können Folgeiterationen optimal auf einem praxiserprobten Vorgehensmodell und den bisherigen Ergebnissen aufsetzen.

Unter dem Oberbegriff Web 2.0 werden Dienste zusammengefasst, die es Nutzern ermöglichen, aktiv an der Erstellung von Inhalten im Web mitzuwirken. Dabei ist die einfache Benutzbarkeit ein wesentliches Kriterium, um eine möglichst große freiwillige Nutzerbeteiligung zu erreichen (Koch et al., 2007). Web 2.0 Dienste werden zunehmend auch für das Wissensmanagement in Unternehmen eingesetzt (Dueck, 2008; Richter et al., 2009).

Insbesondere Wikis (Ebersbach et al., 2005) bieten sich an, um Informationen einfach und intuitiv einer möglichst großen Zielgruppe zugänglich zu machen (Blaschke, 2008; Koch et al., 2009). Dabei wird unter einem Wiki „eine frei editierbare Sammlung verlinkter Webseiten“ (Blaschke, 2008) verstanden. Eine detaillierte Untersuchung über den Einsatz von Wikis als Kommunikationsinstrument für das Projektmanagement findet sich in (Happ et al., 2006). Dabei wurde ein Wiki in einem geografisch verteilten Projekt erfolgreich bei der gemeinsamen Erstellung strukturierter Projektdokumente und einer Wissensdatenbank eingesetzt. Als wesentliche Erfolgsfaktoren wurden eine kritische Masse an Inhalten, ein einfaches Rollenmodell und die Vorgabe einer klaren Grundstruktur identifiziert. Auch umfangreiche Suchfunktionen, RSS Feeds und verfügbare Änderungshistorien wurden als sehr hilfreiche Grundfunktionalitäten eines Wikis erkannt.

Somit stellen Wikis einen erfolgsversprechenden Ansatz für die Darstellung von Vorgehensmodellen dar.

## 7 Das Wiki-basierte Vorgehensmodell für Business Intelligence Projekte

Im folgenden wird ein Vorschlag für ein Wiki-basiertes Vorgehensmodell für BI Projekte auf Basis der Ergebnisse aus den Abschnitten 4 und 5 vorgestellt.

Zum Betrieb des Wikis wurde aufgrund seiner starken Verbreitung das frei verfügbare (GPL Lizenz) Tool MediaWiki<sup>13</sup> ausgewählt.



Abbildung 2: Einstiegsseite in das Wiki-basierte Vorgehensmodell für BI Projekte

In Abbildung 2 ist die Einstiegsseite in das Wiki-basierte Vorgehensmodell für BI Projekte dargestellt. Die Auswahl einer Projektphase bzw. Disziplin führt zu einer Übersichtsseite aller für diese Projektphase bzw. Disziplin relevanten Aktivitäten. Die anschließenden

<sup>13</sup> [www.mediawiki.org](http://www.mediawiki.org)

de Auswahl einer einzelnen Aktivität öffnet eine Seite mit der Detailbeschreibung der entsprechenden Aktivität gemäß dem Template aus Tabelle 3. Das Beispiel der Aktivität *Erstellen Source-to-Target Map* (siehe Tabelle 8) ist in Abbildung 3 dargestellt.

**Vorgehensmodell für BI-Projekte** [Wiki](#) [Eigene Diskussion](#) [Einstellungen](#) [Beobachtungsliste](#) [Eigene Beiträge](#) [Abmelden](#)

Navigation: [BI-PM Wiki](#), [1. Initialisierung](#), [2. Analyse](#), [3. Design](#), [4. Realisierung](#), [5. Test](#), [6. Deployment](#), [7. Produktion](#), [Hilfe](#)

Suche:  [Seite](#) [Suchen](#)

Werkzeuge: [Links auf diese Seite](#), [Änderungen an verlinkten Seiten](#), [Hochladen](#), [Spezialseiten](#), [Druckversion](#), [Permanenter Link](#)

**Erstellen Source-to-Target Map**

Name	Erstellen Source-to-Target Map
ID	AN.ET.2
Phase	Analyse
Disziplin	ETL
Iteration	1

**Beschreibung** [\[Bearbeiten\]](#)

Die Source-to-Target Map gibt einen Gesamtüberblick über die ETL Prozesse und den daran beteiligten operativen Datenquellen und BI Zieldatenbanken. Als wesentliche Grundlage für das folgende ETL Design vermindert sie das Risiko nicht erkannter gravierender Quelldatenprobleme. Die Source-to-Target Map enthält alle im Umfeld der ETL Prozesse erforderlichen wesentlichen Angaben über technische Plattformen, Tabellen, Datenformate, Extraktions- und Transformationsregeln, Abhängigkeiten, Mengengerüste und offene Fragen. Sie eignet sich insbesondere auch, um die Komplexität der ETL Prozesse dem Management zu verdeutlichen.

**Beteiligte Rollen (RACI)** [\[Bearbeiten\]](#)

Abbildung 3: Spezifikation der Aktivität *Erstellen Source-to-Target Map* im Wiki

## 8 Ausblick

Nach einer Validierung des Wiki-basierten BI Vorgehensmodells in der praktischen Projektarbeit mit anschließender Vervollständigung aller Attribute zu allen Aktivitäten wäre es interessant, die Ergebnisse in einem Wiki-basierten BI-PMBok<sup>14</sup> zur Verfügung zu stellen. Durch Anwendung der so dokumentierten und frei zugänglichen bewährten Vorgehensweisen könnte die Qualität zukünftiger BI Projekte trotz komplexer werdenden Anforderungen weiter gesteigert und damit die Akzeptanz von BI Lösungen insgesamt verbessert werden

<sup>14</sup> In Anlehnung an den Project Management Body of Knowledge (PMBok) des Project Management Institutes.

## 9 Literatur

Anahory, S., & Murray, D. (1997). *Data Warehouse: Planung, Implementierung und Administration*. Bonn: Addison-Wesley-Longman.

Bauer, A., & Günzel, H. (2008). *Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung*. Heidelberg: dpunkt-Verlag.

Becker, J., & Knacktedt, R. (2004). Referenzmodellierung im Data-Warehousing. *WIRTSCHAFTSINFORMATIK*, 46(1), 39-49.

Becker, J., Maßing, D., & Janiesch, C. (2006). Ein evolutionäres Vorgehensmodell zur Einführung von Corporate Performance Management Systemen. *Data Warehousing*, 2006, 247-262.

Blaschke, S. (2008). Wikis in Organisationen: Von Kommunikation zu Kollaboration. In Alpar, P., & Blaschke, S. (Hrsg.), *Web 2.0: Eine empirische Bestandsaufnahme* (pp. 184-203). Wiesbaden: Vieweg + Teubner.

Breslin, M. (2004). DWH Battle of the Giants: Comparing the Basics of the Kimball and Inmon Modells. *Business Intelligence Journal*, 9(1).

Brobst, S., McIntire, M., & Rado, E. (2009). Agile Data Warehousing with Integrated Sandboxing. *Business Intelligence Journal*, 13(1).

Chenoweth, T. (2006). Seven Key Interventions for Data Warehouse Success. *Communications of the ACM*, 49(1), 114-119.

Dueck, G. (2008). Bluepedia. *Informatik Spektrum*, 31(3), 262-269.

Ebersbach, A., & Glaser, M. (2005). Wiki. *Informatik Spektrum* 28(2), 131-135.

Füting, U. C. (2000). Projektmanagement und -controlling von DWH Projekten. In Mucksch, H., & Behme, W. (Hrsg.), *Das Data Warehouse Konzept* (pp. 269-289). Wiesbaden: Gabler.

Gardner, S. (1998). Building the DWH. *Communications of the ACM*, 41(9), 52-60.

Giorgini, P. (2005). Goal-Oriented Requirement Analysis for DWH Design. *Data Warehousing and OLAP, Proceedings of the 8th ACM international workshop on Data warehousing and OLAP*.

Gluchowski, P., Gabriel, R., & Dittmar, C. (2008). *Management Support Systeme und Business Intelligence: Computergestützte Informationssysteme für Fach- und Führungskräfte*. Berlin: Springer.

Goeken, M. (2004). Referenzmodellbasierte Einführung von Führungsinformationssystemen. *WIRTSCHAFTSINFORMATIK*, 46(1), 353-365.

Goeken, M. (2007). *Entwicklung von Data-Warehouse-Systemen: Anforderungsmanagement, Modellierung, Implementierung*. Wiesbaden: Deutscher Universitäts-Verlag.

Golfarelli, M., & Rizzi, S. (1998). A methodological framework for data warehouse design. *Data Warehousing and OLAP. Proceedings of the 1st ACM international workshop on Data warehousing and OLAP*.

Happ, S., Wünsche, A., & Henkel, F. (2006). Blogs, Wikis, Webcasts: Utilization of State-of-the-Art Communication Instruments for Project Management. *PMI Global Congress 2006-EMEA. Proceedings*. Project Management Institute.

Hecht, H. (2001). *Einführung und kontinuierliche Adaption von betriebswirtschaftlichen Data-Warehouse-Bibliotheken*. Dissertation. Universität Würzburg.

Heck-Weinhart, G., Mutterer, G., Herrmann, C., & Rupprecht, J. (2003). Entwicklung eines angepassten Vorgehensmodells für Data-Warehouse-Projekte bei der W&W AG. In von Maur, E., & Winter, R. (Hrsg.), *Data Warehouse Management: Das St. Galler Konzept zur ganzheitlichen Gestaltung der Informationslogistik* (pp. 197-220). Heidelberg: Springer.

Herden, O. (2001). *Entwurfsmethodik für DWH*. Dissertation. Universität Oldenburg.

Höhn, R. (2000). *Der Data Warehouse Spezialist: Entwurf, Methoden und Umsetzung eines Data Warehouses*. München: Addison-Wesley.

Holten, R. (1999). *Entwicklung von Führungsinformationssystemen: Ein methodenorientierter Ansatz*. Wiesbaden: Deutscher Universitäts-Verlag.

Hughes, R. (2008). *Agile Data Warehousing: Delivering World-Class Business Intelligence Systems using Scrum and XP*. Bloomington, Ind.: iUniverse.

Inmon, W. H. (2005). *Building the Data Warehouse*. Indianapolis, Ind.: Wiley.

Kemper, H. G., Mehanna, W., & Unger, C. (2006). *Business Intelligence: Grundlagen und praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung*. Wiesbaden: Vieweg.

Keppel, S., Müllenbach, S., & Wölkhammer, M. (2001). Vorgehensmodelle im Bereich Data Warehouse: Das Evolutionary Data Warehouse Engineering (EDE). In Schütte, R., Rotthowe, T., & Holten, R. (Hrsg.), *Data Warehouse Management Handbuch* (pp. 81-106). Berlin: Springer.

Kimball, R. (2008). *The Data Warehouse Lifecycle Toolkit*. Indianapolis, Ind.: Wiley.

Koch, M., Richter, A., & Schlosser, A. (2007). Produkte zum IT-gestützten Social Networking im Unternehmen. *WIRTSCHAFTSINFORMATIK* 49(6), 448-455.

Koch, M., Ott, F., & Richter, A. (2009). Wikis und Weblogs im Wissens- und Innovationsmanagement. *HMD Praxis der Wirtschaftsinformatik*, 267, 47-55.

Konzelmann, R. (2008). Vorgehensmodell zur Erstellung eines Enterprise Data Warehouse. In Töpfer, J., & Winter (Hrsg.), R., *Active Enterprise Intelligence: Unternehmensweite Informationslogistik als Basis einer wertorientierten Unternehmenssteuerung* (pp. 273-312). Berlin: Springer.

Moss, L.T., & Atré, S. (2006). *Business intelligence roadmap: The complete project lifecycle for decision-support applications*. Boston, Mass.: Addison-Wesley.

Oehler, K. (2006). *Corporate Performance Management mit Business-Intelligence-Werkzeugen*. München: Carl Hanser.

Reinschmidt, J., & Allison, F. (2000). *Business Intelligence Certification Guide*. San Jose, Cal.: IBM Redbook.

Richter, A., Kneifel, D., & Ott, F. (2009). Fallstudie: Social Networking bei Accenture. *Wirtschaftsinformatik & Management*, 2009(1), 78-81.

Sammon, D., & Finnegan, P. (2000). The Ten Commandments of DWH. *The DATA BASE for Advances in Information Systems*, 31(4), 82-91.

Schirp, G. (2001). *Anforderungsanalyse im Data-Warehouse-Projekt: Ein Erfahrungsbericht aus der Praxis*. *HMD Praxis der Wirtschaftsinformatik*, 222, 81-87.

Schrödl, H. (2009). *Business Intelligence mit Microsoft SQL Server 2008: BI-Projekte erfolgreich umsetzen*. München: Carl Hanser.



Sen, A., & Sinha, A. P. (2005). A Comparison of Data Warehousing Methodologies. *Communications of the ACM*, 48(3), 79-84.

Staade, M., & Schüler, B. (2006). *SAP BI-Projektmanagement*. Bonn: Galileo-Press.

Strauch, B., & Winter, R. (2002). Vorgehensmodell für die Informationsbedarfsanalyse im Data Warehousing. In von Maur, E., & Winter, R. (Hrsg.), *Vom Data Warehouse zum Corporate Knowledge Center* (pp. 359-378). Heidelberg: Physica.

Turban, E., Aronson, J. E., Liang, T. P., & Sharda, R. (2006). *Decision Support and Business Intelligence Systems*. Upper Saddle River, NJ: Pearson/Prentice Hall.

Westerman, P. (2000). *Data Warehousing*. Frankfurt am Main: Deutscher Fachverlag.



# **Semantische Anreicherung von strukturierten Daten und Prozessen in analytischen Informationssystemen am Beispiel von MUSTANG**

Matthias Mertens<sup>1</sup>, Yvette Teiken<sup>1</sup>, Jürgen Appelrath<sup>2</sup>

<sup>1</sup> OFFIS - Institut für Informatik Oldenburg

<sup>2</sup> Universität Oldenburg, Praktische Informatik, Abteilung Informationssysteme

## **1 Einleitung**

Analytische Informationssysteme ermöglichen es multidimensionale Ad-Hoc-Analysen mittels On-Line Analytical Processing (OLAP) Operationen auf quantitativen Daten eines Data-Warehouses durchzuführen. Hierbei werden die Auswertungen oftmals in Form von Berichten und zugehörigen Berichts-elementen – Tabellen, Diagrammen, Karten, etc. – visualisiert, gespeichert und gegebenenfalls mit zusätzlichen unstrukturierten Informationen zur besseren Verständlichkeit beschrieben. Diese Berichte können im Sinne einer Informationslogistik dazu dienen, betriebliche Fach- und Führungskräfte zeitnah mit allen relevanten Informationen über interne und externe Sachverhalte und deren Zusammenhängen zu versorgen und sie somit bei der Entscheidungsfindung zu unterstützen (Gluchowski, 2006; Winter, 2008). Eine wesentliche Rolle sollte jedoch auch die Dokumentation von Prozessen und Ergebnissen in analytischen Informationssystemen einnehmen. Dies umfasst sowohl die zugrunde liegenden Fragestellungen als auch die Analyse-durchführung und die Ergebnissicherung. So stellen sich die Fragen: Welche Analysten haben Ad-Hoc Analysen zu spezifischen Fragestellungen durchgeführt? Welche Verfahren, Parameter und OLAP-Operationen wurden im Rahmen dieser Analysen in den einzelnen Analyseschritten auf den quantitativen Daten durchgeführt? Welche Erkenntnisse und Zusammenhänge wurden in den quantitativen Daten gewonnen und welche Fragestellungen konnten durch die Analyse be- bzw. widerlegt werden? Des Weiteren könnte auch von Interesse sein, welche Entscheidungen auf Basis eines Analyseberichts getroffen wurden und wie diese sich in weiteren quantitativen Daten des DWH widerspiegeln.

Die Beantwortung solcher und ähnlicher Fragen spielt insbesondere abseits des Standard Reportings bei Untersuchung komplizierter Fragestellungen eine große Rolle. Die Dokumentationsmöglichkeiten bestehender Ad-hoc Reporting Werkzeuge, die allenfalls in Form unstrukturierter Zusatzinformationen erfolgen (Willenborg), stoßen hierbei schnell

an ihre Grenzen. Sie bieten weder Speicher- und Verwaltungs-, noch Such- und Navigationsmöglichkeiten für die beschriebenen Informationen. Daher wäre es wünschenswert, in einem analytischen Informationssystem über eine Metaebene zu verfügen, die es erlaubt, Informationen zu Fragestellungen, zu Analyseprozessen sowie zu Ergebnissen und deren Beziehungen in Form von strukturierten maschinenlesbaren Metadaten festzuhalten. In diesem Zusammenhang wird in diesem Forschungsansatz auch von semantischen Annotationen auf Basis von Semantic Web Technologien gesprochen, die zur Anreicherung von Prozessen und insbesondere quantitativen Daten verwendet werden. Die hierfür zugrunde liegende Wissensbasis wird durch verschiedene Ontologien umgesetzt.

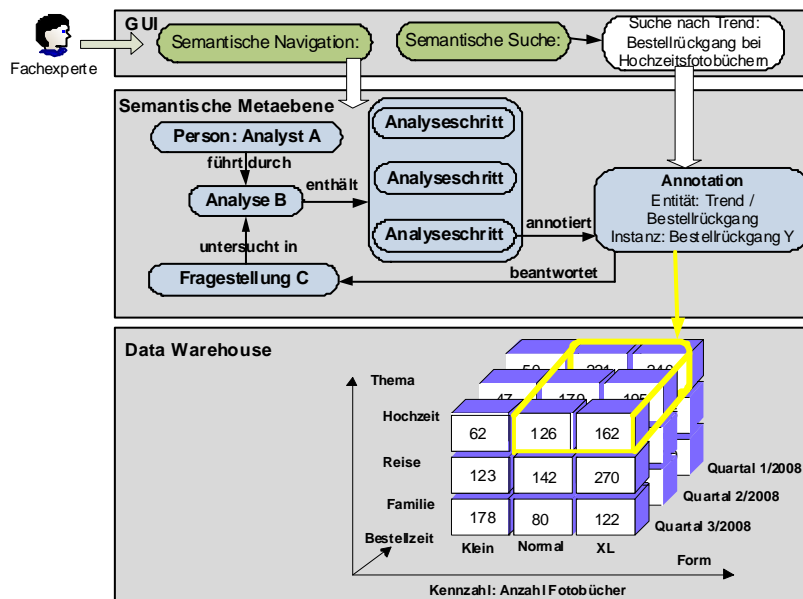


Abbildung 1: Beispiel semantischer Annotationen

Die Vorteile dieses Ansatzes bestehen darin, auf Basis strukturierter semantischer Metadaten auf den quantitativen Daten und Prozessen neue Funktionalitäten in einem analytischen Informationssystem anbieten zu können. Wie das Beispiel in Abb. 1 zeigt, können mit einer semantischen Suche annotierte quantitative Daten des Data-Warehouse mit einer spezifischen Semantik, wie z.B. Trend, Ereignis, Anomalie, Effekt, Muster, etc. gefunden werden. Aber auch die Suche nach komplexeren Annotationen, z.B. zu Analysen, die in einer Beziehung zu einer Fragestellung und zu einer Personen stehen, ist möglich. Des Weiteren ermöglicht dieser Ansatz eine Visualisierung von bereits bestehenden Ergebnissen und hilft so Analysten und Entscheidungsträgern relevante Daten samt ihrer Semantik in Berichtselementen zu erkennen. Darüber hinaus kann, eine semantische Navigation entlang der modellierten „benannten“ Beziehungen in den Metadaten erfolgen.

Ein Beispiel hierfür ist die Navigation von einem Analysten A zu einer Analyse B über den Pfad „führt durch“.

## 2 Verwandte Arbeiten

Schwerpunkt dieser Arbeit ist die Erfassung und Nutzung von Metadaten zu Analyseprozessen und deren Ergebnissen im Business Intelligence (BI) Bereich auf Basis von Semantic Web Technologien. Im Folgenden werden verschiedenen Forschungsbereiche und -arbeiten beschrieben, die sich zum Einen ebenfalls mit der Erfassung und Nutzung solcher Metadaten beschäftigen (Baars, 2006). Zum Anderen werden aber auch Forschungsbereiche aufgezeigt, die nicht unmittelbar die gleiche Forschungsfrage, wie der hier vorgestellte Ansatz verfolgen, jedoch die Potentiale der Verknüpfung des Semantic Webs mit dem BI-Bereich verdeutlichen. Dazu gehören die Bereiche: Verknüpfung von strukturierten und unstrukturierten Daten, Verknüpfung verschiedener BI-Systeme auf Basis einer semantischen Mittelschicht und die Überwindung bestehender Begrenzungen aktueller Data-Warehouse-Systeme.

Erste Synergiepotentiale zwischen dem Semantic Web und der Business Intelligence konnten bereits 2003 von Priebe mit der noch heute relevanten Forschungsrichtung, der Integration und Analyse von strukturierten und unstrukturierten Daten identifiziert werden (Priebe, 2003). Konkret ging es in den Arbeiten um die Verwendung von Ontologien für den Aufbau eines Enterprise Knowledge Portals, das sowohl OLAP-, als auch Information Retrieval- Funktionalitäten in quantitativen Data-Warehouse-Daten und zugehörigen qualitativen Dokument-Daten ermöglicht. Ein ähnlicher Themenschwerpunkt wurde auch in (Haak, 2007) verfolgt. Hier stand die Frage im Fokus, inwieweit sich strukturierte Data-Warehouse-Daten und unstrukturierte Content-Management-Daten auf Basis ihrer Semantik integrieren lassen. Anlog beschrieb Ludwig, dass Wissenrepräsentationssprachen wie RDF unstrukturierte als auch strukturierte Daten (semi-) strukturieren können und so eine Klammer zwischen Dokument und Tabelle, zwischen CMS und DWH, gebildet werden kann (Ludwig, 2005).

Soll ein Corporate Knowledge Center (Klesse, 2003) zur ganzheitlichen Entscheidungsunterstützung des Managements aufgebaut werden, so muss bei der Schaffung einer Integrationsinfrastruktur für die BI-Applikationen eine semantische Verknüpfung strukturierter Daten auf logischer Ebene erfolgen. Für die Umsetzung eines solchen Systems kann eine ontologiebasierte Architektur verwendet werden. In (Spahn, 2008) wird ein ähnlicher Ansatz verfolgt, wobei mit Hilfe einer „business level Ontology“ eine semantischen Mittelschicht gebildet wird, die Endnutzern die Möglichkeit gibt, in einem einheitlichen Vokabular verschiedene BI-Systeme (CRM, ERP, ...) konsistent anzufragen.

Einen besonderen Stellenwert nimmt das Semantic Web in der Überwindung bestehender Begrenzungen von aktuellen Data-Warehouse-Systemen ein, die laut Inmon meist auf die Vernachlässigung von Metadaten zurückzuführen sind (Inmon, 2008). So werden in verschiedenen Forschungsansätzen Business-Ontologien aufgebaut, welche „Business Semantics“ modellieren und diese für das BI-System zur Verfügung stellen (O’Neil, 2007). Sie ermöglichen es, den semantischen Kontext - Annahmen, Definitionen, Business Regeln, Terminologie, Hintergrundinformationen - eines Business Objekts zu berücksichtigen und für neue Funktionen zu nutzen. Xie verwendet diese Technik, um schnell und flexibel die Analysen auf sich dynamisch ändernde Anforderungen anpassen zu können (Xie, 2007). Mit den Cerebra Tools sollen so flexible und erweiterbare Datenrepräsentationen sowie komplexe und feingranulare Business Rules modelliert und berücksichtigt werden (Cerebra, 2004). In Diamantini wird neben einer Domänenontologie für die Business Semantics eine mathematische Ontologie einbezogen, die es erlaubt, die mathematischen Formeln auf den Kennzahlen samt der Semantik von Operatoren und Operanden zu beschreiben, um so neue OLAP Operationen anbieten zu können (Diamantini, 2008). Sell beschreibt mit ihrem Semantic Business Intelligence Framework ebenfalls eine vielversprechende Möglichkeit Business Intelligence mit weiterführenden Semantiken zu verknüpfen. Verbesserte Such- und Navigationsmöglichkeiten in den zu untersuchenden Daten, die Filterung multidimensionaler Daten anhand ihrer Semantik, die Berücksichtigung von Business Rules oder das Vorschlagen weiterführender Analysemöglichkeiten sind nur einige Funktionen einer verbesserten explorativen Analyse. Das automatische Query Rewriting zur semantischen Anreicherung der DWH-Anfrage-Ergebnisse bildet einen Schwerpunkt in diesem Arbeiten (Sell, 2005/2008).

Im Bereich der Dokumentation von Analyseprozessen und -ergebnissen ist die Arbeit von Baars zur Distribution von Business-Intelligence-Wissen zu nennen. Die Arbeit zielt auf die kontrollierte und einfache organisationsweite Verbreitung und Weiterverwendung von Berichten und Analyseansätzen (Templates) durch einrichten der BI-Inhalte in Wissensmanagementsysteme, jedoch ohne auf eine konkrete technische Umsetzung einzugehen (Baars, 2006).

Im eigenen Ansatz wird ebenfalls die Erfassung von maschinenlesbaren Metadaten zur Durchführung des Analyseprozesses sowie zur Ergebnissicherung angestrebt. Allerdings wird der Fokus auf die Semantik der Entitäten sowie deren Beziehungen zueinander gelegt, um darauf aufbauend semantische Such- und Navigationsfunktionen im analytischen Informationssystem zu realisieren, wie sie in den Arbeiten von (Sell, 2005/2008) auf den zu untersuchenden quantitativen Daten möglich sind. Hierbei verbleiben die semantischen Metadaten jedoch im Gegensatz zu (Baars, 2006) im analytischen Informationssystem und werden nicht in Form von Templates für neue ähnliche Analysen an ein Wis-

sensmanagementsystem weitergegeben. Die Verknüpfung von strukturierten und unstrukturierten Daten sowie die Verknüpfung zu anderen BI-Systemen werden in diesem Ansatz nicht angestrebt. Auch spielt die Modellierung und Berücksichtigung der Business Semantics der zu untersuchenden Objekte in diesem Ansatz zunächst keine Rolle.

### 3 Konzept

Ziel dieses Forschungsansatzes ist die Erfassung und Verwaltung maschinenverständlicher Metadaten zu Analyseprozessen und Ergebnissen und deren Nutzung für semantische Such- und Navigationsmöglichkeiten in einem analytischen Informationssystem. Technologisch bilden Ontologien auf Basis der standardisierten Web-Ontology-Language (OWL) die benötigte Metaebene für die strukturierte Dokumentation. Sie definieren sowohl die Entitäten und ihre Semantik als auch ihre hierarchischen und semantischen Beziehungen zueinander.

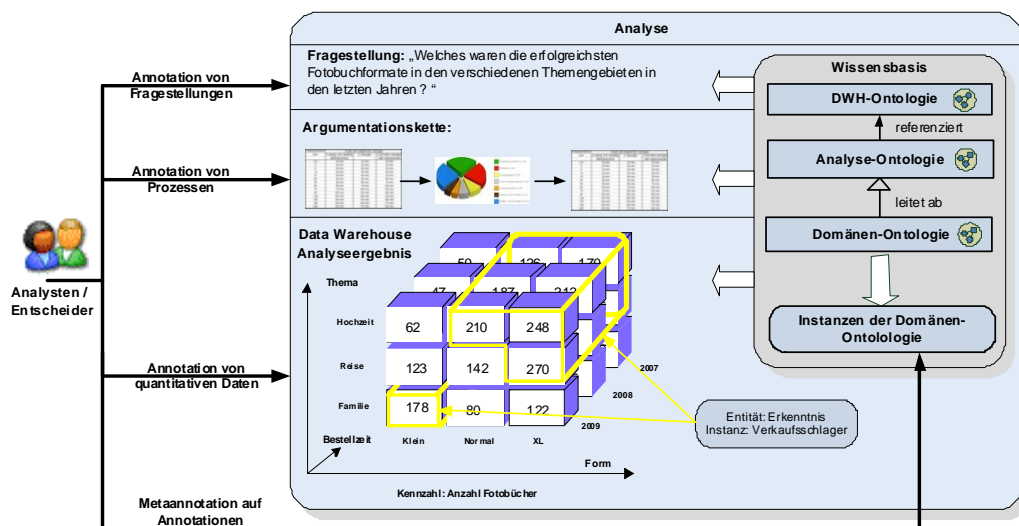


Abbildung 2: Semantische Annotation von Daten und Prozessen

In Abbildung 2 werden die Ontologien unter dem Begriff Wissensbasis zusammengefasst, wobei sie als Grundlage für die semantischen Annotationen dienen. Diese repräsentieren konkrete Instanzen der verschiedenen Entitäten und stellen diese miteinander in Beziehung. So kann z.B. in Abb. 2 eine spezifische Fragestellung erfasst und mit einem konkreten Analyseprozess in eine „untersucht“ Beziehung verknüpft werden. Auch ist es möglich, verschiedene Analyseelemente als Bestandteil eines Analyseprozesses zu kennzeichnen und diese in einer Vorgänger-Nachfolger Beziehung zeitlich anzuordnen. Darüber hinaus können Analyseergebnisse mit einer spezifischen Semantik wie „Verkaufschlager“ auf quantitativen Daten annotiert werden, was in Abb. 2 durch die gelbe Um-

randung auf dem Datenwürfel visualisiert wird. Metadatenannotationen ermöglichen es Aussagen über Annotationen zu treffen und so z.B. eine Ursache / Wirkungsbeziehung zwischen zwei Annotationen – z.B. Marketingaktion und Verkaufsschlager – zu beschreiben. Im Folgenden werden der Aufbau der jeweiligen Ontologien, deren Beziehungen zueinander sowie ihre Entstehung näher erläutert.

### **3.1 DWH-Ontologie**

Um eine Schnittstelle zwischen den Realweltentitäten des Data-Warehouses und den Analyseprozess- und Ergebnisentitäten herstellen zu können, wird eine Data-Warehouse-Ontologie benötigt, die ein Mapping auf die Entitäten des DWH-Schemas, wie Kennzahlen, Dimensionen, Dimensionshierarchien, etc. beschreibt. Hartmann stellt in diesem Zusammenhang eine OWL-Ontologie zur semantisch reichhaltigen Beschreibung von DWH-Metadaten zur Verfügung, die auf dem Semantic-Data-Warehouse-Modell (SDWM) basiert (Hartmann, 2008; Böhnlein, 2001). Ergänzt um weitere Konzepte aus dem am OFFIS entwickelten Metamodell MADEIRA (Wietek, 2000) und reduziert um nicht benötigte Entitäten ergibt sich das in Abb. 3 visualisierte Schema.

Die dargestellten Entitäten und Beziehungen ermöglichen es Metadaten zu den Datenstrukturen des Data-Warehouses sowie deren Analysemöglichkeiten festzuhalten und diese mit weiteren Metadaten des Analyseprozesses in Verbindung zu setzen. Die in Abb.3. visualisierten Entitäten und Beziehungen reichen für das hier verfolgte Ziel aus, jedoch wäre es durch die Erweiterung der DWH-Ontologie theoretisch möglich eine ähnliche Business Ontologie, wie in (Sell ,2005/2008) oder (Xie, 2007) aufzubauen. Hartmann sieht bereits weitere Strukturen zu den Bereichen Terminologie, Organisations- und Systembezug, Datenqualität, Datentransformationen und Historie vor. Daneben wäre es aber auch möglich Konzepte zu modellieren, die spezifische Aussagen über die zu analysierenden quantitativen Daten enthalten. Dazu zählen insbesondere semantische Beziehungen, die in der Realwelt existieren, von bestehenden hierarchischen DWH-Strukturen aber nicht erfasst werden können. So können spezifische Business Semantics (O’Neil, 2007) der im DWH untersuchten Objekte explizit festgehalten werden.





Die Analyse-Ontologie gliedert sich in zwei Bereiche, wobei der erste die allgemeine Definition von Entitäten und Relationen aus dem Umfeld analytischer Informationssysteme fokussiert. Er erlaubt es, Konzepte wie Personen, Analysen, Fragestellungen, Berichtselemente sowie deren semantische Beziehungen zu beschreiben (s. Abb. 4). Darauf basierend können semantische Annotationen zur Dokumentation von Fragestellungen und Analyseprozessen erfolgen. Der zweite Bereich dient der Ergebnissicherung. Er definiert zum Einen Entitäten, die von „direkte Annotation“ ableiten und es dem Analysten erlauben, erkannte Ergebnisse in den quantitativen Daten festzuhalten. Zum Anderen definiert der zweite Bereich aber auch Metadatenannotationen, um Aussagen über die Annotationen zu treffen. Sowohl direkte Annotationen als auch Metadatenannotationen verfügen über eine Vererbungshierarchie, um spezifischere Annotationstypen, wie Effekte, Anomalien, Trends, etc. beziehungsweise Zusammenhänge, Sachverhalte, etc. zu beschreiben. Zu beachten ist, dass diese Entitäten zunächst domänenunabhängig modelliert sind, um die Ontologie in verschiedenen Analyseszenarien einsetzen zu können. Da jedoch in verschiedenen Domänen unterschiedliche Ergebnisse annotiert werden können, werden Domänenontologien definiert, welche die Entitäten der Analyse-Ontologie um domänenspezifische Konzepte erweitern. So ist es möglich in der Domäne epidemiologisches Krebsregister eine Entität „Erhöhung der altersspezifischen Inzidenz“ zu modellieren die

ebenso wie die Entität „Bestelleinbruch“ in der Domäne Fotobuch von der Entität „Trend“ ableitet.

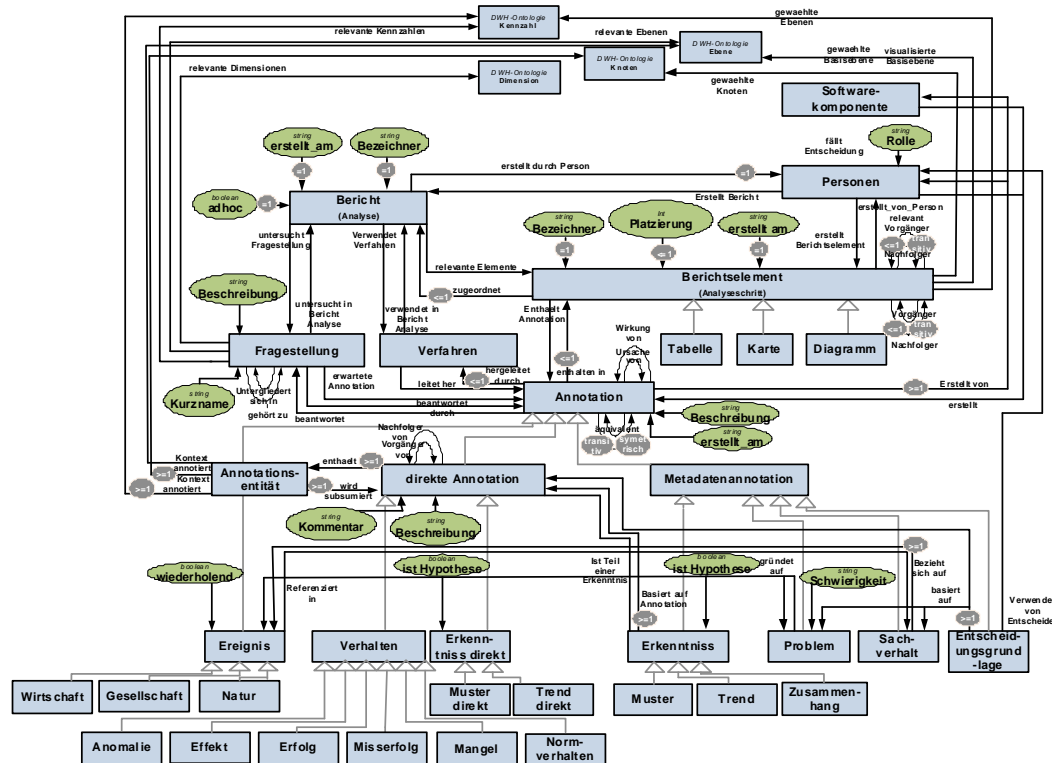


Abbildung 4: Entitäten und Beziehungen der Analyse-Ontologie

Eine Verknüpfung der Analyse-Ontologie und der DWH-Ontologie wird über die semantischen Beziehungen mehrerer Entitäten zueinander hergestellt. Berichtselemente speichern beispielsweise die Informationen, welche Kennzahlen in welchen Dimensionsebenen und Knoten dargestellt werden und welche Dimensionsebenen und Knoten im Kontext gewählt sein müssen. Auch spezifizieren Fragestellungen, die Kennzahlen, Dimensionen, Ebenen und Knoten, die in der DWH-Ontologie instanziiert und im Rahmen einer Untersuchung von Interesse sind. Um semantische Annotationen auf den quantitativen Daten auszuführen, die z.B. einen spezifischen Trend beschreiben, müssen ebenfalls Beziehungen zu der DWH-Ontologie bestehen. Hierbei ist zu beachten, dass eine direkte Annotation sich nicht auf die Faktdaten selbst, sondern auf die DWH-Metadaten (Dimensionen, Ebenen, Knoten), welche die quantitativen Kennzahlausprägungen aufspannen, beziehen. Über das Konzept der Annotationsentitäten ist es möglich, einzelne Kennzahlenbereiche (s. Abb.2) zu wählen. Hierbei werden mehrere Annotationsentitäten in einer direkten Annotation subsumiert.

Da die hier gewählte OWL-DL-Ontologie auf Beschreibungslogik basiert, ist es möglich, Restriktionen auf den Beziehungen der Entitäten zu definieren und diese darüber hinaus als transitiv, symmetrisch und funktional zu markieren. So kann z.B. spezifiziert werden, dass eine Annotation von mindestens einer Person oder Softwarekomponente erstellt werden muss. Dies erlaubt eine Konsistenzprüfung der modellierten Ontologieinstanzen.

### 3.3 Semantische Anwendungen

Auf Basis der vorgestellten Ontologien lässt sich eine semantische Metaebene in das analytische Informationssystem einbinden, die es ermöglicht, Analyseprozesse und Ergebnisse strukturiert und maschinenverständlich zu dokumentieren. Dies ermöglicht nicht nur die Visualisierung bestehender Ergebnisse samt ihrer Semantik während der Exploration, sondern erlaubt auch mit Hilfe einer semantischen Suche komplexe Anfragen an die Wissensbasis zu senden. Hierbei kann nach allen Entitäten der Analyse-Ontologie insbesondere unter Berücksichtigung der ein- und ausgehenden semantischen Beziehungen gesucht werden. Technisch wird eine Umsetzung auf Basis der „SPARQL Query Language for RDF“ erfolgen. Die semantischen Beziehungen zwischen den aufgezeichneten Metadaten erlauben eine explorative Navigation entlang benannter Pfade.

## 4 Implementierung in der MUSTANG Plattform

OFFIS beschäftigt sich seit vielen Jahren mit epidemiologischen Fragestellungen. Ursprung dieser Aktivitäten war die Bereitstellung eines analytischen Informationssystems für das Epidemiologische *Krebsregister* Niedersachsen (EKN). Über die Zeit hat sich dieses Werkzeug als sehr mächtig erwiesen, so dass es neben dem Public Health Bereich vielfältig auch in anderen Projekten beispielsweise als Management Unterstützungssystem zum Einsatz gebracht worden ist. Da diese Analyse-Infrastruktur ursprünglich als Expertensystem für die Epidemiologie konzipiert worden war, stellte sich bei Einsätzen in anderen Domänen die intuitive Bedienbarkeit nicht immer ein. Auch erwies sich die ursprüngliche Systemarchitektur als unflexibel, so dass Anpassungen an neue Bereiche und Anforderungen sich als schwierig erwiesen. Aus diesem Grund wurde mit der Neuentwicklung der multidimensionalen, geografischen statistischen Datenanalyse (MUSTANG) Plattform begonnen. Hierbei spielte der Aspekt der Serviceorientierung eine wesentliche Rolle. Im Folgenden werden kurz die drei Kerngebiete von MUSTANG aufgelistet.

- **Multidimensional:** Daten, die mittels der MUSTANG Plattform zur Analysen verwendet werden sind multidimensional aufbereitet.

- **Erweiterte Statistik:** Für die Analysen können vielfältige erweiterte statistische Verfahren angewendet werden. Diese Eigenschaft hat ihren Ursprung in der Epidemiologie, da hier komplexe Verfahren realisiert werden müssen.
- **Geographisch:** Daten können einen Geographiebezug besitzen, der dann zur räumlichen statistischen Analyse herangezogen wird. Hierbei unterstützen wir sowohl kleinräumige wie Flächenanalysen.

#### 4.1 Beschreibung der MUSTANG Plattform

Die MUSTANG Plattform basiert auf einer rekonfigurierbaren serviceorientierten Architektur, wobei die Services zustandslos sind. Daten werden über sogenannte Datentransfer Objekte ausgetauscht. Diese Objekte beschreiben zusammen den Zustand des Systems. Die Services lassen sich in einem gewissen Rahmen zu Anwendungen komponieren. Services werden abstrakt als Interfaces konzipiert und werden dann durch ggf. verschiedene Realisierungen konkretisiert. Sowohl bei den Services wie auch bei den Datentransfer Objekten muss eine größtmögliche Stabilität gewährleistet werden, da Änderungen an diesen Definitionen große Umstrukturierungen mit sich ziehen.

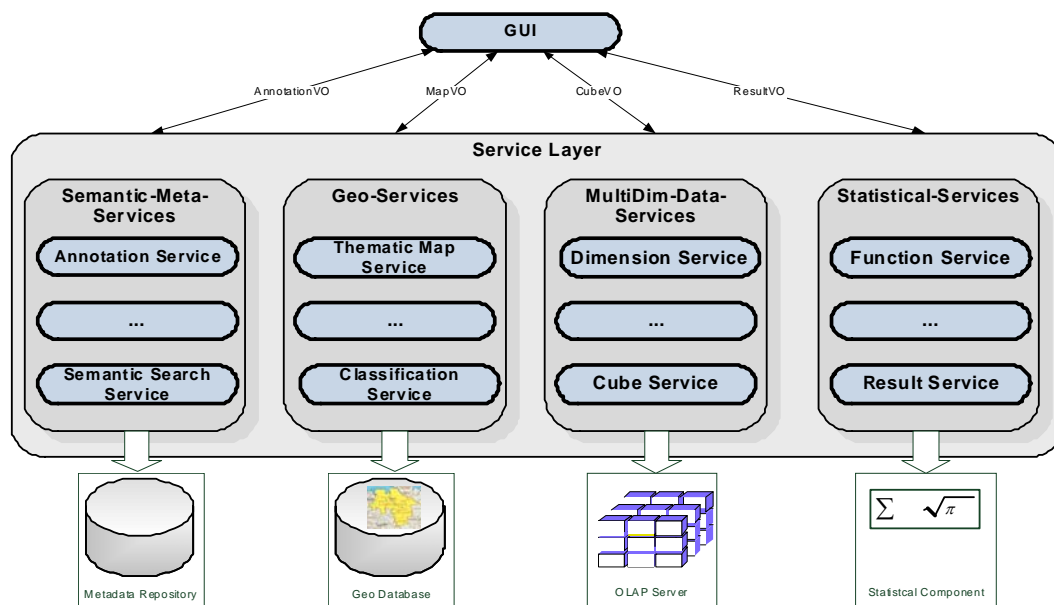


Abbildung 5: Erweiterte MUSTANG Architektur

In Abbildung 5 ist die Architektur von MUSTANG mit ihren verschiedenen Schichten zu sehen. Auf der untersten Schicht befinden sich die Standardkomponenten, wie Geo-Datenbank, OLAP Server und Statistik-Komponente. Diese werden in einer darüber liegenden Service-Schicht in eine einheitliche Sicht integriert, so dass diese Fremdkompo-

nenten transparent zugreifbar sind. Die oberste Schicht bildet schließlich die graphische Benutzungsschnittstelle (GUI).

Die Service-Schicht unterteilt sich in vier voneinander abgegrenzte Anwendungsblöcke mit unterschiedlichem Funktionsumfang. Die Realisierung dieser Blöcke erfolgt durch konkrete Services.

- **Geo-Services:** Hier werden Funktionen zum Umgang mit geografischen Daten umgesetzt, aktuell der Classification- Service zur Anfrage von Geo-Objekten aus einer Geodatenbank, sowie der Thematic-Map-Service zum Erzeugen thematischer Karten. Zurzeit wird PostGIS als Standard-Geodatenbank integriert.
- **Multidimensional-Data-Services:** Diese Services kapseln den Zugriff auf den OLAP-Server. Der Dimension-Service ist für die Abfragen von Dimensionen und deren Elementen zuständig, der Cube-Service für das Abfragen von Cubes innerhalb des OLAP-Servers. Der eigentliche Zugriff auf den OLAP-Server erfolgt mittels der Abfragesprache XMLA. Dies ermöglicht es, auf einfache Weise andere OLAP-Server an die Plattform anzubinden. Unterstützte Standard-OLAP-Server sind zurzeit Microsoft Analysis Services und Mondrian.
- **Statistical-Services:** Hier werden die von MUSTANG unterstützten statistischen Verfahren, für komplexe Berechnungsmethoden auf OLAPCubes, realisiert. Hierbei ist, auch über die Grenzen einzelner OLAPCubes hinweg, eine große Vielzahl unterschiedlicher statistischer Verfahren und Berechnungsmöglichkeiten realisierbar. Der Function-Service stellt dabei die im System integrierten statistischen Verfahren bereit. Der Result-Service dient zum Abruf der um statistische Daten angereicherten OLAP-Cubes bzw. Kennzahlen. Die Berechnungen werden nicht innerhalb der Plattform durchgeführt, sondern durch Einbindung der Standard-Statistik-Komponente RProject.
- **Semantic-Meta-Services:** Realisiert die in diesem Artikel vorgestellte Verfahren zur strukturierten Erfassung semantischer Metadaten im Analyseprozess. Die Einbettung in die MUSTANG Plattform wird im folgenden Abschnitt erläutert.

Der Service-Layer ist flexibel gehalten, so dass auch weitere Standardkomponenten, wie beispielsweise weitere Geo- Datenbanken oder andere OLAP-Server, integriert werden können.

## 4.2 Einbettung des Forschungsansatzes in die MUSTANG Plattform

Mit dem hier vorgestellten Forschungsansatz wird ein neuer Anwendungsblock in den Servicelayer eingeführt, der mit neuen Funktionalitäten bestehende Probleme analytischer Informationssysteme im Allgemeinen und MUSTANG im Speziellen im Bereich der Dokumentation relevanter Zusatzinformationen zu Fragestellungen, Analyseprozessen und Ergebnissen sowie deren Semantiken adressiert. Durch die Erfassung solcher Informationen in Form von strukturierten Metadaten, die als Instanzen eines DWH-, BI- und Domänenmodells in einen Kontext eingebettet werden, ergeben sich neue Visualisierungs-, Such- und Navigationsmöglichkeiten, die insbesondere eine bessere Nachvollziehbarkeit von Analysewegen zur Ergebnisfindung, den Ergebnissen selbst und darauf basierenden Entscheidungen ermöglichen.

Die Semantic-Meta-Services:

- **OntologyManagementService:** Dieser Service dient der Verwaltung der in Kapitel 3. vorgestellten Ontologien, wobei er das Erzeugen und Ändern sowie Laden und Speichern von Ontologiemodellen, sowie konkreten Entitäten und Instanzen kapselt. So kann im Hintergrund ein beliebiges Ontologie-Verwaltungs-System eingesetzt werden, ohne das weitere Services angepasst werden müssen. Eine Umsetzung erfolgt durch die Jena2 Ontology API.
- **AnalyseManagementService:** Dieser Service dient unter anderem dem Erfassen strukturierter Metadaten zum Einloggen von spezifischen Nutzern, zum Erstellen von Analysen und dem Erstellen von Fragestellungen.
- **DWHOntologyService:** Dieser Service verwaltet die Data-Warehouse-Ontologie. Durch ihn werden alle relevanten Instanzen des DWH-Schema in die Ontologie geladen, verwaltet und wieder auf das DWH-Schema abgebildet.
- **AnnotationService:** Dieser Service erlaubt das Erstellen von direkten Annotationen auf den quantitativen Daten und das Erstellen von Metaannotationen auf diesen.
- **LoggingService:** Mit diesem Service werden die strukturierten Daten rund um die Analysedurchführung strukturiert festgehalten. Dies umfasst unter anderem die gewählten Parameter, die durchgeführten OLAP Operationen sowie die gewählten Visualisierungen der Daten.

- **VisualisationService:** Dieser Service realisiert die Visualisierung bestehender Ergebnisse in Form von Annotationen auf quantitativen Daten während der Exploration im analytischen Informationssystem.
- **SearchService:** Dieser Service realisiert die Suche nach Metadaten mit einer spezifischen Semantik. Intern werden Fragen in der SPARQL Query Language for RDF formuliert und an das Ontologie-Verwaltungs-System geschickt. Die zurückgesendeten Antworten werden verarbeitet und für den Systembenutzer aufbereitet.
- **NavigationService:** Dieser Service erlaubt die Durchführung von Navigationen entlang semantischer Pfade, wie sie in den zugrunde liegenden Analyse- und Domänen-Ontologien modelliert sind.

Alle Services können intern und mit der GUI über Datentransferobjekte kommunizieren. Die hier aufgelisteten Services sind dabei nicht isoliert zu betrachten sondern stehen ebenfalls mit den MultiDim- Data-Services in Verbindung. Analog zu den Services werden entsprechende GUI Elemente realisiert, welche die im Forschungsansatz beschriebenen Funktionalitäten umsetzen.

## 5 Fazit und Ausblick

Ziel des vorgestellten Forschungsansatzes ist es, strukturierte Dokumentationen von Analysefragestellungen, Analyseprozessen und erkannten Erkenntnissen und Zusammenhängen auf Basis semantischer Metadaten in analytischen Informationssystemen umsetzen zu können. Insbesondere abseits des Standardreportings kann mit diesem Ansatz die Nachvollziehbarkeit komplexer Analysen, deren Ergebnissen und darauf basierender Entscheidungen verbessert werden, indem semantische Such- und Navigationsfunktionalitäten in dem analytischen Informationssystem angeboten werden. Es wurde beschrieben, dass mit Hilfe von Semantic Web Technologien eine maschinenlesbare und -verständliche Metaebene aufgebaut werden kann, die es erlaubt, die semantischen Annotation zu erfassen, zu speichern und zu verwalten und schließlich in Form neuer Funktionalitäten dem Analysen oder auch der Fach- oder Führungskraft wieder zur Verfügung zu stellen. In diesem Zusammenhang wurden die Data-Warehouse- und die Analyse-Ontologie vorgestellt, die den Kern der Metadatenverwaltung bilden. Erstere bildet Entitäten und deren Zusammenhänge des DWH ab, um diese in anderen Ontologien zugreifen zu können und eignet sich darüber hinaus zur Modellierung sogenannter Business Semantics. Die Analyse-Ontologie hingegen modelliert Entitäten und deren hierarchische und semantische Beziehungen, um Fragestellungen und Prozesse strukturiert festhalten zu können. Des Weiteren definiert sie domänenunabhängige Entitäten, um erkannte Analyseergebnisse in quantita-

tiven DWH-Daten dokumentieren zu können. Spezifischere Domänenontologien leiten sich von der Analyse-Ontologie ab, um konkrete Annotationsmöglichkeiten für die jeweilige zu untersuchende Domäne zu modellieren.

Eine Umsetzung der semantischen Metaebene zur Speicherung und Verwaltung erfasster Metadaten sowie darauf aufbauende Such-, Navigations- und Visualisierungsmöglichkeiten werden in der am OFFIS entwickelten MUSTANG-Plattform prototypisch umgesetzt. Eine Implementierung von entsprechenden Services und GUI-Elementen wird derzeit realisiert.

Weiterführende Arbeiten werden sich mit der intelligenten Nutzung erhobener Analyse- und Ergebnismetadaten, insbesondere in Kombination mit modellierten Business Semantics der zu untersuchenden quantitativen Daten, beschäftigen. Im Fokus steht zum Einen die Bildung von Klassifikatoren auf Basis bestehender Analyseergebnisse zur semiautomatischen Exploration und Annotation von neuen Analyseergebnissen während des Analyseprozesses. Zum Anderen wird eine zielgerichtete und weiterführende Analyseunterstützung für Ad-Hoc Analysen im analytischen Informationssystem angestrebt, von der insbesondere Fach- und Führungskräfte, aber auch Analysten profitieren sollen.

## 6 Literatur

Baars, H. (2006). Distribution von Business-Intelligence-Wissen: Diskussion eines Ansatzes zur Nutzung von Wissensmanagement-Systemen für die Verbreitung von Analyseergebnissen und Analysetemplates. In P. Chamoni, & P. Gluchowski, *Analytische Informationssysteme* (pp.409-424). Berlin, Heidelberg: Springer.

Böhnlein, M. & vom Ende, A.U. (2001). Semantisches Data Warehouse-Modell (SDWM): Ein konzeptuelles Modell für die Erstellung multidimensionaler Datenstrukturen. Retrieved February 8, 2009 from [http://www.ceushb.de/forschung/downloads/SDWM\\_Rundbrief.pdf](http://www.ceushb.de/forschung/downloads/SDWM_Rundbrief.pdf)

Diamantini, C. & Potena, D. (2008). Semantic Enrichment of Strategic Datacubes. In I.Song & A. Abelló (Eds.), *DOLAP 2008, ACM 11th International Workshop on Data Warehousing and OLAP*, ACM

Gluchowski, P. (2006). Techniken und Werkzeuge zum Aufbau betrieblicher Berichtssysteme In P.Chamoni, & P.Gluchowski, *Analytische Informationssysteme* (pp.207-226). Berlin, Heidelberg: Springer.

Haak, L. (2007). *Semantische Integration von Data Warehousing und Wissensmanagement*. Berlin: dissertation.de – Verlag im Internet GmbH.

Hartmann, S. (2008). *Überwindung semantischer Heterogenität bei multiplen Data-Warehouse-Systemen*, Bamberg: University of Bamberg Press.

Inmon, W.H., Strauss, D. & Neushloss, G. (2008). *DW 2.0: The Architecture for the Next Generation of Data Warehousing* (Morgan Kaufman Series in Data Management Systems), Burlington: Elsevier Science.



Klesse, M., Melchert, F. & von Maur, E. (2003). Corporate Knowledge Center als Grundlage integrierter Entscheidungsunterstützung In U. Reimer, A. Abecker, S. Staab & G. Stumme (Eds.), WM 2003: Professionelles Wissensmanagement - Erfahrungen und Visionen. GI.

Ludwig, L. (2005). Business Intelligence und das Semantic Web: ein Traumpaar. Retrieved May 15, 2009 from <http://www.competence-site.de/business-intelligence-hr/Business-Intelligence-und-das-Semantic-Web-39683>

Network Inference Inc. (2004). Ontology And Data Warehousing: How Data Warehousing Investments can deliver continous ROI by augmenting them with ontologies and inference capabilities Retrieved March 22, 2009 from [http://me.jtpollock.us/pubs/2005.10-Whitepaper\\_DataWarehouse.pdf](http://me.jtpollock.us/pubs/2005.10-Whitepaper_DataWarehouse.pdf)

O'Neil, B. (2007). Semantics and Business In The Data Administration Newsletter. Retrieved September 9, 2009 from <http://www.tdan.com/view-articles/4934>

Priebe, T. & Pernul, G. (2003). Ontology-based Integration of OLAP and Information Retrieval. In V. Marik, W. Retschitzegger & O. Stepankova (Eds.), 14th International Workshop on Database and Expert Systems Applications (pp. 610-614). Heidelberg: Springer.

Sell, D., Caroso da Silva, D., Beppler, F.B., Napoli, M., Ghisi, F.B., Pacheco, R.C.S. & Todesco, J.L. (2008). SBI: a semantic framework to support business intelligence. In A. Duke, M. Hepp, K. Bontcheva & M.B. Vilain (Eds.), Proceedings of the First International Workshop on Ontology-supported Business Intelligence. ACM.

Sell, D., Cabral, L., Motta, E., Domingue, J., Hakimpour, F. & Pacheco, R. (2005). A Semantic Web based Architecture for Analytical Tools In 7th IEEE International Conference on E-Commerce Technology (CEC 2005)

Spahn, M., Kleb, J., Grimm, S. & Scheidl, S. (2008). Supporting Business Intelligence by Providing Ontology-Based End-User Information Self-Service intelligence. In A. Duke, M. Hepp, K. Bontcheva & M.B. Vilain (Eds.), Proceedings of the First International Workshop on Ontology-supported Business Intelligence. ACM.

Wietek, F. (2000). Intelligente Analyse multidimensionaler Daten in einer visuellen Programmierungsumgebung und deren Anwendung in der Krebsepidemiologie. Universität Oldenburg.

Willenborg, K., Berichtswesen - Vereinigung von Gegensätzen. Retrieved July 10, 2009 from <http://www.sapdesignguild.org/editions/edition2/willenborg/index.html>

Winter, R., Schmaltz, M., Dinter, B. & Bucher, T. (2008). Das St. Galler Konzept der Informationslogistik. In B. Dinter & R. Winter, Integrierte Informationslogistik. (pp.1-16). Berlin, Heidelberg: Springer.

Xie, G., Yang, Y., Liu, S., Qiu, Z., Pan, Y. & Zhou, X. (2007). EIAW: Towards a Business-Friendly Data Warehouse Using Semantic Web Technologies In K. Aberer, K. Choi, N. Fridman Noy, D. Allemang, K. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, P. Cudré-Mauroux (Eds.), The Semantic Web, 6th International Semantic Web Conference. Berlin, Heidelberg: Springer.



# **Business Intelligence bei Entscheidungen im Human Resource Management**

Andreas Hilbert, Alexander E. Müller

TU Dresden – Fakultät Wirtschaftswissenschaften, Professur für Wirtschaftsinformatik,,  
insb. Informationssysteme im Dienstleistungsbereich

## **1 Einleitung**

Die dynamische Situation, in der sich Unternehmen heutzutage befinden, fordert effiziente Entscheidungsprozesse. Dazu werden präzise und solide Analysen der Geschäftsdaten benötigt. Die fortlaufende Entwicklung und das Bestreben, immer mehr Daten zu erfassen, schafft riesige Datenmengen (Köster 2002). Dies zeigt sich ebenfalls im Personalwesen bzw. im Human Resource Management (HRM), wo eine Vielzahl an Daten vorhanden ist (Jung 2008, S. 655). Aus dieser Datenflut resultiert im HRM – wie bereits in anderen Bereichen des Unternehmens – die Notwendigkeit, relevante Informationen zu generieren und diese für Entscheidungen nutzbar zu machen (Lory 2005, S. 44-45). Hierzu kann Business Intelligence einen wertvollen Beitrag leisten. Es stellt sich folglich die Frage, welche Techniken der BI hierfür besonders geeignet sind und ob es Bereiche gibt, die besser oder stärker als andere unterstützt werden können.

Aus dieser Fragestellung heraus lässt sich das Forschungsziel der vorliegenden Arbeit als eine Systematisierung von Unterstützungspotentialen definieren, welche die Business Intelligence bei Entscheidungen im Human Resource Management bieten kann. Der in Abbildung 1 dargestellten Systematisierung der Aufträge und Ziele der Wirtschaftsinformatik gemäß Becker et al. 2004, S. 347 folgend, lässt sich die Aussage treffen, dass die gestaltungszielorientierte Erstellung des angestrebten Modells zur Zielerreichung auf der erkenntniszielgeleiteten Analyse der Entscheidungen des Personalbereichs und der Methoden der Business Intelligence aufbaut.

Der methodische Auftrag (Becker et al. 2004, S. 347) steht bei der Erreichung des erklärten Forschungsziels im Vordergrund. Die zur Erreichung des Forschungsziels eingesetzten Forschungsmethoden sind zum einen die in der Wirtschaftsinformatik vorherrschende argumentativ-deduktive Analyse (Wilde und Hess 2007, S. 284). Zusätzlich wird diese

stellenweise mit der als subjective-argumentative bezeichneten Forschungsmethode der Informationssystemforschung (Hars 2002, S. 42) kombiniert.

	Erkenntnisziel	Gestaltungsziel
Methodischer Auftrag	Verständnis von Methoden und Techniken der Informationssystemgestaltung	Entwicklung von Methoden und Techniken der Informationssystemgestaltung
Inhaltlich-funktionaler Auftrag	Verständnis von betrieblichen Informationssystemen und ihrer Anwendungsbereiche	Bereitstellung von IS-Referenzmodellen für einzelne Betriebe oder Branchen

*Abbildung 1: Aufträge und Ziele der Wirtschaftsinformatik  
(nach Becker et al. 2004, S. 347)*

In Kapitel 2 findet eine Betrachtung von Entscheidungen im Hinblick auf ihre Zugehörigkeit zum operativen oder zum strategischen Bereich eines Unternehmens statt. Die Analyse des personalwirtschaftlichen Entscheidungsprozesses mit seinen verschiedenen Phasen und die Auswahl derer, die für das weitere Vorgehen dieses Artikels relevant sind, beschreibt das dritte Kapitel. Gefolgt wird Kapitel 3 von der Erörterung des für das weitere Vorgehen Anwendung findende Verständnis der Business Intelligence (Kapitel 4). Im Kapitel 5 werden schließlich ausgewählte Business Intelligence Komponenten auf ihr Unterstützungspotential hin untersucht. Abschließend werden die bis dato erstellten Teilmodelle in ein Gesamtmodell integriert, welches einen ersten Systematisierungsansatz darstellt. Das Modell wird als eine State-of-the-Art-Lösung betrachtet, dass die Grundlage für weiterführende Untersuchungen bilden. Zusätzlich bietet es eine Hilfestellung bei der Entscheidung, welche Bereiche des HRM eines Unternehmens intensiver auf eine Unterstützung durch BI-Techniken hin untersucht werden sollten.

## 2 Merkmale strategischer und operativer Entscheidungen

Strategische und operative Entscheidungen werden häufig dahingehend unterschieden, dass strategische Entscheidungen mehr auf das Erreichen von Effektivität und operative Entscheidungen auf das Erreichen von Effizienz ausgerichtet sind. Eine Klassifikation von Entscheidungen bietet sich folglich an, um das Unterstützungspotential bei Entscheidungen spezifischer herauszuarbeiten. Dies lässt sich idealtypisch anhand von Merkmalen durchführen, die eine Entscheidung entweder mehr der operativen Ebene oder mehr der strategischen Ebene zuordnen. Eine solche Einordnung von Entscheidungen im Allgemeinen lässt sich deduktiv auf personalwirtschaftliche Entscheidungen übertragen und

ermöglicht im Rahmen dieser Arbeit eine Zuordnung von Entscheidungen zur strategischen bzw. operativen Ebene des HRM.

Entscheidungen können sich einerseits auf Ziele und andererseits auf Funktionen oder Prozesse im Sinne von Mitteln zur Erreichung von Zielen beziehen (Jaeschke 2007, S. 9). Routineentscheidungen, die durch regelmäßiges Auftreten sowie einen hohen Bekanntheits- und Gewohnheitsgrad charakterisiert sind und eher in den untergeordneten Hierarchieebenen anfallen, unterscheiden sich von einmaligen Entscheidungen, die eine höhere Unsicherheit aufweisen und aufgrund ihrer meist größeren Bedeutung vor allem in der Spitze der Unternehmensführung getroffen werden (Müller 1992, S. 60). Führungsentscheidungen unterscheiden sich von Ausführungsentscheidungen durch den eingeräumten Freiheitsgrad. Bei Führungsentscheidungen ist dieser sehr groß, was gleichzeitig bedeutet, dass Entscheidungen auf den unteren Unternehmensebenen stärker durch Regelungen eingeengt sind und dadurch eine begrenzte Folgewirkung haben. Mehrheitsentscheidungen finden, im Gegensatz zu Einzelentscheidungen, eher auf den oberen Unternehmensebenen statt (Korndörfer 1999, S. 65). Aufgrund der hohen Bedeutsamkeit bestimmter Entscheidungen sind diese meist nicht delegierbar und werden anders als delegierbare Entscheidungen auf der obersten Unternehmensebene getroffen (Gemünden 1983, S. 52). Der Entscheidungsweg verläuft in der Regel von der obersten Unternehmensebene hierarchisch nach unten ab (top-down). In seltenen Fällen, wenn z. B. Lösungsvorschläge aus den unteren Ebenen einbezogen werden, kann der Entscheidungsweg auch beginnend in einer unteren Unternehmensebene (bottom-up) nach oben erfolgen (Korndörfer 1999, S. 66). Sind Auswirkungen von Entscheidungen als langfristig zu betrachten, so fallen diese Entscheidungen in den strategischen Bereich. Entscheidungen mit kurzfristigen Auswirkungen sind im Gegensatz eher operativ. Ebenso sind Entscheidungen mit geringen Vorbereitungszeiten oder solche, die sogar in Echtzeit gefällt werden müssen, im operativen Personalmanagement anzusiedeln. Entscheidungen mit langen Vorbereitungszeiten sind dem strategischen Personalmanagement zuzuweisen (Jaeschke 2007, S. 9). Als strukturierte Entscheidungen werden Entscheidungen aufgefasst, die einen geringen Spielraum haben und beinahe automatisiert gefällt werden können. Unstrukturierte Entscheidungen hingegen sind durch Neuartigkeit und starke Komplexität gekennzeichnet, wodurch diese Entscheidungen dem strategischen Bereich zuzuordnen sind (Gluchowski et al. 2008, S. 25).

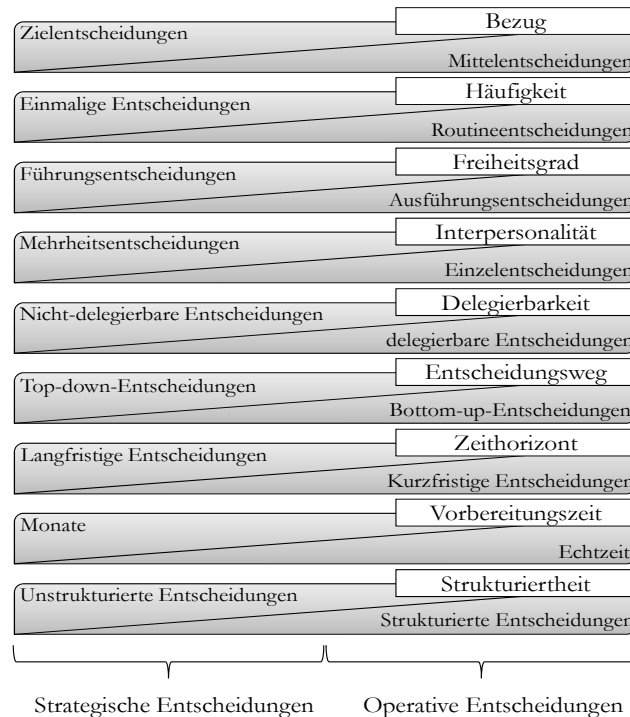


Abbildung 2: Merkmale von (Personal-)Entscheidungen

Daraus folgend definieren sich strategische Personalentscheidungen u. a. durch Zielbezug, Einmaligkeit, einen hohen Freiheitsgrad, Interpersonalität, Nicht-delegierbarkeit, einen Top-down-Entscheidungsweg, Langfristigkeit, längere Vorbereitungszeit und Unstrukturiertheit. Operative Personalentscheidungen sind, wie in Abbildung 2 dargestellt, durch eine entgegengesetzte Ausprägung der Entscheidungsmerkmale charakterisiert.

### 3 Der Entscheidungsfindungsprozess im Personalwesen

Nach Hentze und Kammel (2001, S. 67) setzt sich der personalwirtschaftliche Entscheidungsprozess aus fünf aufeinander folgenden Phasen zusammen (vgl. Abbildung 3). Dieser Entscheidungsprozess lässt erkennen, dass der eigentlichen Entscheidungsphase zwei weitere Phasen vorgelagert sind. Auch nach Berthel und Becker (2007, S. 170) sind die Ermittlung der personalwirtschaftlichen Problemstellung, die Analyse der Durchführungsmöglichkeiten und die Prognose von Konsequenzen sowie die Bewertung der alternativen Maßnahmen Bestandteil der Schaffung einer problemadäquaten Informationsbasis für Entscheidungen. Um eine Entscheidung durch die Bereitstellung von Informationen unterstützen zu können, ist es somit sinnvoll, die Entscheidungsphase sowie die beiden vorgelagerten Phasen der Entscheidungsvorbereitung (Anregungs- und Suchphase) in die Betrachtung einzubeziehen.

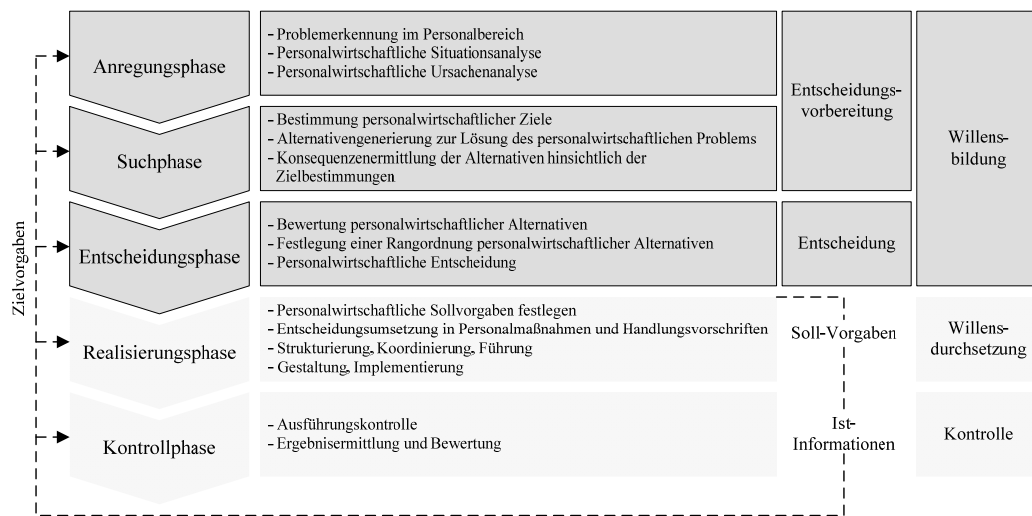


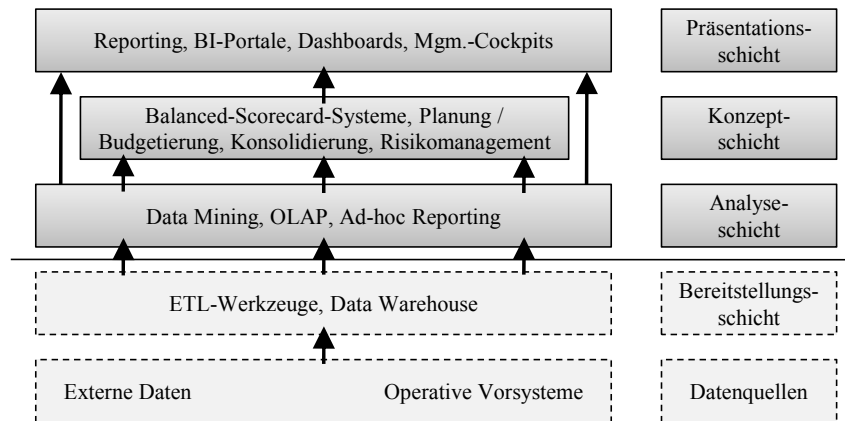
Abbildung 3: Der personalwirtschaftliche Entscheidungsprozess  
(in Anlehnung an Hentze und Kammel 2001, S. 67; Jung 2008, S. 19)

Im Folgenden werden demnach die Anregungs-, Such- und Entscheidungsphase fokussiert, welche in ihrer Gesamtheit als Willensbildungsprozess (Hentze und Kammel 2001, S. 69) verstanden werden. Die rein operativ-dispositive Umsetzung der getroffenen Entscheidungen ist somit kein Gegenstand der Betrachtung.

## 4 Verständnis von Business Intelligence

Der Begriff Business Intelligence wird im Rahmen dieser Arbeit wie folgt verstanden: Business Intelligence ist ein integriertes IT-Gesamtkonzept, das unterschiedliche Technologien und Konzepte zusammenfasst, die der Sammlung und Aufbereitung unternehmensinterner und -externer Daten dienen und diese in Form von geschäftsrelevanten Informationen transparent und verständlich zur Analyse und Entscheidungsunterstützung bereitstellen. Dieses Verständnis erfolgt in Anlehnung an Mertens 2002, S. 65 f., Chamoni und Gluchowski 2004, S. 119 f., Gluchowski und Kemper 2006, S. 12 und Gluchowski et al. 2008, S. 93. Die wesentlichen Komponenten des in Abbildung 4 dargestellten BI-Schichtenmodells sind im Folgenden dargestellt.

Die Aufgaben der Datenbereitstellung erstrecken sich entlang des Datenflusses von der Lieferung interner und externer Daten bis hin zur dauerhaften Ablage und Verwaltung dieser Daten in den entscheidungsorientierten Datenspeichern. Diese Aufbereitung der Daten wird durch sogenannte ETL-Werkzeuge (extract, transform, load) unterstützt. Als Komponenten zur Speicherung von Daten werden heute separate Datenbanksysteme genutzt, die Analytisch-orientiert sind. Diese Lösungen werden als Data Warehouse bezeichnet (Gluchowski und Kemper 2006, S. 14).



*Abbildung 4: Business-Intelligence-Schichtenarchitektur*

*(in Anlehnung an Gluchowski und Kemper 2006, S. 14 f.; Gluchowski et al. 2008, S. 108)*

Die Analyseschicht stellt dem Benutzer Werkzeuge zur Verfügung, mit denen er die Daten nach seinen Anforderungen untersuchen und auswerten kann (Gluchowski et al. 2008, S. 111). Die aktuellen Konzepte der Analysewerkzeuge lassen sich in Ansätze zur Hypothesengenerierung und zur Hypothesenverifizierung unterteilen. Die Hypothesengenerierung vereint Methoden und Verfahren, die im Zusammenhang mit dem Knowledge Discovery in Databases und dem Data Mining stehen. Die Ansätze zur Hypothesenverifizierung, die die Gültigkeit formulierter Hypothesen untersuchen, werden aktuell durch das On-Line Analytical Processing (OLAP) repräsentiert (Gluchowski et al. 2008, S. 143).

Konzeptorientierte Systeme greifen angebotene Funktionalitäten und Konzepte der Analyseschicht auf und nutzen diese für spezielle betriebswirtschaftliche Anwendungsbereiche (Gluchowski und Kemper 2006, S. 17 f.). Dazu zählen vor allem Balanced-Scorecard-Systeme und Tools für das analytische Customer-Relationship-Management sowie für die Planung/Budgetierung, Konsolidierung und das Risikomanagement (Gluchowski et al. 2008, S. 223).

Der Präsentationsschicht werden Präsentations- und Zugangssysteme zugeordnet, die dem Anwender adäquat und arbeitsplatzspezifisch den Zugriff auf Funktionalitäten und Informationen ermöglichen. Zu den BI-spezifischen Bausteinen zählen Management Cockpits und Dashboards, die Entscheidungsträgern eine Navigation durch die Daten des gesamten Unternehmens ermöglichen (Gluchowski und Kemper 2006, S. 17 f.), sowie das Reporting (Gluchowski et al. 2008, S. 205 f.). Performance Dashboards und BI-Portale hängen sind auf Nutzung am Bildschirm ausgerichtet.



## 5 Entscheidungsunterstützung mit Business Intelligence

Die Einsatzpotentiale für Techniken der Business Intelligence zur Entscheidungsunterstützung im Human Resource Management werden im folgenden Kapitel erarbeitet. Den ersten Schritt hierzu stellt die Benennung der Informationen dar, die für die zu treffenden Entscheidungen notwendig sind. Im Anschluss wird überprüft, inwieweit die Business Intelligence Werkzeuge (1) Data Mining, (2) OLAP, (3) Balanced Scorecard-Systeme, (4) Dashboards und (5) Reporting die auf den beiden herausgestellten Ebenen (strategisch & operativ) und in den betrachteten Phasen (Anregungs-, Such- & Entscheidungsphase) des Entscheidungsprozesses im HRM benötigten Informationen liefern können. Die Auswahl dieser fünf Werkzeuge erfolgte in Anlehnung an das in Kapitel 4 dargestellte Schichtenmodell. Aus jeder der drei oberen Ebenen wurde mindestens ein populärer Vertreter gewählt. Das Data Warehouse wurde nicht separat betrachtet. Es wird jedoch unterstellt, dass allein durch das Wesen des im Hintergrund stehenden Datenlieferanten in jedem betrachteten Bereich eine Unterstützung gegeben ist.

### 5.1 Entscheidungsrelevante Informationen

Aufgrund der in Kapitel 2 dargelegten Wesensmerkmale der Entscheidungen und darauf aufbauend einer Unterscheidung in eine strategische und eine operative Ebene sollten auch die auf diesen Ebenen benötigten Informationen getrennt betrachtet werden.

#### *Strategische Ebene*

Das strategische HRM (S-HRM) benötigt in der Anregungsphase aggregierte Daten über Gesamtrisikopositionen (z. B. das kumulierte Abwanderungsrisiko hochqualifizierter Mitarbeiter), um Probleme zu erkennen, die das gesamte Unternehmen betreffen. In der Suchphase werden Informationen benötigt, die Prognosen und Simulationen längerfristiger personalbezogener Entwicklungen ermöglichen und so die Entscheidungsphase direkt mit identifizierten Alternativen beliefern. Die Entscheidungsphase ihrerseits benötigt anschließend Informationen, die eine Bewertung der Alternativen zulassen und das Erstellen einer Rangfolge der Möglichkeiten bis hin zur endgültigen Entscheidung ermöglichen (Hentze und Kammel 2001, S. 67ff.).

### Operative Ebene

Die auf der operativen Ebene (O-HRM) benötigten Informationen lassen sich, wie in Tabelle 1 dargestellt, ebenfalls phasenspezifisch erfassen.

	Informationen Anregungsphase	Informationen Suchphase	Informationen Entscheidungsphase
Akquisition	<ul style="list-style-type: none"> <li>▪ Personalbedarf</li> <li>▪ Personalbestand</li> </ul>	<ul style="list-style-type: none"> <li>▪ Beschaffung intern oder extern möglich</li> <li>▪ Anzahl der benötigten Mitarbeiter (MA)</li> <li>▪ Ziele der MA</li> </ul>	<ul style="list-style-type: none"> <li>▪ Stellung auf dem Arbeitsbeschaffungsmarkt</li> </ul>
Placement		<ul style="list-style-type: none"> <li>▪ Anforderungen der Stelle</li> <li>▪ Können der MA</li> <li>▪ Karriereziele der MA</li> <li>▪ Entwicklungsmöglichkeiten</li> </ul>	
Entwicklung		<ul style="list-style-type: none"> <li>▪ Kenntnis- &amp; Fähigkeitsdefizite der MA</li> <li>▪ Kenntnis- &amp; Fähigkeitsbestand der MA</li> </ul>	<ul style="list-style-type: none"> <li>▪ Entwicklungsziel des Unternehmens</li> <li>▪ Entwicklungsbedarf der MA</li> <li>▪ Ergebnis des Abgleichs</li> </ul>
Motivation		<ul style="list-style-type: none"> <li>▪ Individuelle Bedürfnisse</li> <li>▪ Individuelle Motive</li> </ul>	
Dispensation		<ul style="list-style-type: none"> <li>▪ Konjunkturdaten / -prognosen</li> <li>▪ Rationalisierungsmaßnahmen</li> <li>▪ Technologieeinführungen</li> </ul>	<ul style="list-style-type: none"> <li>▪ Personalbestandsstatistik</li> <li>▪ Arbeitsmarktstruktur</li> </ul>

*Tabelle 1: In den Phasen benötigte Informationen (Nicolai 2009, S. 38-40; Jung 2008, S. 250-257 & S. 315-317; Fargel 2006, S. 29-38)*

Zusätzlich ist eine, in Anlehnung an den Personallebenszyklus durchgeführte, weitere Unterteilung in Akquisition, Placement, Entwicklung, Motivation und Dispensation möglich (Ringlstetter und Kaiser 2008, S. 5). Insbesondere in der Such- und in der Entscheidungsphase erlaubt diese feinere Granularität eine spezifischere Ausdifferenzierung der benötigten Informationen. In der Anregungsphase stellt sich diese Unterscheidung als nicht relevant heraus, da die Identifikation eines Problems beim Personalbestand immer auf dieselben Informationen angewiesen ist.

Um herauszufinden, ob ein Unterstützungspotential gegeben ist, werden die herausgestellten Informationsbedürfnisse in den folgenden Abschnitten damit abgeglichen, ob Business Intelligence diese Information liefern kann.

## 5.2 Unterstützung durch Data Mining

Data Mining ist aufgrund seiner analytischen Natur in der Lage, alle Phasen der strategischen Ebene mit den relevanten entscheidungsunterstützenden Informationen zu versorgen. Am besten lässt sich dies am Beispiel der Abwanderung hochqualifizierter Mitarbeiter verdeutlichen. Mittels Data Mining ist es u. a. möglich mehr über die Hintergründe zu erfahren und die weitere Entwicklung dieses Problems zu analysieren (Strohmeier 2008, S. 164f.).

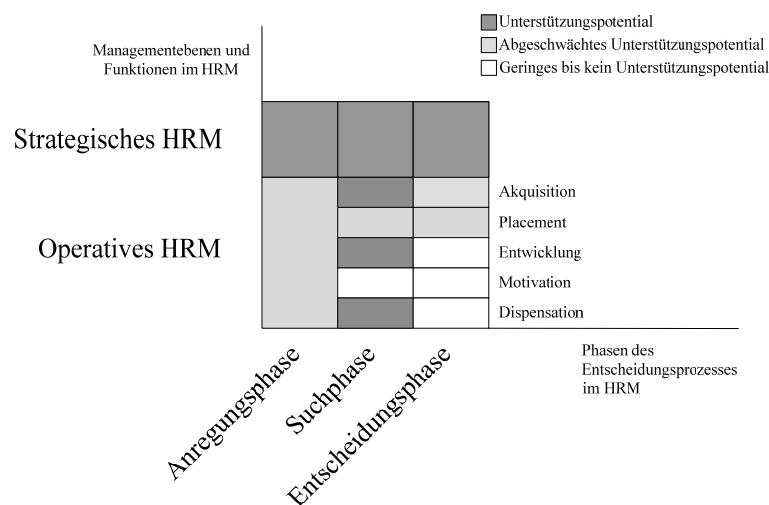


Abbildung 5: Unterstützungspotentiale Data Mining

Ähnlich gut stellt sich die Situation in der operativen Ebene dar. Hier kann das Data Mining u. a. bei der Frage ob die Mitarbeiterbeschaffung intern oder extern erfolgen soll, eine Unterstützung der Entscheidung beispielsweise durch eine Analyse ähnlicher Vergangenheitsdaten liefern. Auch bei der Erhebung des Kenntnis- und Fähigkeitsbestands kann das Data Mining hilfreich sein. Insbesondere moderne Anwendungen des Data Mining, die auch mit unstrukturierten Daten arbeiten können, sind in diesem Schritt hilfreich. In den verbleibenden Bereichen lässt sich nur noch unterscheiden, ob ein vollausgeprägtes Unterstützungspotential (dunkelgrau dargestellt), wie in der Suchphase der Dispensation bei den Konjunkturdaten und –prognosen, oder ein teilausgeprägtes Unterstützungspotential (hellgrau dargestellt), wie bei den Fragestellungen der Anregungsphase, vorliegt. Nur die Entscheidungen weniger Bereiche lassen sich nicht durch Data Mining unterstützen. So kann beispielsweise in der Entscheidungsphase der Entwicklung und der Dispensation

Data Mining die Informationen zur Personalbestandsstatistik oder dem Entwicklungsziel des Unternehmens nicht liefern. Ebenso problematisch sind die Informationen zu den individuellen Bedürfnissen und Motiven in der Such- und Entscheidungsphase der Motivation, wobei sich im Folgenden noch zeigen wird, dass die Motivation generell ein kritischer Bereich für eine Unterstützung durch Business Intelligence ist.

### 5.3 Unterstützung durch Online Analytical Processing

Online Analytical Processing ist ein weiterer Vertreter der Analyseschicht. Das Online Analytical Processing bietet, wie in Abbildung 6 dargestellt, weniger Unterstützungspotential als das Data Mining. Durch sog. What-Analysen eignet sich OLAP u. a. auf der strategischen Ebene für die Anregungs- und Suchphase (Strohmeier 2008, S. 154f.). In der operativen Ebene kann OLAP die Such- und Entscheidungsphase des Placement unterstützen indem es erlaubt, die gespeicherten Daten nach spezifischen Merkmalen zu selektieren und auch das Aufspüren von Korrelationen gestattet (Schüle 2006, S. 224). Als nicht unterstützbar fällt u. a. die Such- und Entscheidungsphase der Akquisition und Dispensation auf, da sich OLAP auf die unternehmensinterne Datenbasis stützt. Ebenso stellen die individuellen Bedürfnisse und Motive der Such- und Entscheidungsphase der Motivation ein Problem für OLAP dar.

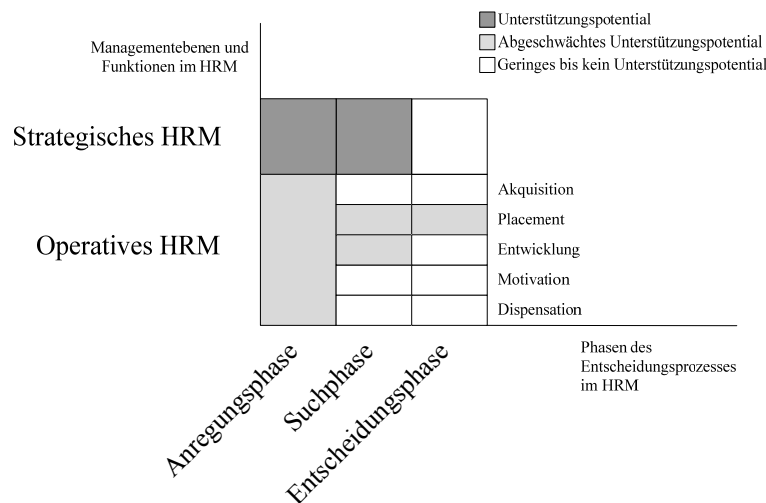


Abbildung 6: Unterstützungspotentiale OLAP

### 5.4 Unterstützung durch Balanced-Scorecard-Systeme

Für die Fragestellungen der strategischen Ebene sind Balanced-Scorecard-Systeme aufgrund ihrer Strategieorientierung geradezu prädestiniert (Kunz 2001, S.61-62). Aus diesem Grund wurden sie als Repräsentant der Konzeptschicht ausgewählt. Einzig in der Entscheidungsphase ist, wenn es um die Bewertung von Alternativen geht, eine Ab-

schwächung zu erkennen (in Abbildung 7 erkennbar durch die schwache Einfärbung des Bereichs). In der operativen Ebene lässt sich aufgrund des Fehlens von strategischen Fragestellungen kein nennenswertes Unterstützungspotential ausmachen.

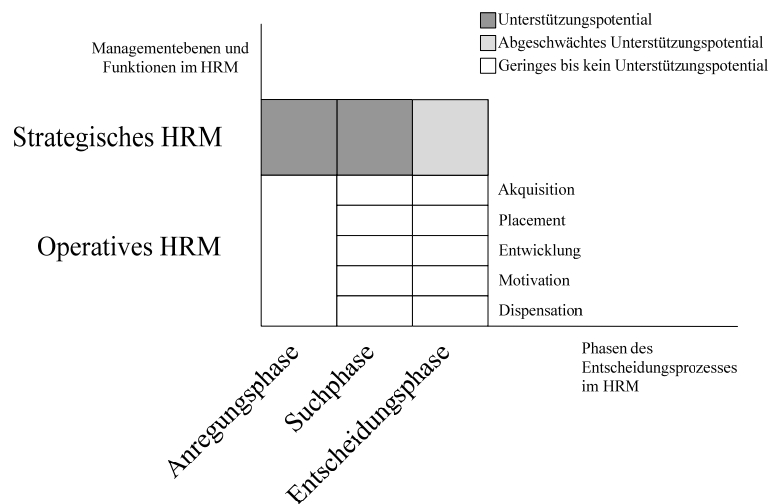


Abbildung 7: Unterstützungspotentiale Balanced-Scorecard-Systeme

## 5.5 Unterstützung durch Dashboards

Dashboards erweisen sich vornehmlich in der Anregungsphase als nützlich. Im Allgemeinen weisen strategische Dashboards eine größere Verbreitung als operative Dashboards auf und zeichnen sich durch die Darstellung stark verdichteter Kennzahlen und vielseitige Interaktionsmöglichkeiten aus (Gluchowski et al. 2008, S. 218).

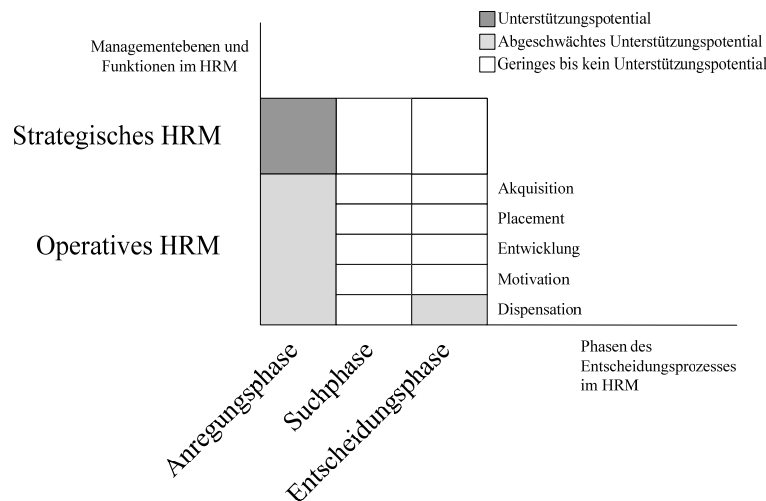


Abbildung 8: Unterstützungspotentiale Dashboards

Die Fragestellungen bzgl. Personalbedarf bzw. -bestand und Gesamtrisikopositionen auf der strategischen Ebenen entsprechen dem, was ein Dashboard an unterstützenden Informationen liefern kann (Strohmeier 2008, S. 155). Aufgrund der hohen Aggregationsstufe eines Dashboard bietet dieses auf der operativen Ebene ein nicht vollausgeprägtes Unterstützungspotential (dargestellt durch den helleren Grauton).

Zusätzlich bieten Dashboards in der Entscheidungsphase der Dispensation und bei der Personalbestandsstatistik ein teilausgeprägtes Unterstützungspotential. In den anderen Feldern der in Abbildung 8 dargestellten Matrize bieten Dashboards aufgrund des Wesens der Fragestellungen im Regelfall kein Unterstützungspotential.

## 5.6 Unterstützung durch Reporting

Reports weisen ähnlich den Dashboards, insb. die Fragestellungen bzgl. Personalbedarf bzw. -bestand und Gesamtrisikopositionen auf der strategischen Ebenen hohes Potential auf (Strohmeier 2008, S. 155). Durch individuellere Konfigurationsmöglichkeiten zeigt das Reporting auch im operativen Bereich das volle Unterstützungspotential (siehe Abbildung 9).

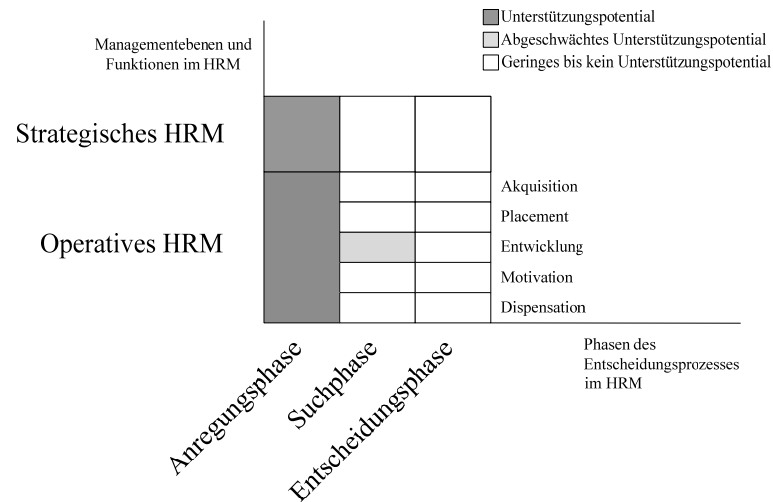


Abbildung 9: Unterstützungspotential Reporting

Außerdem bieten Reports in der Suchphase der Entwicklung bei Bestands- und Defizitfragestellungen ein teilausgeprägtes Unterstützungspotential.

### 5.7 Zusammenführung der Einzelergebnisse

Die Einzelbetrachtung der Unterstützungspotentiale im HRM lässt sich, wie in Abbildung 10 dargestellt, in eine dreidimensionale Darstellung eines Würfels integrieren. Hierfür werden die einzelnen Matrizen der Abbildungen 5 bis 9 perspektivisch hintereinander angeordnet. Zur besseren Übersicht musste die Anordnung der strategischen und der operativen Ebene geändert werden. In der integrierten Gesamtabbildung steht der operative Bereich oben.

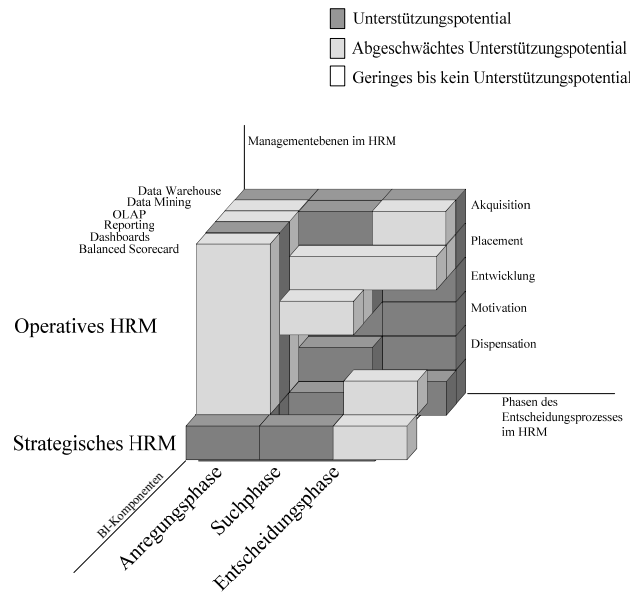


Abbildung 9: Integration der Teilergebnisse

Ausgangspunkt der Überlegungen, die Unterstützungspotentiale der BI bei Entscheidungen im HRM zu identifizieren, war die systematische Aufgliederung des Verlaufs von Entscheidungen durch die Betrachtung des Entscheidungsprozesses im HRM. Die unterschiedlichen Informationsbedarfe, die sich ebenen-, und phasenspezifisch identifizieren ließen, boten die Möglichkeit, bedarfsspezifisch Unterstützungspotentiale einzelner BI-Komponenten auszumachen.

Es zeigte sich, dass die Potentiale in verschiedenen starken Ausprägungen vorliegen und dass es noch einige Bereiche gibt, in denen nur ein abgeschwächtes oder gar kein Potential identifiziert werden konnte. Erkennbar ist dies an den hellgrauen oder gar nicht ausgefüllten Bereichen des Würfels.

## 6 Zusammenfassung

Das Ziel einer anschaulichen Systematisierung wurde anhand der oben stehenden Integration der Teilergebnisse des vorangegangenen Kapitels erreicht. Ob die Darstellungsart in Form eines Würfels die beste Visualisierungsmöglichkeit für diese Systematisierung ist, muss aufgrund der aufgetretenen Überlagerungen (z. B. bei der Kombination Data Warehouse - Placement) in Frage gestellt werden. Weiter muss angemerkt werden, dass sich bei der Feststellung und Unterscheidung der Informationsbedürfnisse die Schwierigkeit gezeigt hat, trennscharfe Zuordnungen vorzunehmen.



Die vorliegende Betrachtung kann keine Auskunft darüber geben, welches Unterstützungspotential der BI in einem bestimmten Unternehmen existiert. Ein ähnlich systematisches Vorgehen anhand des aufgegliederten Entscheidungsprozesses, indem der Informationsbedarf des HRM eines konkreten Unternehmens in Zusammenhang mit einer umfangreichen Informationsbedarfsanalyse (Gluchowski et al. 2008, S. 265) identifiziert wird, gibt jedoch auch einzelnen Unternehmen die Möglichkeit, unternehmensspezifische Unterstützungspotentiale der BI-Komponenten auszumachen.

Aus der Perspektive der Forschung bietet diese Systematisierung einen Ausgangspunkt für weiterführende Betrachtungen. Bei den teilausgeprägten Unterstützungspotentialen können ggf. mehr BI-Techniken der jeweiligen Schicht mit in die Betrachtung aufzunehmen werden, um zu evaluieren, ob durch diese ein vollausgeprägtes Unterstützungspotential erreichbar ist. In den aus verschiedenen Gründen als nicht unterstützbar identifizierten Bereichen können Überlegungen zu einer Erweiterung der Datenbasis sinnvoll sein, um auf diesem Weg Unterstützungspotentiale zu schaffen. Ebenso ist es sinnvoll, die in Tabelle 1 als für die Entscheidungen notwendig erachteten Informationen nicht mit dem Anspruch auf Vollständigkeit zu sehen. Diese Sammlung ist durch eine weiterführende Literatur- oder eine Feldstudie erweiterbar. Eine Ausweitung der als benötigt definierten Informationen in einzelnen Phasen-Ebenen-Kombinationen bietet dann weitere Ansatzpunkte für eine Unterstützung durch Business Intelligence.

## 7 Literatur

Becker J, Holten R, Knackstedt R, Niehaves B (2004) Epistemologische Positionierungen in der Wirtschaftsinformatik am Beispiel einer konsensorientierten Informationsmodellierung. In: Frank U (Hrsg) Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik. Theoriebildung und -bewertung, Ontologien, Wissensmanagement. Gabler, Wiesbaden.

Berthel J, Becker FG (2007) Personal-Management - Grundzüge für Konzeptionen betrieblicher Personalarbeit. Schäffer-Poeschel, Stuttgart.

Chamoni P, Gluchowski P (2004) Integrationstrends bei Business-Intelligence-Systemen. WIRTSCHAFTSINFORMATIK 46(2):119-128.

Fargel YM (2006) Mitarbeiter-Placement – Eine fit-orientierte Perspektive. Deutscher Universitäts-Verlag, Wiesbaden.

Gemünden H (1983) Echte Führungsentscheidungen - Empirische Beobachtungen zu Gutenbergs Idealtypologie. DIE BETRIEBSWIRTSCHAFT 43(1):49-64.

Gluchowski P, Gabriel R, Dittmar C (2008) Management Support Systeme und Business Intelligence - Computergestützte Informationssysteme für Fach- und Führungskräfte. Springer, Berlin.

Gluchowski P, Kemper H (2006) Quo Vadis Business Intelligence? BI-SPEKTRUM (1):12-19.

Hars A (2002) Wissenschaftstheorie für Wirtschaftsinformatiker. Tutorial im Rahmen der Multikonferenz Wirtschaftsinformatik. Nürnberg.

Hentze J, Kammel A (20001) Personalwirtschaftslehre 1. Paul Haupt, Bern.

Jaeschke P (2007) Business Intelligence - Wieviel Intelligence braucht das Business wirklich? In: Deutsche Oracle-Anwendergruppe (DOAG) e.V. (Hrsg): Business Intelligence & Data Warehousing, Tagungsband zum Special Interest Day der Deutschen Oracle-Anwendergruppe (DOAG) e.V., Berlin.

Jung H (2008) Personalwirtschaft. Oldenbourg, München.

Korndörfer W (1999) Unternehmensführungslehre - Einführung, Entscheidungslogik, soziale Komponenten. Gabler, Wiesbaden.

Köster M (2002) Business-Intelligence und Data-Warehouse: Datenfriedhof oder Schatztruhe? [http://www.contentmanager.de/magazin/artikel\\_235\\_business\\_intelligence\\_data\\_warehouse.html](http://www.contentmanager.de/magazin/artikel_235_business_intelligence_data_warehouse.html). Abruf am 2009-01-10.

Kunz G (2001) Die Balanced Scorecard im Personalmanagement. Campus, Frankfurt/Main.

Lory R (2005) Data Warehousing - Warehouse tönt schon fast wie Wahrheit. SCHWEIZER BANK 20(3):44-45.

Mertens P (2002) Business Intelligence - Ein Überblick. INFORMATION MANAGEMENT & CONSULTING 17(Sonderausgabe):65-73.

Müller A (1992) Informationsbeschaffung in Entscheidungssituationen. Wirtschaft & Praxis, Ludwigsburg.

Nicolai C (2009) Personalmanagement. UTB, Stuttgart.

Ringlstetter M, Kaiser S (2008) Humanressourcen-Management. Oldenburg, München.

Schüle H (2006) Strategisches Wissensmanagement als Aufgabe der Managemententwicklung. In: Riekhof H-C (Hrsg) Strategien der Personalentwicklung. Gabler, Wiesbaden.

Strohmeier S (2008) Informationssysteme in Personalmanagement. Vieweg+Teubner, Wiesbaden.

Thom N, Zaugg RJ (2007) Thesen zur Personalentwicklung. In: Thom N, Zaugg RJ (Hrsg) Moderne Personalentwicklung. Gabler, Wiesbaden.

Wilde T, Hess T (2007) Forschungsmethoden der Wirtschaftsinformatik – Eine empirische Untersuchung. WIRTSCHAFTSINFORMATIK 49(4):280-287.

# **Computer Aided Warehouse Engineering auf Basis der Model Driven Architecture**

Christian Kurze, Peter Gluchowski

Technische Universität Chemnitz – Fakultät für Wirtschaftswissenschaften  
Professur Systementwicklung / Anwendungssysteme

## **1 Einleitung**

### **1.1 Motivation und Problemstellung**

Die Model Driven Architecture (MDA) ist ein Standard der Object Management Group zur modellgetriebenen Software-Entwicklung (Object Management Group 2003). Hinter dem Ansatz verbirgt sich im Kern die Idee einer strikten Trennung von Spezifikation und Implementierung eines Systems. Modelle auf verschiedenen Abstraktionsebenen und automatisierte Transformationen zwischen Modellen setzen diese Idee um. Als neuerer Ansatz im Kontext der MDA adressiert die Architecture Driven Modernization (ADM) die Modernisierung bzw. das Reengineering von Systemen sowohl aus fachlicher als auch aus technischer Perspektive und stellt formale Modelle für diesen Kontext bereit (Khusidman und Ulrich 2007, Khusidman, 2008).

Data Warehouses stellen wichtige Informationen zur Unterstützung strategischer und taktischer Entscheidungen bereit, und verkörpern somit zentrale Komponenten moderner Entscheidungsunterstützungssysteme (Inmon 2005). In den letzten zehn Jahren hat das Data Warehousing einen hohen Reifegrad erreicht und flächendeckende Akzeptanz gefunden. Datenschutz, Datensicherheit und Compliance erfordern zunehmend höhere Anstrengungen in der Entwicklung und Dokumentation von Data Warehouses sowie im Metadatenmanagement. Zusätzlich ist eine Schwerpunktverschiebung der Initiativen auf die fachliche Seite – weg von der IT – zu verzeichnen (Kimball, Ross, Thornthwaite, Mundy und Becker 2008). Diese Punkte stellen den Mehrwert einer modellgetriebenen Data Warehouse-Entwicklung heraus: Das Data Warehouse muss Daten in einer für Fachanwender verständlichen Art und Weise bereitstellen. Eine detaillierte Systembeschreibung erhält nunmehr zwei Bedeutungen: Zum einen dient sie als Zusammenstellung der Anforderungen an das System und damit einhergehend als Dokumentation für Endnutzer und Entwickler; gleichzeitig aber auch als solide Basis für die Implementierung

des Zielsystems, da die Befolgung des MDA-Paradigmas die Transformation in ein reales System erlaubt.

Die aktuelle Literatur, die MDA auf die Entwicklung von Data Warehouses anwendet, berücksichtigt den ADM-Ansatz nur unzureichend. Dennoch erweist sich vor allem das Framework von (Mazón und Trujillo, 2008) als ein guter Ausgangspunkt für die Entwicklung von CAWE-Werkzeugen auf Basis des MDA-Paradigmas. Die Autoren stellen eine Fallstudie vor, um die Machbarkeit des vorgestellten Ansatzes aufzuzeigen, geben allerdings keine konkreten Hinweise zur Implementierung. Ähnlich der genannten Arbeit konzentriert sich der vorliegende Beitrag auf die Datenhaltungsschicht einer Data Warehouse-Lösung – speziell auf die Modellierung mehrdimensionaler Datenstrukturen. Dabei repräsentiert der ADM-Ansatz ein solides und tragfähiges Fundament, um auf bestehenden Data Warehouse-Systemen ein Reverse-Engineering durchzuführen.

Das hier vorgestellte Werkzeug ist sowohl für die Wissenschaft als auch für Praktiker relevant. Es stellt eine Plattform für Forscher bereit, um aktuelle Methoden des Software-Engineerings auf die Spezifika des Data Warehousing anzugleichen; Praktikern hilft es bei der Bewältigung von komplexen Fragestellungen im Rahmen der (Weiter-) Entwicklung von Data Warehouse-Systemen.

Aus den bisherigen Ausführungen ergeben sich unmittelbar zwei zentrale Forschungsfragen: Wie sind die Ansätze MDA und ADM in eine Architektur für ein CAWE-Werkzeug, das zunächst die Aspekte der Datenmodellierung abdeckt, zu integrieren? Wie lässt sich eine solche Architektur implementieren?

Die vorliegende Arbeit stellt sich der Aufgabe, diese Lücke zu schließen. Die präsentierte Software-Architektur integriert sowohl MDA als auch ADM. Bereits existierende Frameworks zum effizienten Umgang mit Modellen werden auf das Problem angewendet, um durch eine prototypische Implementierung die Umsetzbarkeit der Architektur zu zeigen.

## **1.2 Forschungsmethodik**

Der Design Science Ansatz verfolgt die Zielstellung, praktischen und theoretischen Herausforderungen durch die Erstellung und Evaluation von IT-Artefakten zu begegnen. Die so erstellten Artefakte lösen die aufgezeigten organisatorischen Probleme (March und Smith 1995, Hevner, March, Park und Ram 2004, Peffers, Tuunanen, Rothenberger und Chatterjee 2008). Der von (Peffers et al. 2008) vorgestellte Forschungsprozess findet in der vorliegenden Arbeit wie folgt Anwendung:

- *Problemidentifikation und Motivation:* Das Einleitungskapitel legte die Problemstellung immer komplexerer Data Warehouse-Architekturen dar und motivierte die Verwendung von MDA-Konzepten als tragfähigen Lösungsansatz.
- *Definition der Zielstellungen der Lösung:* Die gegebenen Forschungsfragen repräsentieren die Zielstellung: Es ist eine Software-Architektur zu entwickeln, die es erlaubt, MDA und ADM für die Modellierung mehrdimensionaler Datenstrukturen einzusetzen.
- *Design und Entwicklung:* In dieser Phase finden verwandte Arbeiten Anwendung, um die gesetzten Ziele zu erreichen. Die Architektur wird durch einen Prototypen exemplarisch umgesetzt.
- *Demonstration:* Ein Proof-of-Concept zeigt die Eignung des erstellten Prototyps zur Lösung der gestellten Problematik.
- *Evaluation:* Im Rahmen der Evaluation ist der Prototyp auf Realwelt-Probleme anzuwenden. Die hier identifizierten Ergänzungen dienen als Input für einen Kreislauf zur weiteren Verbesserung und Erweiterung.

## 2 Verwandte Arbeiten und Zielstellungen von CAWE

### 2.1 Model Driven Architecture (MDA) und Architecture Driven Modernization (ADM)

Die Object Management Group (OMG) vereint als non-profit Industriekonsortium Unternehmungen um den ganzen Globus. Der Zweck seiner Gründung besteht in Beschleunigung sowie Senkung von Komplexität und Kosten in der Softwareentwicklung. Detaillierte Spezifikationen führen zu interoperablen, wiederverwendbaren sowie portablen Softwarekomponenten auf Basis standardisierter Modelle (Object Management Group 2003).

Der Mehrwert der MDA liegt in der Möglichkeit, maschinenlesbare Applikations- und Datenmodelle zu erstellen, die eine langfristige Flexibilität von Implementierung, Integration, Wartung, Test und Simulation sicherstellen. Die bereits genannte strikte Trennung zwischen funktionaler Systemspezifikation und Implementierungsdetails einer konkreten Zielplattform mündet in der Einführung von drei sogenannten Viewpoints auf ein System. Im Kontext der MDA ist ein Viewpoint als Abstraktionsstufe zu verstehen, die für die jeweilige Stufe unnötige Details zur Erlangung eines vereinfachten Modells ausblendet. Jeder Viewpoint wird durch ein korrespondierendes Modell repräsentiert. Das

*Computation Independent Model* (CIM), auch Domänen-Modell genannt, nutzt als implementierungsunabhängige Sicht das Vokabular der Anwendungsdomäne. Hauptzweck ist es, eine Brücke zwischen Domänenexperten und Entwicklern zu schlagen. Das *Platform Independent Model* (PIM) bleibt in dem Sinne plattformunabhängig, dass es für verschiedene Plattformen ähnlichen Typs zur Anwendung kommt. Die Ergänzung des PIM um zusätzliche Details bezüglich der Umsetzung auf einer konkreten Zielplattform führt zum *Platform Specific Model* (PSM) (Object Management Group 2003). Aufgrund verschieden starker Ausdrucksstärken von Modellierungssprachen ist es durchaus möglich, dass mehrere, zueinander in Beziehung stehende, Modelle zur Beschreibung eines Systems herangezogen werden. Dieser Sachverhalt ist vor allem auf CIM-Ebene von Relevanz. Ergänzend zu Modellen auf verschiedener Abstraktionsebene spielen Modelltransformationen eine zentrale Rolle in der MDA: Sie konvertieren verschiedene Modelle des gleichen Systems ineinander; beispielsweise das CIM in ein PIM und dieses wiederum in ein PSM und anschließend in Code. Abbildung 1 fasst das beschriebene Gesamtkonzept zusammen.

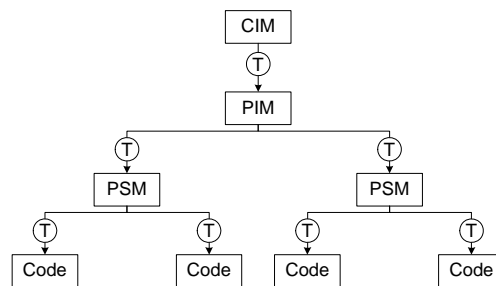


Abbildung 1: Grundkonzept der MDA (Mazón und Trujillo 2008)

Die Meta Object Facility (MOF) stellt eine vierstufige Metamodellierungsarchitektur als Rahmen für MDA bereit. Jede Ebene definiert die jeweils darunter liegende: Die unterste Ebene enthält die zu beschreibenden Daten. Auf zweiter Ebene finden sich Modelle zur Beschreibung der Daten – sogenannte Metadaten. In der dritten Ebene sind Metamodelle anzusiedeln, die wiederum die Struktur der Metadaten definieren. Das Meta-Metamodell auf vierter Ebene beinhaltet die Beschreibung von Struktur und Semantik der Metamodelle. Es stellt also die abstrakte Sprache zur Beschreibung verschiedener Arten von Metamodellen bereit (Object Management Group 2002, Object Management Group 2006).

Die Architecture Driven Modernization (ADM)-Task Force der OMG setzt sich eine Spezifikation zur Modernisierung bereits existierender Applikationen zum Ziel (Object Management Group 2009). Den Kern des Ansatzes bildet das in Abbildung 2 dargestellte Hufeisenmodell (Kazman, Woods und Carrière 1998, Khusidman und Ulrich 2007, Khusidman 2008). Es teilt sich horizontal in drei Ebenen: *Business Architecture*, *Applica-*

tion and Data Architecture sowie *Technical Architecture*. Vertikal ist zwischen IST- und SOLL- bzw. Quell- und Zielsystem zu unterscheiden. Das Hufeisen beschreibt verschiedene Pfade von der ursprünglichen Lösung zur neuen Zielarchitektur. Der kürzeste Weg verbleibt dabei innerhalb einer Ebene. Veränderungen innerhalb der technischen Architektur, z.B. neue Hardware, beeinflusst lediglich eine Ebene. Der längste Transformationspfad, der in Abbildung 2 gezeigt ist, verläuft durch alle drei Ebenen. Unabhängig von dessen Länge ist jeder Transformationspfad durch drei Schritte gekennzeichnet: Strukturextraktion aus bestehenden Systemen, Definition der Zielarchitektur sowie die Transformation des IST-Systems in das spezifizierte SOLL-System.

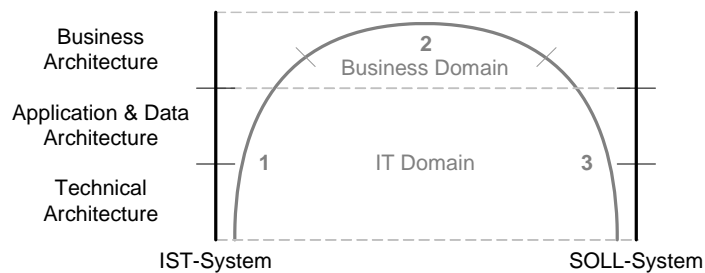


Abbildung 2: Prinzip der Architecture Driven Modernization  
(Khusidman und Ulrich 2007)

## 2.2 Data Warehousing und MDA/ADM

Im Kontext des Data Warehousings existieren verschiedene CIMs, beispielsweise ADAPT (Bulos und Forsman 2006), DFM (Golfarelli, Maio und Rizzi 1998) oder ME/R (Sapia, Blaschka, Hofling und Dinter 1998). Allerdings konnte sich bislang keiner dieser Ansätze als allgemein akzeptierte multidimensionale Modellierungstechnik durchsetzen. Insgesamt erweist sich die Definition eines Modellierungskonzeptes, welches alle mehrdimensionalen Besonderheiten abdeckt, als schwierig. Auf PIM/PSM-Ebene konkurrieren drei Ansätze gegeneinander: Das auf relationalen Datenbanken basierende ROLAP, auf multidimensionalen Datenbanken basierendes MOLAP sowie eine hybride Kombination beider Herangehensweisen, HOLAP (Chaudhuri und Dayal 1997). Der Unterschied zwischen PIM und PSM für ein ROLAP-System besteht beispielsweise darin, dass ein PIM weder Datentypen noch konkrete Indexierungsverfahren beschreibt; ein PSM hingegen auf die speziellen Gegebenheiten eines konkreten Systems eingeht.

In diesem Punkt unterscheidet sich die Auffassung der Autoren gegenüber der von (Mazón und Trujillo 2008). Grafische Notationen wie DFM oder ME/R sind in der vorliegenden Arbeit den CIMs zugerechnet. Gemäß Definition sind diese Modelle zur Kommunikation zwischen Endanwendern und Entwicklern gedacht (Object Management Group 2003, S. 2-5). PIMs sind für eine Reihe verschiedener Plattformen ähnlichen Typs

konzipiert (Object Management Group 2003, S. 2-6). Ein relationales Star-Schema beispielsweise ist für verschiedene konkrete relationale Datenbanksysteme geeignet. Das PSM reicht ein PIM mit Spezifika einer konkreten Plattform (Object Management Group 2003, S. 2-6) dahingehend an, dass, am Beispiel des Star-Schemas, spezifische Datentypen, Indexierungsmechanismen, Tabellenpartitionierungen und weitere Merkmale eines konkreten Datenbanksystems angegeben sind.

Das Framework von (Mazón und Trujillo 2008) deckt verschiedene Ebenen des Data Warehousings ab: Datenquellen, ETL-Prozesse, mehrdimensionale Datenmodellierung eines relationalen Data Warehouses, mehrdimensionale Würfel sowie Applikationsmodelle, beispielsweise für das Reporting, die freie Datenanalyse oder das Data Mining. Im Gegensatz zur genannten Arbeit fassen die Autoren die Modellierung eines relationalen Warehouses und mehrdimensionaler Würfel in eine Ebene der mehrdimensionalen Datenmodellierung zusammen. Diese Zusammenfassung resultiert aus der Unterscheidung zwischen ROLAP, MOLAP und HOLAP. Jede Realisierungsform erfordert verschiedene Beschreibungen. Im Falle von HOLAP existieren beide Beschreibungsformen und bedingen einander. Jede Schicht einer Data Warehouse-Anwendung wird von allen drei Viewpoints (CIM, PIM, PSM) aus modelliert und anschließend in Code transformiert. Die CIM-Ebene stellt dabei den wesentlichen Input für mehrdimensionale Daten- und Applikationsmodelle auf PIM-Ebene dar. Abbildung 3 fasst das Rahmenkonzept zusammen.

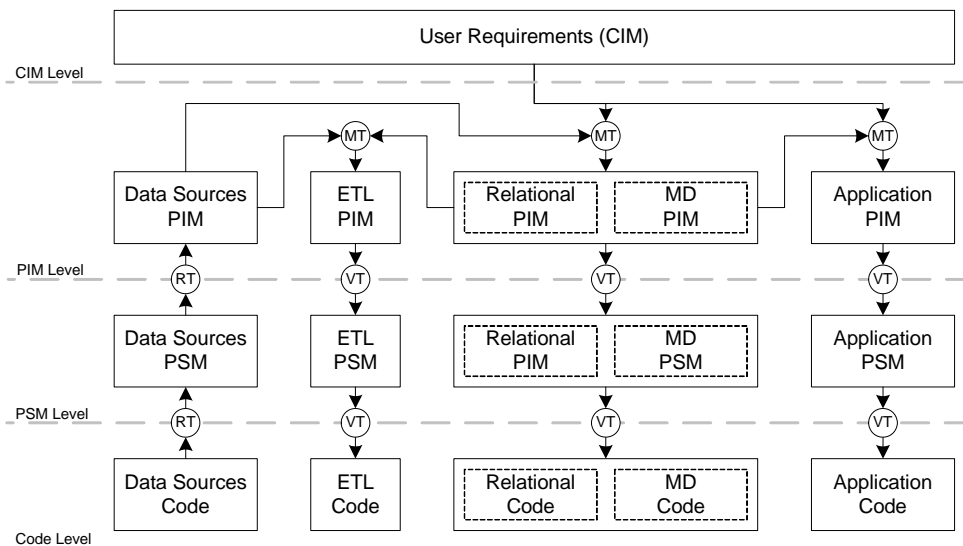


Abbildung 3: MDA-Framework für die Entwicklung von Data Warehouses  
(Mazón und Trujillo 2008)



Die Überführung von Modellen ineinander erfordert eine Reihe von Transformationen innerhalb des Frameworks. Zur Verwendung in einem modellgetriebenen Vorgehen ist es notwendig, zunächst sämtliche Quellsysteme durch Modelle zu beschreiben. In den meisten Fällen existieren diese nicht und sind, ausgehend von Quelltexten bzw. operationalen Datenbanksystemen, zu entwickeln. Diese Transformationen sind als RT (Reverse Transformation) in Abbildung 3 dargestellt. Transformationen, die mehrere Modelle zu einem Zielmodell zusammenführen, sogenannte Merging Transformations, sind als MT abgebildet. Ein ETL-Modell auf PIM-Ebene vereint beispielsweise Modelle der Datenquellen mit Modellen der mehrdimensionalen Zielsysteme. Vertikale Transformationen (VT) transformieren Modelle in darunterliegende Abstraktionsebenen in Richtung Code.

Das beschriebene Framework beinhaltet ADM-Mechanismen lediglich in der Quellsystem-Ebene bei der Erstellung von PIMs aus operativen Systemen heraus. Aus Sicht der Autoren ist jedoch ein formaler und systematischer Weg zur Unterstützung des Verstehens, Anpassens und Weiterentwickelns von gesamten Data Warehouse-Lösungen nötig. Diese Argumentation führt zur Postulation eines ADM-Ansatzes auf allen Ebenen des Data Warehousings: ETL, mehrdimensionale Daten und Applikationen. Ein derartiges Vorgehen ist auch für die Einführung eines modellgetriebenen Entwicklungsvorgehens für bereits im Einsatz befindliche Data Warehouse-Systeme unerlässlich: Im ersten Schritt sind für alle Systemelemente geeignete Modelle zu erstellen. Aufgrund der semantischen Lücke zwischen CIM und PIM stellt sich diese Aufgabe als schwierig heraus. Um Data Warehouse-Entwickler bestmöglich zu unterstützen, ist ein automatisiertes Verfahren für diesen Arbeitsschritt zu entwickeln.

Die Ausführungen zeigen eine weitere Anforderung an die Architektur auf: Sie muss komplexe Algorithmen für das Entdecken von mehrdimensionalen Elementen innerhalb von PIMs und deren Transformation in Elemente eines CIMs verwalten und ausführen können. Da den einzelnen Elementen zusätzliche Semantik hinzuzufügen ist, sind die Ergebnisse verschiedener Algorithmen zu evaluieren und in ein resultierendes CIM zu integrieren.

### **2.3 Fusioniertes CIM**

Das von uns vorgeschlagene fusionierte, implementierungsunabhängige Modell basiert auf der Analyse verschiedener Quellen zur konzeptionellen multidimensionalen Modellierung und fußt auf dem UML-Profil von (Luján-Mora, Trujillo und Song 2006). Die darin enthaltenen Abstraktionsstufen sind, speziell für sehr große Modelle, von essenzieller Bedeutung (Kurze und Gluchowski 2009). Die aktuelle Implementierung lässt die Abstrahierung jedoch zur Reduktion der Komplexität vorerst außen vor. Das von (Pedersen, Jensen und Dyreson 2001) aufgestellte konzeptionelle mehrdimensionale Modell gilt

als eines der umfangreichsten Modelle auf implementierungsunabhängiger Ebene (Mansmann und Scholl 2007, S. 36). Die darin aufgestellten elf Regeln, die ein mehrdimensionales Modell zu befolgen hat, werden zur Erweiterung des Ausgangsmodells herangezogen. Eine formale Beschreibung stellen (Mansmann und Scholl 2007) vor. Weiterhin finden Klassifikationen von Hierarchien (Malinowski und Zimányi 2004, Malinowski und Zimányi 2006) sowie Kennzahlen-Dimensions-Beziehungen (Mazón, Lechtenböcker und Trujillo 2008, Mazón, Lechtenböcker und Trujillo 2009) Anwendung. Hauptziel ist das Abdecken von nahezu jeder möglichen Besonderheit der mehrdimensionalen Datenmodellierung. In realen Anwendungen finden idealtypische Schemata aufgrund komplizierter Anforderungen nur selten Anwendung.

In der ersten Iteration der Metamodell-Entwicklung liegt der Fokus auf der granularsten Ebene von (Luján-Mora et al. 2006). Die Konzepte degenerierter Fakten und Dimensionen finden im Ansatz des vorliegenden Papers keine Betrachtung, da es sich hierbei aus Sicht der Autoren um logische Konstrukte handelt, die auf CIM-Ebene noch nicht relevant sind.

Folgende Anforderungen von (Pedersen et al. 2001) sind in der ersten Iteration nicht integriert worden: Unterstützung der Aggregationssemantik von Kennzahlen bezüglich der Dimensionen, symmetrische Behandlung von Kennzahlen und Dimensionen (die beispielsweise die Verwendung von Kundenalter als Kennzahl zur Durchschnittsbildung, aber auch zum Filtern einzelner Kunden erlaubt), temporale Aspekte wie beispielsweise die Veränderung von Dimensionsinstanzen, Umgang mit verschiedenen Granularitäten sowie das Handling von unpräzisen Daten.

### 3 Architektur und Implementierung von CAWE

#### 3.1 Konzeptionelle Architektur

Aus den zugrunde gelegten Aspekten der Metamodellierung ergeben sich die wesentlichen Anforderungen an CAWE-Werkzeuge. Ausgangspunkt der in diesem Kapitel vorgestellten Architektur bildet deshalb ein generischer Ansatz zur Anordnung von Systemkomponenten von (Karagiannis und Kühn 2002). Wie *Abbildung* zeigt, stellt das *Meta<sup>2</sup>-Modell* die Kernkonzepte zur Erzeugung von Metamodellen und Mechanismen bereit. Es steht im Zentrum der Architektur und ist mit allen weiteren Komponenten verbunden. Modelle und Metamodelle sind im *Modell-* bzw. *Metamodell-Repository* abgelegt. Beide Repositories stehen zueinander in Beziehung, um Veränderungen an Metamodellen an die entsprechenden Modelle zu propagieren, d.h. sie synchron zu halten. Auf Modelle und Metamodelle anwendbare Funktionen sind im *Mechanismen-Repository* gespeichert. Ihre Ablage erfolgt entweder komplett oder als externe Referenz im Repository. Im zweiten

Fall befindet sich zusätzlich eine Schnittstellenbeschreibung zur Interaktion mit dem Baustein im Repository. Der *Persistenzierungsdienst* ist aus den einzelnen Datenspeichern ausgegliedert: Er verwaltet die persistente Ablage sämtlicher (Meta-) Modelle und Mechanismen. Der Zugriff erfolgt transparent und unabhängig von der gewählten physischen Speichermethode wie beispielsweise Datenbanksystemen oder dem Dateisystem. Zusätzlich ermöglicht der Dienst die Verteilung von (Meta-) Modellen, d.h. sehr große Modelle können in kleineren Teilmodellen persistenziert werden. Zugriff auf die einzelnen Komponenten erfolgt über die *Zugriffsdienste* per API oder Dateiaustausch. *Viewer und Builder-Komponenten* an der Spitze der Architektur erlauben Nutzung und Wartung der gesamten Plattform.

Die gestellten Anforderungen der Abschnitte 2.1 und 2.2 sind durch die vorgestellte Architektur erfüllt. Sie dient als Basis für die Entwicklung von Metamodellierungswerkzeugen und deckt somit die benötigten Metamodellierungsaspekte ab. Modelltransformationen sind im Mechanismen-Repository abgelegt und lassen sich auf (Meta-) Modelle anwenden. ADM-Spezifika sind ebenfalls abbildbar: Modell-Discovery aus bestehenden Systemen erweist sich als Mechanismus; die Integration komplexer Algorithmen für diese Aufgabe erfolgt aus externer Ablage. Transformationen von PSMs in PIMs sind ebenfalls im Mechanismen-Repository gespeichert. Ein fusioniertes CIM, wie in Abschnitt 2.3 vorgestellt, ist im Metamodell-Repository bzw. dessen Instanz im Modell-Repository abzulegen.

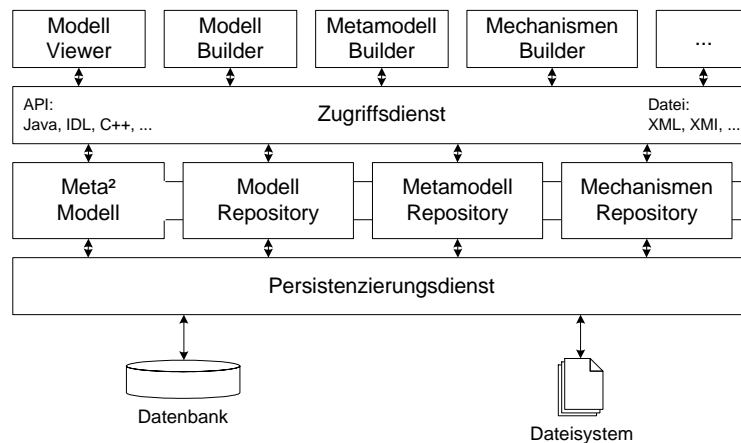


Abbildung 4: Konzeptionelle CAWE-Architektur (Karagiannis und Kühn 2002)

### 3.2 Implementierung der Komponenten

Die gewählte Implementierung basiert auf dem Eclipse Modeling Project (Gronback 2009). Es handelt sich dabei um ein Open Source-Projekt mit einer großen Gemeinde an Entwicklern, die permanent an dessen Evolution arbeiten. Als Meta²-Modell findet das

Ecore-Metamodell des Eclipse Modeling Frameworks (EMF) Anwendung (Steinberg, Budinsky, Paternostro und Merks 2009). Die Modell- und Metamodell-Repositories sind derzeit durch einen dateibasierten Persistenzierungsdienst unterstützt. Das Mechanismen-Repository dient hauptsächlich zur Ablage von Modelltransformationen. Komplexere Algorithmen, insb. zum Modell-Discovery, sind in Java geschrieben. Sämtliche Metamodelle können in Form von Eclipse-Plugins bereitgestellt werden und unterstützen somit einen transparenten Zugriff per API im Sinne der Zugriffsdienste; ein Datei-Export ist ebenso möglich. Viewer- und Builder-Komponenten sind im Rahmen von EMF ebenfalls vorhanden. Integrierte, baumbasierte Editoren lassen sich automatisch aus Metamodellen generieren (Abbildung 7 zeigt Beispiele); ein baumbasierter wie auch ein grafischer Editor für Metamodelle stehen zur Verfügung.

Abbildung 5 zeigt das Prinzip der Modell-Discovery: Das Schema relationaler Datenbanksysteme ist in Form von Modellen – korrespondierend zu deren Metamodellen – abzubilden. Der Discoverer liest das Datenbankschema und legt ein Modell entsprechend der gefundenen Strukturen an. Die Implementierung ist der MoDisco-Toolbox des Eclipse Modeling Projects entnommen und entsprechend angepasst.

Das erstellte Schemamodell, das in relationalen Datenbanksystemen aufgrund deren ausgeprägten Metadaten-Repositories stets erstellt werden kann, erlaubt die Anwendung von Regeln zum Aufdecken der typischen Snowflake-Struktur.

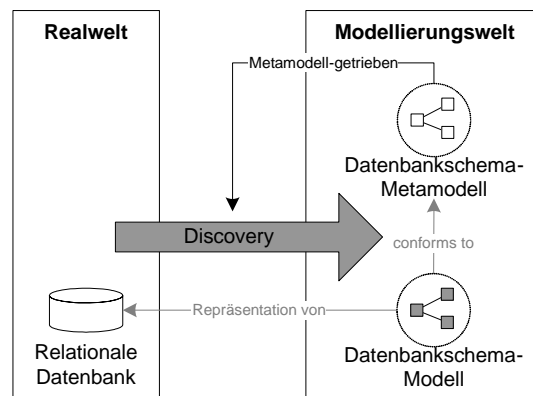


Abbildung 5: Prinzip von Model-Discovery für relationale Datenbanken

## 4 Anwendung einer Fallstudie

Zur Evaluation der vorgestellten Lösung kommt die Fallstudie eines Unternehmens zum Einsatz, das seine Absatzzahlen (Absatzmenge, Preis und Umsatz) hinsichtlich verschiedener Produkte, Organisationseinheiten und Kunden über die Zeit hinweg analysieren möchte. Dieses Beispiel lässt sich mithilfe eines Würfels (Sales) mit vier Dimensionen

(Organisational Unit, Customer, Product und Time) modellieren. Die gesamte Kette verschiedener (Meta-) Modelle ist in Abbildung 6 aufgeführt. Metamodelle sind durch Kreise mit weißen Quadraten, Modelle mit grauen Quadraten dargestellt. Die Ausführung einer Transformation repräsentiert eine Raute mit grauen Quadraten, wobei MM eine Modell zu Modell-Transformation, MT eine Modell zu Text-Transformation und MS Modell-Discovery (im Sinne einer System zu Modell-Transformation) kennzeichnet.

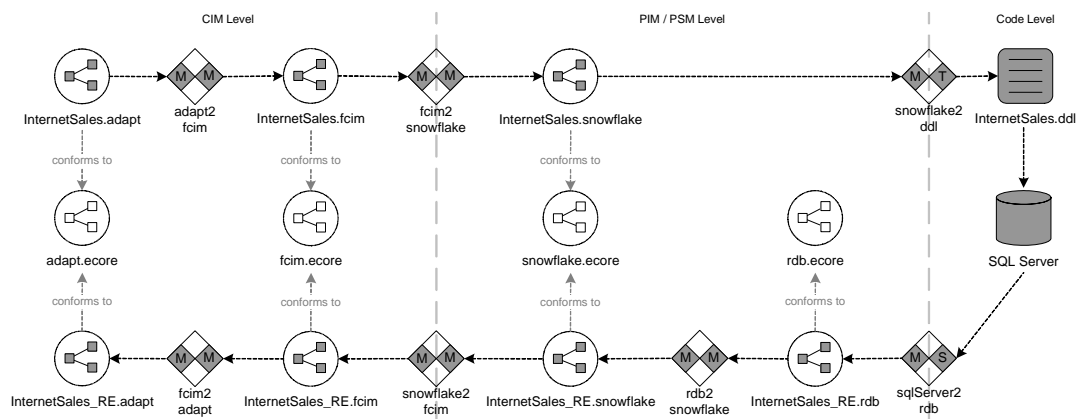


Abbildung 6: Kette von (Meta-) Modellen der Fallstudie

Das ADAPT-Modell ist zur Erhöhung der Wiederverwendbarkeit, gemäß den Ausführungen in Kapitel 2.3, zunächst in ein übergreifendes CIM zu transformieren. Aus diesem generalistischen Modell entsteht durch Transformation ein Snowflake-Schema-Modell. Das zugehörige Metamodell ist eine Erweiterung des relationalen Modells um eine zusätzliche Klassifikation der Tabellen, d.h. Fakten- bzw. Dimensionstabellen, und entstammt maßgeblich aus dem Relational Package des CWM. Im nächsten Schritt sind den einzelnen Attributen der Relationen Datentypen zuzuweisen. Dieser Schritt erfolgt derzeit manuell. Aufgrund systemspezifischer Datentypen stellt dieser Schritt die Transformation des PIM in ein PSM dar; das InternetSales.snowflake-Modell ist deshalb in den Ebenen PIM und PSM zugeordnet. Ein Java Emitter Template (JET) generiert DDL-Code zur Implementierung in Microsoft SQL Server. Es erstellt Tabellen, Spalten und Fremdschlüsselbeziehungen zwischen den verschiedenen Tabellen. In der Reverse-Richtung ist das Ergebnis des Modell-Discoverers eine relationale Datenbankbeschreibung, in der durch eine zusätzliche Transformation einzelne Tabellen entsprechend ihrer Funktion (Fakt oder Dimension) klassifiziert sind.

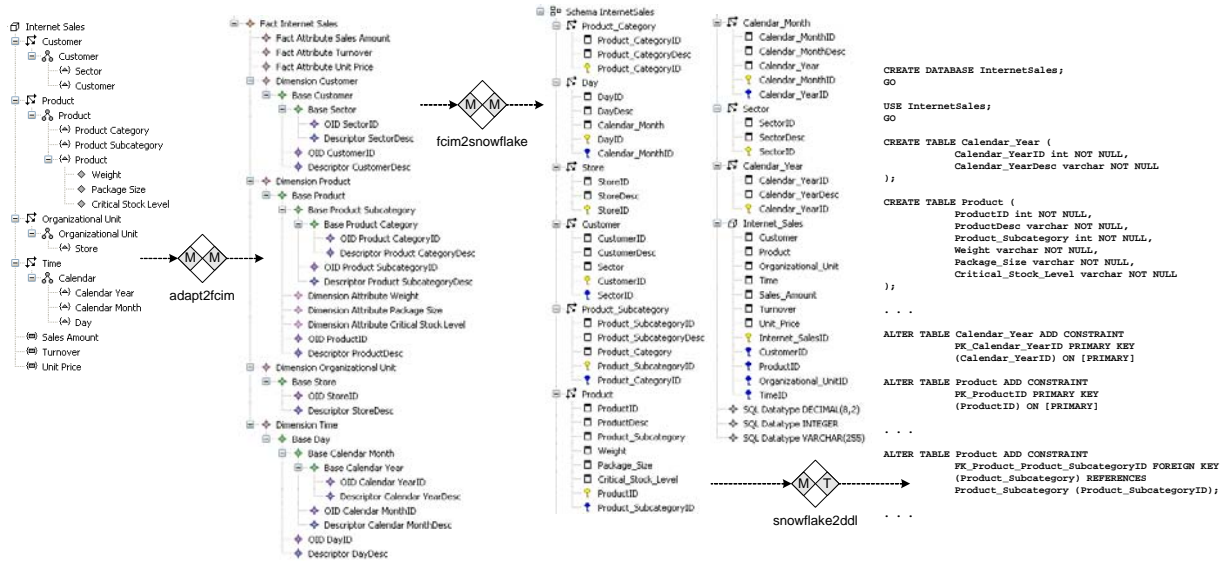


Abbildung 7: Forward-Engineering der Fallstudie

## 5 Kritische Würdigung und weitere Arbeitsschritte

Das vorliegende Paper stellt Forschungsergebnisse dar, die in weiteren Arbeiten zu vertiefen sind. Dabei spielen vor allem die Erweiterung des fusionierten CIM sowie die Verbesserung der Transformation von Modellen relationaler Datenbanken in mehrdimensionale Modelle eine entscheidende Rolle. Obwohl die vorliegende Arbeit nur beschränkt generalisierbar ist, zeigt sie das Potenzial von CAWE sowie dessen Umsetzbarkeit. Weitere Arbeitsschritte müssen folgen, um die prototypische Anwendung in einem breiten Rahmen einsetzen zu können.

## 6 Literatur

Bulos, D. und Forsman, S. (2006) Getting Started with ADAPT. Whitepaper, [http://symcorp.com/downloads/ADAPT\\_white\\_paper.pdf](http://symcorp.com/downloads/ADAPT_white_paper.pdf). Letzter Abruf am 13. September 2009.

Chaudhuri, S. und Dayal, U. (1997) An overview of data warehousing and OLAP technology. *ACM SIGMOD Record*, 26(1), 65-74.

Golfarelli, M., Maio, D. und Rizzi, S. (1998) The Dimensional Fact Model: A Conceptual Model for Data Warehouses. *International Journal of Cooperative Information Systems*, 7(2-3), 215-247.

Gronback, R. C. (2009). Eclipse Modeling Project: A Domain-Specific Language (DSL) Toolkit. Addison-Wesley.

Hevner, A. R., March, S. T., Park, J. und Ram, S. (2004) Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105.

Inmon, W. H. (2005) Building the Data Warehouse, 4. Aufl. Wiley.

Karagiannis, D., & Kühn, H. (2002). Metamodelling Platforms. In *E-Commerce and Web Technologies*, Springer, 451-464.

Kazman, R., Woods, S. G. und Carrière, S. J. (1998) Requirements for Integrating Software Architecture and Reengineering Models: CORUM II. In *Proceedings of Working Conference on Reverse Engineering*, 154-163.

Khusidman, V. (2008) ADM Transformation. Whitepaper, <http://www.omg.org/cgi-bin/doc?admtf/2008-06-10>. Letzter Abruf am 13. September 2009.

Khusidman, V. und Ulrich, W. (2007) Architecture-Driven Modernization: Transforming the Enterprise. Whitepaper, <http://www.omg.org/cgi-bin/doc?admtf/2007-12-01>. Letzter Abruf am 13. September 2009.

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. und Becker, B. (2008) *The Data Warehouse Lifecycle Toolkit*, 2. Aufl. Wiley.

Kurze, C. und Gluchowski, P. (2009) Towards Principles for Managing and Structuring Very Large Semantic Multidimensional Data Models. In *Proceedings of the 15th Americas Conference on Information Systems*.

Luján-Mora, S., Trujillo, J. und Song, I. (2006) A UML profile for multidimensional modeling in data warehouses. *Data & Knowledge Engineering*, 59(3), 725-769.

Malinowski, E., & Zimányi, E. (2004). OLAP Hierarchies: A Conceptual Perspective. In *Advanced Information Systems Engineering*, Springer, 477-491.

Malinowski, E. und Zimányi, E. (2006) Hierarchies in a multidimensional model: from conceptual modeling to logical representation. *Data & Knowledge Engineering*, 59(2), 348-377.

Mansmann, S. und Scholl, M. H. (2007) Empowering the OLAP Technology to Support Complex Dimension Hierarchies. *International Journal of Data Warehousing and Mining*, 3(4), 31-50.

March, S. T. und Smith, G. F. (1995) Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251-266.

Mazón, J., Lechtenbörger, J. und Trujillo, J. (2008) Solving summarizability problems in fact-dimension relationships for multidimensional models. In *Proceeding of the ACM 11th international workshop on Data warehousing and OLAP*, 57-64.

Mazón, J., Lechtenbörger, J. und Trujillo, J. (2009) A survey on summarizability issues in multidimensional modeling. *Data & Knowledge Engineering*, In Press, doi: 10.1016/j.datak.2009.07.010.

Mazón, J. und Trujillo, J. (2008) An MDA approach for the development of data warehouses. *Decision Support Systems*, 45(1), 41-58.

Object Management Group. (2003) MDA Guide V1.0.1. <http://www.omg.org/cgi-bin/doc?omg/03-06-01>. Letzter Abruf am 13. September 2009.

Object Management Group. (2002) Meta Object Facility (MOF) Specification Version 1.4. <http://www.omg.org/spec/MOF/1.4/PDF/>. Letzter Abruf am 19. Juli 2009.

Object Management Group. (2006) Meta Object Facility (MOF) Core Specification Version 2.0. <http://www.omg.org/docs/formal/06-01-01.pdf>. Letzter Abruf am 19. Juli 2009.

Object Management Group. (2009) Architecture-Driven Modernization Task Force. <http://omg.org/adm/>. Letzter Abruf am 19. Juli 2009.

Pedersen, T. B., Jensen, C. S. und Dyreson, C. E. (2001) A foundation for capturing and querying complex multidimensional data. *Information Systems*, 26(5), 383-423.

Peffer, K., Tuunanen, T., Rothenberger, M. und Chatterjee, S. (2008) A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77.

Sapia, C., Blaschka, M., Hofling, G. und Dinter, B. (1998) Extending the E/R Model for the Multidimensional Paradigm. In *Proceedings of the Workshops on Data Warehousing and Data Mining: Advances in Database Technologies*, Springer, 105-116.

Steinberg, D., Budinsky, F., Paternostro, M. und Merks, E. (2009) *EMF: Eclipse Modeling Framework*, 2. Aufl. Addison-Wesley



# **Erweiterungspotenziale für die Modellierung mit „H2 for Reporting“ am Beispiel von Forschungsportalen**

Ralf Knackstedt, Łukasz Lis

European Research Center for Information Systems (ERCIS)  
University of Muenster

## **1 Fachkonzeptionelle Modellierung Data-Warehouse-basierter OLAP-Systeme**

Für die Modellierung der fachlichen Anforderungen an Data-Warehouse-basierte OLAP-Systeme wurde eine Vielzahl unterschiedlicher Modellierungstechniken entwickelt (vgl. Böhnlein 2001): Klassische Ansätze lehnen sich eng an Modellierungssprachen an, die für die Fachkonzeption auf relationalen Datenbanken basierender, operativer Anwendungen etabliert sind, wie z. B. das Entity-Relationship-Modell (Chen 1976) oder die Objekttypenmethode (Wedekind 1981). Erweiterungen klassischer Datenmodellierungsansätze modifizieren die etablierten Datenmodellierungstechniken, wodurch insbesondere den rechentechnischen und sachlogischen Kennzahlenbeziehungen sowie der Modellierung von Bezugsobjekten auf Ausprägungsebene eine stärkere Beachtung geschenkt wird. Rein multidimensionale Modellierungsansätze verzichten auf eine Anlehnung an bestehende Datenmodellierungsansätze. Andere Modellierungsansätze entstammen dem Bereich der Scientific and Statistical Data Bases, der sich bereits seit den achtziger Jahren mit der Spezifikation multidimensionaler Datenräume befasst. Für die Spezifikation der bereitzustellenden Data-Warehouse-Daten werden darüber hinaus auch objektorientierte und erweiterte objektorientierte Modellierungsansätze vorgeschlagen. Einige dieser Ansätze zielen speziell auf eine Erweiterung der Unified Modeling Language (Booch et al. 1999) ab.

Angesichts der Vielzahl bereits bestehender Modellierungstechniken und vorhandener Standardisierungsbestrebungen erscheint es fraglich, ob die Weiterentwicklung einzelner Modellierungstechniken zukünftig eine gleich bleibend wichtige Aufgabenstellung für die Forschung im Bereich Business Intelligence darstellen wird. Anhand ausgewählter *Erweiterungspotenziale* plädiert dieser Beitrag dafür, die Entwicklung fachkonzeptioneller Modellierungstechniken für Data-Warehouse-basierte OLAP-Systeme nicht als abgeschlos-

sen anzusehen, sondern als ein interessantes Forschungsfeld, in dem sich weiterhin relevante Ergebnisse erzielen lassen.

Als Ausgangsbasis für die Weiterentwicklung wird die Modellierungstechnik H2 for Reporting gewählt, die im Folgenden kurz charakterisiert wird (vgl. Abschnitt 2). Die in diesem Beitrag vorgestellten Entwicklungspotenziale dieser Modellierungstechnik lassen sich auf andere Modellierungsansätze übertragen und kommen in unterschiedlichen Anwendungsbereichen Data-Warehouse-basierter OLAP-Systeme zum Tragen. Um die Entwicklungspotenziale trotz ihrer allgemeinen Anwendbarkeit konkretisieren zu können, wird ihre Bedeutung exemplarisch für die Fachkonzeption von Forschungsportalen erläutert. Die Auswahl dieses Anwendungsgebiets ist durch die Zielstellung des Workshops, in dessen Rahmen dieser Beitrag präsentiert wird, motiviert. Sie umfasst die Entwicklung einer „Forschungslandkarte für Business Intelligence“. Eine Forschungslandkarte kann als eine Ausprägung von Forschungsportalen angesehen werden. Im Anschluss an die Erörterung der Bedeutung von Business Intelligence für Forschungsportale (Abschnitt 3), werden konstruktionsorientierte Forschungsperspektiven aufgezeigt, die in einem engen bis weiteren Zusammenhang mit H2 for Reporting derzeit verfolgt werden bzw. werden sollen. Die Weiterentwicklungsperspektiven werden jeweils im Kontext des forschungsporalbasierten Controlling konkretisiert. Ein kurzes Fazit schließt den Beitrag ab (vgl. Abschnitt 5).

## **2 H2 for Reporting als Ausgangspunkt**

Die Modellierungssprache H2 for Reporting ist in ihrer aktuellen Fassung aus einem mehrjährigen Entwicklungsprozess hervorgegangen (vgl. Holten 2003). Mit den anderen für die Fachkonzeption Data-Warehouse-basierter OLAP-Systeme vorgeschlagenen Modellierungstechniken weist sie wesentliche Gemeinsamkeiten, aber auch Besonderheiten auf.

Als gemeinsamer konzeptioneller Kern der Modellierungstechniken lassen sich Dimensionen und Kennzahlensysteme identifizieren, die zu Navigationsräumen zusammengefasst werden. Die Navigationsräume beschreiben die Datenbestände, die das OLAP-System auswertbar machen sollen. Die einzelnen Modellierungstechniken verwenden teilweise unterschiedliche Bezeichnungen für gleiche oder ähnliche Konstrukte. Darüber hinaus unterscheiden sie sich in weiteren Details, die darauf zurückzuführen sind, welche Zielsetzungen bei der Konstruktion der Modellierungstechniken schwerpunktmäßig verfolgt wurden.

H2 for Reporting zeichnet sich im Vergleich zu anderen Modellierungssprachen insbesondere dadurch aus, dass mit ihr neben den auszuwertenden Navigationsräumen auch die

Tabellenstrukturen beim Aufbau der OLAP-Berichte spezifiziert werden können. Darüber hinaus steht mit dem H2-Toolset ein Metamodellierungswerkzeug zur Verfügung, mit dem sich H2 for Reporting-Modelle erstellen und verwalten lassen.

Die inhaltliche Spezifikation der OLAP-Berichte erfolgt, indem über die Angabe von Zeilen- und Spalteninhalten eine zweidimensionale Projektion auf einen Navigationsraum festgelegt wird (vgl. Abbildung 1). Die Zeilen- und Spaltenstruktur legt den tabellari-schen Aufbau des Berichts fest. Navigationsräume werden von Dimension(sausschnitt)en (z. B. Tage nach Jahren, Institutionen nach Ländern und Forschungsthemen nach Themengruppen) und Kennzahlensystemen (z. B. Anzahl der Forschungsergebnisse) aufgespannt. Der Navigationsraum macht auf diese Weise deutlich, welche Kennzahlen nach welchen inhaltlichen Kriterien analysierbar sein sollen.

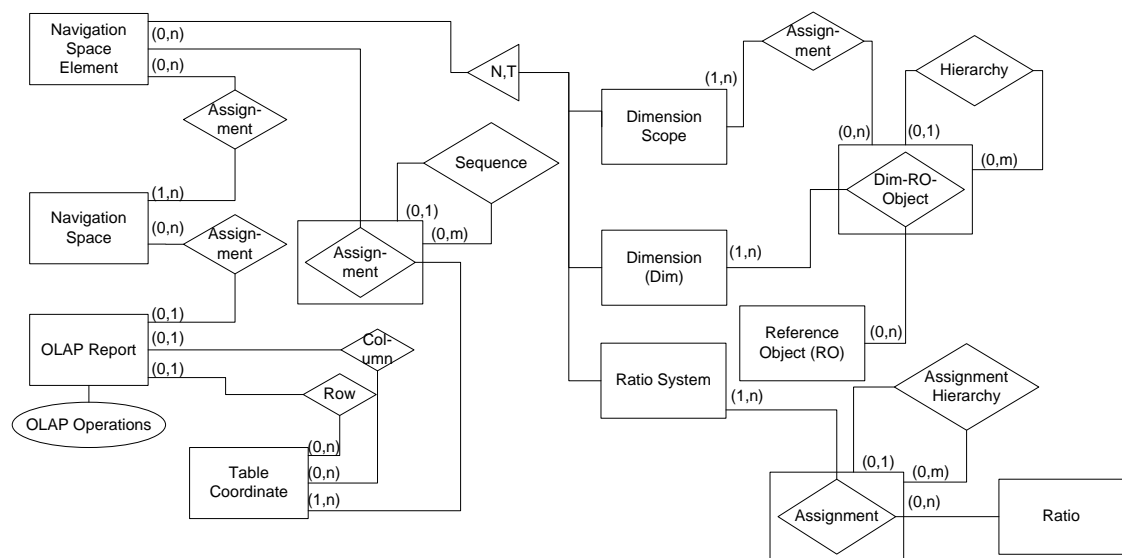


Abbildung 1: Metamodell von H2 for Reporting (Ausschnitt)

### 3 Fachkonzeption von Forschungsportalen als Gegenstand der BI-Forschung

Unter (internetbasierten) *Forschungsportalen* werden themenspezifische Websites verstanden, die Informationen, die an anderer Stelle in ausführlicher Form bereits vorhanden sind, spezifisch strukturiert zur Verfügung stellen, mit dem hauptsächlichen Ziel, die Zusammenarbeit in Forschungsprojekten zu fördern oder die Bekanntheit geplanter bzw. erzielter Forschungsergebnisse zu steigern. Um dies zu erreichen, dokumentiert die Website, wer Forschung betreibt und/oder welche Ziele in Forschungsprojekten verfolgt und/oder welche Ergebnisse in Forschungsprojekten erzielt wurden bzw. werden sollen. Forschung umfasst dabei sowohl akademische Forschung an Hochschulen und sonstigen

staatlichen Forschungseinrichtungen als auch Forschung in Unternehmen, öffentlichen Verwaltungen etc. Um die beschriebenen Ziele zu erreichen, stellt eine geographische Darstellung der Forschungsstandorte häufig eine wünschenswerte aber nicht immer notwendige Funktionalität von Forschungsportalen dar. Im Zusammenhang mit der geographischen Darstellung der Forschungsstandorte wird auch der Begriff Forschungslandkarte verwendet. Teilweise wird der Begriff der Forschungs(land)karte auch für die Visualisierung von Themengebieten verwendet, welche von dem Forschungsportal adressiert werden. Die thematische Eingrenzung spezialisiert den Zweck der Website. Beispiele sind die Förderung der Krebsforschung (<http://www.cancermap.org>) oder die Erforschung des Einsatzes von Windenergie (<http://forschungslankarte-windenergie.de>). Dabei sind auch geografische oder organisatorische Eingrenzungen des Betrachtungsgegenstands des Forschungsportals möglich. Ein Beispiel stellt die Beschränkung der Betrachtung der Stammzellenforschung auf diejenige im Bundesland Nordrhein-Westfalen dar (vgl. z. B. <http://stammzellen.nrw.de>). In Abhängigkeit von der themenspezifischen Zielstellung strukturiert die Website die zur Verfügung gestellten Informationen. Diese Struktur ist charakteristisch für die Website. Der Detaillierungsgrad und Facettenreichtum der Strukturierung ist dabei allerdings vom Qualitätsanspruch der Website und ihres Informationsumfangs abhängig.

Forschungsportale stellen eine Datenbasis zur Verfügung, deren Auswertung mit Business Intelligence (BI)-Systemen einen wichtigen Beitrag zum *Forschungs-Controlling* leisten kann. Ein forschungsportalbasiertes Controlling liefert dabei unterschiedlichen Anspruchsgruppen wichtige Informationen:

- *Einzelne Forscher* können z. B. analysieren, wer sich mit ähnlichen oder ergänzenden Themen beschäftigt, um Kooperationspartner für zukünftige Projekte zu identifizieren.
- *Forschungsabteilungen oder -institute* können mit entsprechenden Auswertungen z. B. dokumentieren, welchen Themen sie sich im Zeitverlauf mit welcher Intensität gewidmet haben und mit welchen anderen Forschern sie in welchem Umfang kooperiert haben.
- *Hochschulen und außeruniversitäre Forschungseinrichtungen* können z. B. die Forschungsleistungen ihrer Mitglieder in verschiedenen Themenbereichen untersuchen.
- *Staatliche und private Fördereinrichtungen* können z. B. untersuchen, zu welchen Forschungsbereichen bereits viele Ergebnisse vorliegen, um so gegebenen-

falls vorhandene Lücken zu identifizieren, für welche die Intensivierung der Förderung angezeigt wäre.

Zu den Dimensionen, nach denen im Rahmen des forschungsportalbasierten Controllings beispielsweise ausgewertet werden kann, zählen: Forschungsergebnistyp, Realisationsgrad, Praxiseinsatz, Funktionsbereich, Organisationsklasse, Anwendungsbranchen und Standardisierung. Als Kennzahl wird häufig die Anzahl der im Portal beschriebenen Forschungsergebnisse ermittelt. Im Zuge einer leistungsabhängigen Mittelvergabe an Forschungseinrichtungen hat sich eine Vielzahl weiterer Kennzahlen entwickelt, deren Ermittlung von Forschungsportalen ebenfalls unterstützt werden kann. Viele der Bewertungsansätze ermitteln dabei nicht allein die Zahl der Publikationen, sondern bewerten diese entsprechend des Stellenwerts des jeweiligen Publikationsorgans. Beispielsweise wird von Garfield (1955) der Impact Factor (IF) propagiert, der im Rahmen des (Social) Science Citation Index bis heute aktiv berechnet wird. Dabei wird gemessen, wie oft Artikel einer Fachzeitschrift in anderen Fachzeitschriften zitiert werden und diese Zahl in Verhältnis zur Gesamtheit zitierbarer Artikel gesetzt. Perspektiv ist davon auszugehen, dass an Hochschulen und in Forschungsfördereinrichtungen über Forschungsportale Datenbanken zunehmend größeren Datenvolumens aufgebaut werden, in denen die Forschungsleistungen (Publikationen, Patente, Preise, Projekte) dokumentiert werden. Die Auswertung dieser Datenbasis zu unterschiedlichen Zwecken (strategische Ausrichtung der Forschungsbemühungen, Identifikation von Kooperationspartnern und Synergiepotenzialen, Mittelvergabe an Institute, Unterstützung von Gehaltsverhandlungen etc.) anhand jeweils verschiedener Dimensionen und Kennzahlen motivieren, die Datenbasis eines Forschungsportals in Anlehnung an die Data Warehouse-Architektur aufzubauen (vgl. Becker et al. 2008).

Mit yourResearchPortal.com wird ein softwaretechnisches System entwickelt, mit Hilfe dessen Forschungsportale automatisiert generiert, administriert und betrieben werden können. Bei der Ausgestaltung des Systems wurden zwei Hauptkomponenten unterscheiden. Die Datenpflegekomponente umfasst Funktionalitäten, die für das Erstellen, Betrachten, Aktualisieren und Löschen von konkreten Daten gebraucht werden. Typisch hierfür sind ein hoher Detaillierungsgrad der Daten und die Betrachtung von einzelnen Inhalten. Die Analysekomponente andererseits ermöglicht einen nur lesenden Zugriff auf die gesamte Wissensbasis eines Portals. Als Datenpflegekomponente fungiert das Open Source Content Management System Drupal (<http://www.drupal.org>), welches modular aufgebaut ist. Zusätzlich zum für die grundlegende Funktionalität zuständigen sogenannten Core können weitere Funktionalitäten durch die Installation verschiedener Module hinzugefügt werden. Die Analysekomponente wird in yourResearchPortal.com mit Hilfe des Open Source OLAP-Server Mondrian realisiert. Dieser ermöglicht sowohl die kom-

fortable Durchführung quantitativer Analysen als auch ein effizientes und interaktives Navigieren durch die mehrdimensionale Wissensbasis bis hin zur Ebene der einzelnen Inhalte. Die beiden Hauptkomponenten sind in einer webbasierten Benutzeroberfläche vollständig integriert. Sowohl die Datenpflege- als auch die Analysekomponente greifen auf die gleiche relationale MySQL-Datenbank (<http://www.mysql.com>) zu.

Obwohl die beiden Komponenten eine unterschiedliche Ausrichtung aufweisen (OLTP bei Drupal und OLAP bei Mondrian) wurde bei der DV-technischen Umsetzung darauf geachtet, ein relationales Schema zu entwickeln, das beiden Aufgaben gerecht wird. Wichtige Herausforderung war es hierbei, das System so zu gestalten, dass Änderungen in Drupal (z. B. das Hinzufügen einer neuen Dimension oder eines neuen Wertes) von Mondrian direkt übernommen werden, ohne dass eine zusätzliche Konfiguration der Analysekomponente nötig ist. Dies wäre im klassischen Star-Schema der Fall, wo jede Dimension durch eine eigene Tabelle repräsentiert wird (Kemper et al. 2006, S. 61-64). In Abbildung 2 wird ein Schemaausschnitt präsentiert, der den in yourResearchPortal.com verfolgten Ansatz schildert. Durch Rechtecke werden Tabellen samt ihren Namen und Spalten (Primärschlüssel unterstrichen in Fettschrift) dargestellt. Pfeile zeigen auf die von Fremdschlüsseln referenzierten Spalten. Neben der Faktentabelle Einordnung, in der die Zuweisung von Artefakten (Forschungsergebnissen) zu Dimensionsobjekten in Anhängigkeit vom Portal festgehalten wird, existieren drei Dimensionstabellen Portal, Dimensionsobjekt und Artefakt.

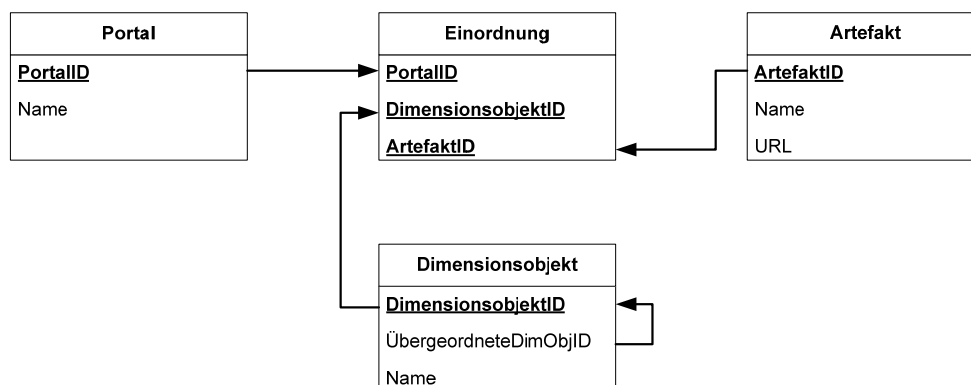


Abbildung 2: Generisches multidimensionales Analyseschema

Dimensionen werden durch Wurzel-Dimensionsobjekte abgebildet, die kein übergeordnetes Dimensionsobjekt besitzen. Diese werden wiederum von den Werten der ersten Hierarchieebene referenziert usw. Der OLAP-Server Mondrian ist so konfiguriert, dass den Benutzern eines Portals die entsprechenden Dimensionen samt ihren Werten und eventuellen untergeordneten Organisationen oder Projekten für Analysezwecke zur Verfügung stehen. Im Hintergrund wird ein Filter aktiviert, welcher nur Daten des aktuellen Portals

darstellen lässt. Hierfür werden die Ausprägungen der Analysedimension Portal auf die entsprechende ID eingeschränkt.

Infrastrukturen wie yourResearchPortal.com ermöglichen es, dass sich Betreiber von Forschungsportalen bei deren Realisierung auf die fachkonzeptionelle Ausgestaltung des Portals konzentrieren können. Die Spezifikation der Anforderungen an die Controllingauswertungen und damit die Festlegung der auch in der Datenpflegekomponente verwendeten Dimensionen steht dabei im Vordergrund. Die Diskussion und Dokumentation der gewünschten Ausgestaltung des Forschungsportals stellt daher ein Anwendungsgebiet für die fachkonzeptionelle Modellierung Data-Warehouse-basierter OLAP-Systeme dar.

#### **4 Entwicklungsperspektiven für H2 for Reporting am Beispiel des forschungportalbasierten Controlling**

Im Folgenden wird ein Überblick über weiterführende Entwicklungen gegeben, die für die Modellierungssprache H2 for Reporting untersucht wurden bzw. werden. Die vorgestellten Entwicklungspotenziale adressieren dabei jeweils konstruktionsorientierte Forschung. Im Sinne von Design Science beinhaltet diese die Entwicklung von Artefakten, die einen Problemlösungsbeitrag leisten, und die Evaluation dieses Problemlösungsbeitrags (vgl. z. B. Hevner et al. 2004). Zu den im Rahmen der Wirtschaftsinformatik besonders relevanten Artefakttypen zählen Sprachkonstrukte, Methoden, Modelle und Implementierungen. Die konstruktionsorientierten Forschungsfragen werden zunächst allgemein eingeführt und anschließend konkret auf das forschungportalbasierte Controlling angewandt (vgl. Tabelle 1).

<b>Entwicklungsperspektive</b>	<b>Fragestellungen</b>	<b>Anwendungsbeispiele im forschungportalbasierten Controlling</b>
Kollaborative Entwicklung von H2 for Reporting-Modellen	Welche Mechanismen unterstützen die Konsensbildung und wie lassen sich diese softwaretechnisch umsetzen?	Kollaborative Diskussion von Dimensionen, die zur Beschreibung von Forschungsergebnissen vom Forschungsportal genutzt werden

Entwicklungsperspektive	Fragestellungen	Anwendungsbeispiele im forschungsportal-basierten Controlling
Integration von H2 for Reporting-Modellen mit weiteren Modelltypen	Wie lassen sich die Metamodelle unterschiedlicher Modellierungstechniken integrieren und wie kann die gestiegene Komplexität z. B. durch Sichtenkonzepte gehandhabt werden?	Integration von H2 for Reporting mit einer Technik zur Spezifikation von Websites
Referenzmodellierung zur Wiederverwendung von H2 for Reporting-Modell(teil)en	Mit welchen Mechanismen lässt sich die Wiederverwendung unterstützen und wie lassen sich diese softwaretechnisch umsetzen? Welche Berichtsspezifikationen eignen sich in bestimmten Anwendungsdomänen für eine Wiederverwendung?	Varianten von Fachkonzepten für Forschungsportalen inklusive ihrer Auswertungsfunktionalität und -inhalte spezifizieren und generieren
Optimierung einer Menge Berichtsspezifikationen z. B. durch Vergleich der Berichtsspezifikationen auf Ähnlichkeit	Wie lässt sich der Vergleich von Modell(teil)en auf Ähnlichkeit automatisieren z. B. durch die Adaption der Clusteranalyse?	Vergleich von Berichtsdefinitionen im Forschungs-Controlling zur Identifikation von ähnlichen Berichtsanforderungen
Automatisierung der Implementierung von Berichten auf der Basis H2 for Reporting-Modell(bestandteil)en	Welche Algorithmen sind geeignet, um OLAP-Systeme auf der Basis der fachkonzeptionellen Modelle automatisiert einzurichten?	Generierung von Auswertungen für Forschungsportale oder weiterer Bestandteile auf der Basis fachkonzeptioneller Modelle



Konstruktion von Reifegradmodellen zur Beurteilung von BI-Systemen bzw. BI-Komponenten enthaltenen Systemen	Welche Inhalte von H2 for Reporting (Referenz)modellen bzw. welche methodischen Aspekte der Modellierung mit H2 for Reporting eignen sich zur Beschreibung von Reifegraden?	Entwicklung eines Reifegradmodells für Forschungsportale
---	---	--

*Tabelle 1: Erweiterungspotenziale von H2 for Reporting*

*Kollaborative Entwicklung:* Fachkonzepte werden in aller Regel nicht durch eine einzelne Person, sondern unter Beteiligung mehrerer Anspruchsgruppen entwickelt, die möglichst einen Konsens herbeiführen. Eine erste Forschungsfrage, die über die reine Definition der OLAP-Spezifikationssprache hinausführt, aber zugleich auch eng mit ihr verbunden ist, stellt daher die Entwicklung einer Diskussionsumgebung dar, mit der sich fachkonzeptionelle Modelle kollaborativ erarbeiten lassen. Im Kontext der Prozessmodellierung sind derartige Ansätze schon relativ weit verbreitet. Die „Open Model Initiative“ (vgl. Frank, Strecker 2007) adressiert die gemeinschaftliche Erstellung beliebiger Modelltypen. Für den OLAP-Kontext ist zu untersuchen, welche Mechanismen sich zur Koordination der Diskussionen übernehmen und auf welche Modellbestandteile sich diese erfolgreich anwenden lassen. Für H2 for Reporting wurde eine Erweiterung des H2-Toolsets entwickelt, mit der sich Diskussionsprozesse unterstützen lassen, die in verschiedenen Anwendungskontexten evaluiert wurde. Im Rahmen des forschungsportalbasierten Controlling wurde diese Erweiterung genutzt, um eine funktionale Dimension zu entwickeln, der später Forschungsergebnisse zugeordnet werden konnten. Die diskutierte Dimension wurde im Anschluss in einen funktionalen Ordnungsrahmen umgesetzt, der in der DIN PAS 1094 dokumentiert wird.

*Integration mit weiteren Modelltypen:* OLAP-Berichte werden in bestimmten Kontexten verwendet. Beispielsweise lässt sich beschreiben, für welche betriebswirtschaftlichen Aufgaben OLAP-Berichte eingesetzt werden sollen und welche Organisationseinheiten, die Berechtigung zum Aufruf eines bestimmten Berichts erhalten sollen. Von besonderer Bedeutung für die Verbreitung fachkonzeptioneller Modelle erscheint daher die Integration der OLAP-Modelle mit anderen verbreiteten Modelltypen (vgl. Knackstedt 2006). Durch die Annotation von OLAP-Berichten an Funktionen in Prozessmodellen kann z. B. ausgedrückt werden, in welchen Arbeitskontexten OLAP-Berichte verwendet werden sollen. Insbesondere bei wiederkehrenden Aufgaben, wie z. B. der Sortimentsgestaltung im Handel, bietet sich eine Dokumentation in Prozessmodellen an. Die Integrationsauf-

gabe umfasst unter anderem die Konstruktion aufeinander abgestimmter Metamodelle mit geeigneten Sichtenkonzepten, welche die steigende Komplexität der integrierten Modelle handhabbar machen. Im Kontext des forschungsportalbasierten Controlling wurde die Integration der H2 for Reporting-Sprachkonstrukte in die Internetspezifikationsprache eW3DT nach Scharl (vgl. Scharl 1997) untersucht (vgl. Knackstedt et al. 2009). Dabei wurden OLAP-Berichtsdefinitionen als Inhalte einzelner Web-Seiten in die Spezifikationsprache integriert. Außerdem wurde erkannt, dass auch einzelne Dimensionen, Kennzahlensysteme und Navigationsräume zur Spezifikation von Bestandteilen spezifischer Web-Seitentypen, wie z. B. Index und Suche, herangezogen werden können (vgl. Abbildung 3). Die integrierte Modellierungssprache eignet sich dazu, Forschungsportale fachkonzeptionell zu spezifizieren und dabei sowohl die Seitenstrukturen der Forschungsportale abzubilden, als auch die OLAP-Berichte zur Auswertung der durch die Forschungsportale erfassten Daten zu spezifizieren.

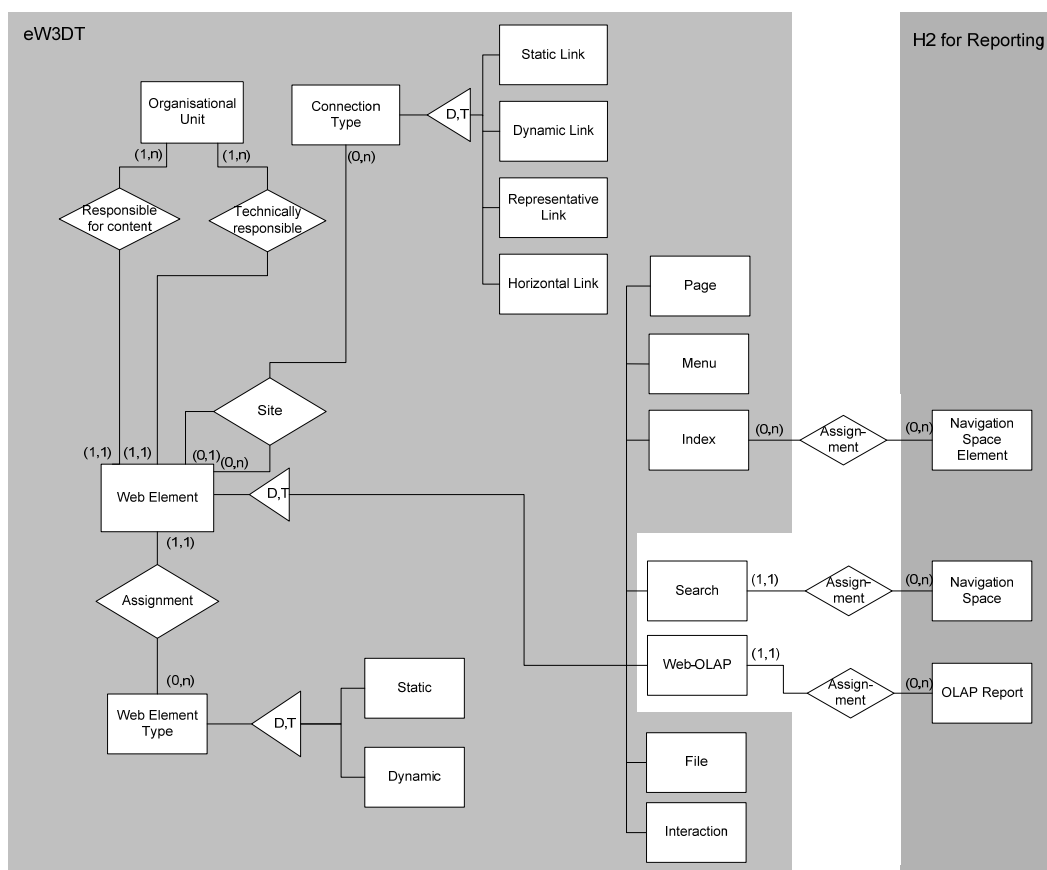


Abbildung 3: Integration von H2 for Reporting mit eW3DT (Ausschnitt)

**Referenzmodellierung:** Modellierungswerkzeuge profitieren in besonderem Maße davon, wenn sie nicht allein mit vordefinierten bzw. adaptierbaren Modellierungssprachen zur Verfügung gestellt werden, sondern auch mit anpassbaren, wiederverwendbaren Model-

len. Dieser Umstand und weitere Vorteile der Referenzmodellierung haben motiviert, zu untersuchen, welche Spracherweiterungen notwendig sind, um die Wiederverwendung von OLAP-Fachkonzepten zu unterstützen. Grundsätzlich lassen sich zur Unterstützung der Wiederverwendung Mechanismen unterscheiden, die auf Aggregation, Spezialisierung, Analogiebildung, Instanziierung und Konfiguration beruhen (vgl. Becker, Knackstedt 2004). Für H2 for Reporting wurde eine konfigurierbare Sprachversion entwickelt, die es ermöglicht, Regeln zu formulieren, die festlegen, wie ein fachkonzeptionelles Modell an bestimmte Anwendungskontexte anzupassen ist (vgl. Knackstedt, Seidel, Janiesch 2006). Im Kontext des forschungsportalbasierten Controlling wird eine Referenzmodell für Forschungsportale entwickelt. Hierzu wurden bisher im Internet 66 Forschungsportale identifiziert und hinsichtlich ihrer Inhalte analysiert. Die im Sinne eines Common Sense üblichen Inhalte wurden mit einer Kombination aus H2 for Reporting und eW3DT modelliert (vgl. Abbildung 4). In weiteren Forschungsarbeiten soll die empirische Basis zur Entwicklung des Referenzmodells erweitert werden. Hierbei wird untersucht, welche typischen Varianten von Forschungsportalen existieren. Es wird z. B. vermutet, dass sich die Forschungsportale einzelner Forscher von denen von Forschungsnetzwerken durch spezifische Inhalte bzw. Funktionen unterscheiden. Des Weiteren sind Unterschiede zwischen Forschungsportalen, die von der Wirtschaft bzw. der Wissenschaft betrieben werden, wahrscheinlich. Die Arbeiten zur Spracherweiterung ermöglichen die adäquate Dokumentation der Varianten in einem konsistenten Gesamtmodell. Mittels Auswertung der Konfigurationsregeln lassen sich dann kontextspezifische Varianten aus dem Gesamtmodell ableiten.

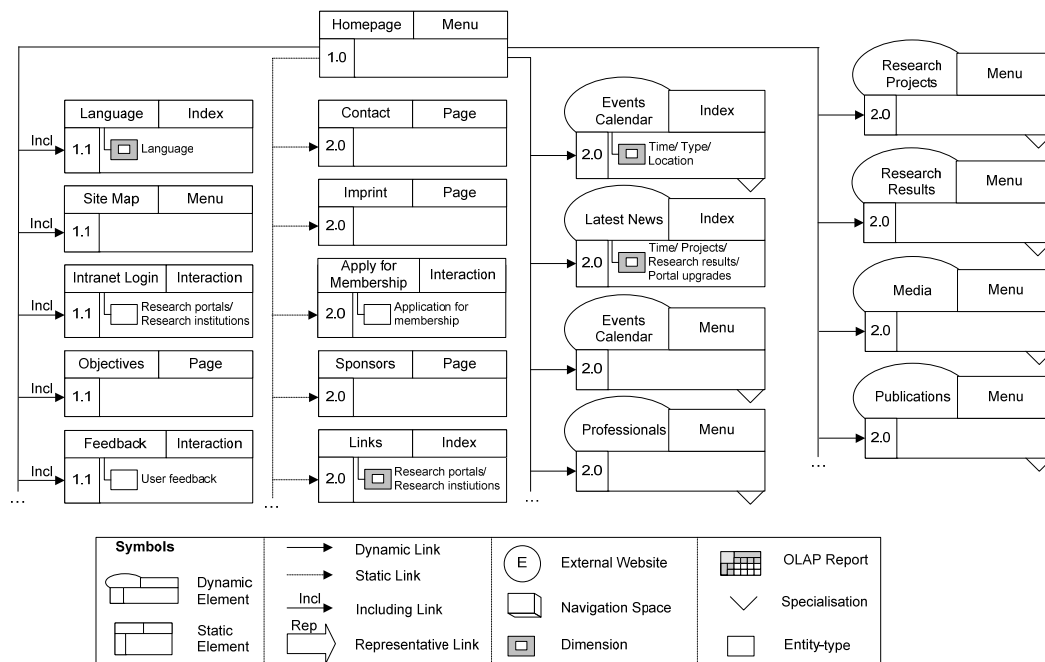


Abbildung 4: Forschungsportal-Referenzmodell (Ausschnitt)

*Optimierung einer Menge OLAP-Berichtsspezifikationen:* Der Abgleich eines bestehenden Berichtssystems mit einem Referenzmodell stellt einen möglichen methodischen Ansatz zu dessen inhaltlichen Verbesserung dar. Die Orientierung an den Begriffen eines Referenzmodells kann z. B. zu einer Vereinheitlichung des Sprachgebrauchs im Berichtswesen eines Unternehmens beitragen. Die Präzisierung und Vereinheitlichung der Kennzahldefinitionen, die bessere Abstimmung von Informationsangebot und Informationsnachfrage, die Verringerung der Berichtszahl sind weitere wesentliche Ansatzpunkte für die Verbesserung eines bestehenden Berichtswesens. Zur Ermittlung des Informationsbedarfs tragen auch die zahlreichen Ansätze zur Entwicklung von Soll-Modellen bei, die sich z. B. in datenangebots-, aufgaben- bzw. zielanalytische und datennachfrageorientierte Ansätze unterteilen lassen. Die Entwicklung eines konsistenten Sollkonzepts und die Konsolidierung eines historisch gewachsenen Berichtswesens lassen sich darüber hinaus durch statistische Ansätze unterstützen, die Verbesserungs- und Vereinfachungspotenziale in den OLAP-Berichten aufzeigen bzw. identifizieren helfen (vgl. Delfmann et al. 2008). Im Rahmen der konstruktionsorientierten Forschung ist zu untersuchen, welche statistischen Methoden sich auf dieses Problem anwenden lassen und welche Anpassungen gegebenenfalls an den Modellierungssprachen bzw. -editoren vorzunehmen sind. Für H2 for Reporting und das H2 Toolset wurde die Clusteranalyse adaptiert, um im Rahmen einer Konsolidierung umfangreicher Modellsysteme automatisiert Hinweise auf ähnliche Modelle generieren zu können (vgl. Knackstedt, Deinert, Becker 2009). Für die ähnlichen Modelle soll dann geprüft werden, ob sich die Berichtsspezifikationen durch geeignet

parametrisierbare Berichte ersetzen lassen, so dass die Komplexität des Berichtsystems reduziert werden kann. Im Rahmen der Entwicklung eines Referenzmodells für das forschungsportalbasierte Controlling kann dieses Verfahren genutzt werden, um Ähnlichkeiten in den Berichten unterschiedlicher Forschungsportale zu identifizieren. Hierbei werden die in Forschungsportalen verwendeten Bereichsspezifikationen miteinander verglichen. Eine geeignete Datengrundlage wird sich in diesem Bereich derzeit vor allem im Forschungscontrolling von Hochschulen finden lassen. Neben den Berichtsspezifikationen lassen sich aber auch die Modelle der Forschungsportale selbst einer Clusteranalyse unterziehen. Auf diese Weise könnten Gruppen von Forschungsportalen gleicher bzw. ähnlicher Funktionalität identifiziert werden, die im Rahmen der Referenzmodellierung als Varianten ausgewiesen werden können.

*Automatisierung der Implementierung:* Ein wesentlicher Beitrag zur Reduktion des Aufwandes der Einführung von Data-Warehouse-gestützten OLAP-Systemen der Fachkonzeption stellt die automatische Überführung der fachkonzeptionellen Modelle in techniknähere (Teil-)spezifikationen bzw. Implementierungen dar, wie sie z. B. von MDA (Model Driven Architecture) adressiert wird. Im Kontext von OLAP-Modellierungsmethoden wurde diese Zielsetzung bereits frühzeitig aufgegriffen. Auch im Rahmen der Entwicklung von H2 for Reporting wurden Algorithmen zur Automatisierung der Implementierung entwickelt (vgl. Holten 2003). YourResearchPortal.com ermöglicht die Generierung von Forschungsportalen. Das System erfragt hierzu wesentliche Auswertungsdimensionen vom zukünftigen Administrator des zu generierenden Forschungsportals. In Übereinstimmung mit den Strukturen der H2 for Reporting-Sprache kann der Nutzer über das Internet interaktiv die Auswertungsdimensionen anlegen. Im Anschluss wird ein spezifisches Forschungsportal mit angepassten OLAP-Berichten generiert. Auch die übrige Funktionalität des generierten Forschungsportals lässt sich über die Administrationskomponente anpassen. Ein erstes Forschungsportal, das mit diesem Werkzeug generiert wurde, stellt die „Forschungslandkarte zur hybriden Wertschöpfung“ (<http://www.forschungslandkarte-hybridewertschoepfung.de>) dar (vgl. Abbildung 5). Die Dimensionsspezifikationen der zu generierenden Forschungsportale ließen sich prinzipiell auch aus H2 for Reporting-Modellen auslesen, was derzeit allerdings noch nicht unterstützt wird. Mit den Arbeiten zur Referenzmodellierung von Forschungsportalen sind weitere wichtige Grundlagen zur Realisierung einer modellgetriebenen Generierung von Forschungsportalen gelegt. Das Modell, das der Generierung zu Grunde liegt, könnte dabei zunächst durch die Adaption eines Forschungsportal-Referenzmodells erzeugt werden.

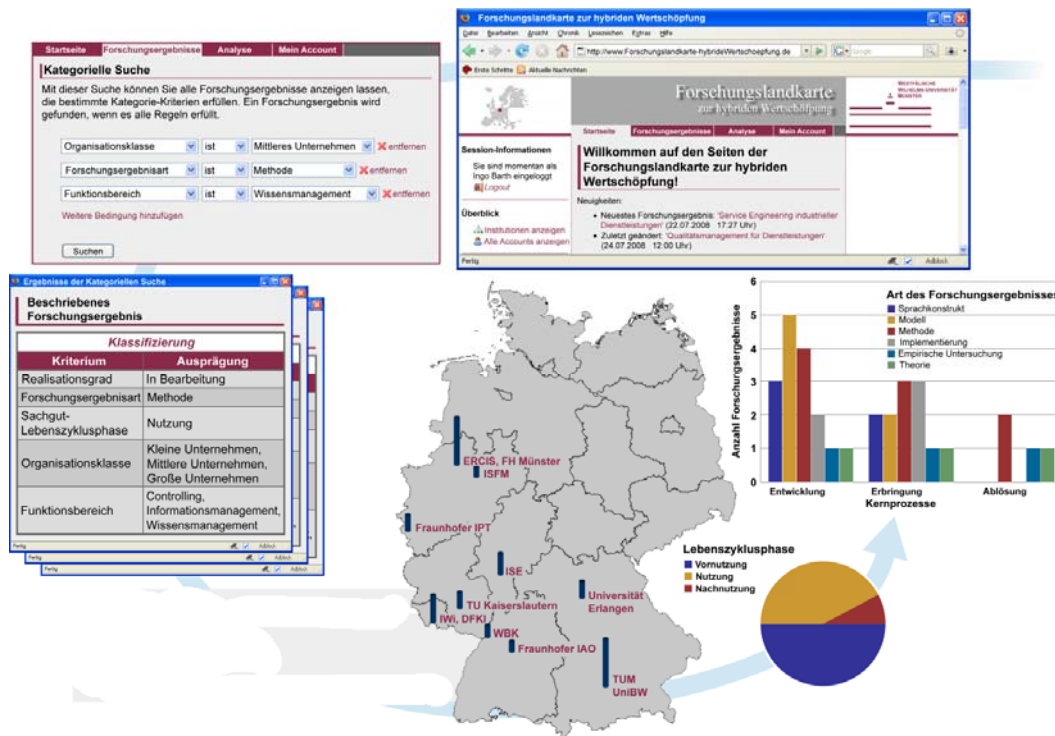


Abbildung 5: Mittels yourResearchPortal.com generierte „Forschungslandkarte zur hybriden Wertschöpfung“

**Konstruktion von Reifegradmodellen:** Zur Beurteilung des Entwicklungsstands des BI-Systems eines Unternehmens werden verschiedene Reifegradmodelle vorgeschlagen (vgl. Chamoni, Gluchowski 2004). In der Regel wird in Reifegradmodellen die Beurteilung nach unterschiedlichen Sichten vorgenommen. Die Verwendung von Modellierungstechniken zur Fachkonzeption des BI-Systems geht in die organisatorische Beurteilung der Prozesse der Entwicklung und des Betriebs eines BI-Systems ein. Die technische Dimension beurteilt die von den Auswertungssystemen zur Verfügung gestellte Funktionalität und die Infrastruktur zur Vereinheitlichung und Verteilung der Daten. Die inhaltliche Beurteilungsdimension untersucht, ob die für das Unternehmen relevanten Informationen bereitgestellt werden. Neben der allgemeinen Beurteilung des BI-Systems einer Unternehmung als ganzes, können Reifegradmodelle auch spezifische Controllingbereiche, wie z. B. das IT-Controlling, adressieren (vgl. Becker, Knackstedt, Pöppelbuß 2009). Je spezifischer die Controlling-Bereiche gewählt werden, für die ein Reifegradmodell entwickelt wird, desto konkreter können die bereitgestellten Inhalte von BI-Systemen beurteilt werden. Für das forschungsportalbasierte Controlling wird ein Reifegradmodell entwickelt, das neben forschungsportalspezifischer Funktionalität insbesondere auch die Abdeckung relevanter Inhaltsbereiche, wie z. B. Forschungsprojekte, -organisationen, -ziele, -ergebnisse und -publikationen untersucht. Für höhere Reifegrade ab Stufe 3

sieht das Reifegradmodell vor, dass das Forschungsportal die Analyse der Datenbasis mittels OLAP-Funktionalität unterstützt. Dementsprechend ergibt sich eine Überschneidung der Beurteilungskriterien des Reifegradmodells für Forschungsportale mit den BI-spezifischen Reifegradmodellen.

## 5 Fazit

Angesichts der Vielzahl an Modellierungstechniken, die für die fachkonzeptionelle Spezifikation Data-Warehouse-basierter OLAP-Systeme bereits vorgeschlagen worden sind, erscheint es nahe liegend, daran zu zweifeln, dass die (Weiter)entwicklung dieser Modellierungstechniken noch immer eine viel versprechende Forschungsperspektive darstellt. Die fachkonzeptionelle Modellierung ist eine aufwändige Aufgabe. In der Praxis wird diese häufig vernachlässigt, um vermeintlich Aufwand und Zeit zu sparen. Die Einsicht in den Nutzen der fachkonzeptionellen Modellierung steigt, wenn die Modelle in mehreren Kontexten verwendet werden. Am Beispiel des forschungsportalbasierten Controlling wurden sechs ausgewählte Themenbereiche vorgestellt, die zeigen, dass die fachkonzeptionelle Modellierung durchaus weiterhin interessante Perspektiven für konstruktionsorientierte Forschung bietet. Die diskutierten Themenbereiche erheben dabei keinen Anspruch auf Vollständigkeit, sondern repräsentieren priorisierte Arbeitsbereiche, die im Rahmen der Fortsetzung der Entwicklung der Modellierungssprache H2 for Reporting gewählt wurden.

## Danksagung

Dieser Beitrag wurde ermöglicht durch die Förderung des Projektes "Flexible Informationssystemarchitekturen für hybride Wertschöpfungsnetzwerke" (FlexNet) durch das BMBF (Förderkennzeichen 01FD0629) im Rahmen des Programms "Innovationen mit Dienstleistungen". Wir danken an dieser Stelle auch dem Projektträger Deutsches Zentrum für Luft- und Raumfahrt (DLR) für die Betreuung.

## 6 Literatur

Becker, J.; Beverungen, D.; Knackstedt, R.; Glauner, C.; Stypmann, M.; Rosenkranz, C.; Schmitt, R.; Hatfield, S.; Schmitz, G.; Eberhardt, S.; Dietz, M.; Thomas, O.; Walter, P.; Lönngren, H.; Leimeister, J.M. (2008): Ein Plädoyer für die Entwicklung eines multidimensionalen Ordnungsrahmens zur hybriden Wertschöpfung. In: Arbeitsberichte des Instituts für Wirtschaftsinformatik der Universität Münster, Nr. 119. Editors: Becker, J.; Grob, H.L.; Klein, S.; Kuchen, H.; Müller-Funk, U.; Vossen, G. 2008.

- Becker, J.; Knackstedt, R. (2004): Referenzmodellierung im Data-Warehousing – State-of-the-Art und konfigurative Ansätze für die Fachkonzeption. In: *Wirtschaftsinformatik*, 46 (2004) 1, S. 39-49.
- Becker, J.; Knackstedt, R.; Pöppelbuß, J. (2009): Entwicklung von Reifegradmodellen für das IT-Management – Vorgehensmodell und praktische Anwendung. In: *Wirtschaftsinformatik*, 51 (2009) 3, S. 249-260.
- Böhnlein, M. (2001): Konstruktion semantischer Data-Warehouse-Schemata, 1. Auflage, Deutscher Universitäts-Verlag, Wiesbaden, 2001.
- Booch, G.; Rumbaugh, J.; Jacobson, I. (1999): UML – The Unified Modeling Language Reference Manual. Reading 1999.
- Chamoni, P.; Gluchowski, P. (2004): Integrationstrends bei Business-Intelligence-Systemen. *Wirtschaftsinformatik*, 46 (2004) 2, S. 119-128.
- Chen, P.P.-S. (1976): The Entity-Relationship-Model – Toward a Unified View of Data. *ACM Transactions on Database Systems*, 1 (1976) 1, S. 9-36.
- Delfmann, P.; Knackstedt, R.; Niehaves, B.; Winkelmann, A. (2008): Berichtswesenvereinfachung: Perspektiven für eine gestaltungsorientierte Wirtschaftsinformatikforschung. In: vom Brocke, J.; Becker, J. (Hrsg.): *Einfachheit in Wirtschaftsinformatik und Controlling*. Festschrift für Heinz Lothar Grob. München 2008. S. 189-206.
- Frank, U., Strecker, S. (2007). Open Reference Models – Community-driven Collaboration to Promote Development and Dissemination of Reference Models. *Enterprise Modelling and Information Systems Architectures* 2 (2), pp. 32-41.
- Garfield, E. (1955): Citation indexes for science: a new dimension in documentation through association of ideas. *Science* Vol. 122, S. 108–111.
- Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design science in information systems research. In: *MIS Quarterly*, 28 (2004) 1, S. 75-105.
- Holten, R. (2003): Specification of management views in information warehouse projects. *Information Systems*, 28 (2003) 7, S. 709-751.
- Kemper H-G, Mehanna W, Unger C (2006) *Business Intelligence – Grundlagen und Praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung*. 2. Aufl. Wiesbaden.
- Knackstedt, R. (2006): Fachkonzeptionelle Referenzmodellierung einer Managementunterstützung mit quantitativen und qualitativen Daten. *Methodische Konzepte zur Konstruktion und Anwendung*. Berlin 2006.
- Knackstedt, R.; Deinert, M.-O.; Becker, J. (2009): Modellvergleich mittels Clusteranalyse am Beispiel einer automatisierten Ähnlichkeitsanalyse für OLAP-Berichtsspezifikationen. In: Hansen, H. R.; Karagiannis, D.; Fill, H.-G. (Hrsg.): *Business Services: Konzepte, Technologien, Anwendungen*. 9. Internationale Tagung Wirtschaftsinformatik. 25. – 27. Februar 2009, Wien. Band 2, Wien, S. 339-348.
- Knackstedt, R.; List, L.; Stein, A.; Becker, J.; Barth, I. (2009): Towards A Reference Model for Online Research Maps. In: *Proceedings of the 17th European Conference on Information Systems (ECIS 2009)*. Verona, Italy, 2009.
- Knackstedt, R.; Seidel, S.; Janiesch, C. (2006): Konfigurative Referenzmodellierung zur Fachkonzeption von Data-Warehouse-Systemen mit dem H2-Toolset. In: J. Schelp, R. Winter, U. Frank, B. Rieger, K. Turowski (Hrsg.): *Integration, Informationslogistik und Architektur*. DW2006, 21.-22. Sept. 2006, Friedrichshafen. *Lecture Notes in Informatics*. Bonn 2006, S. 61-81.



Scharl, A. (1997): Referenzmodellierung kommerzieller Masseninformati<sup>o</sup>nssysteme. Idealtypische Gestaltung von Informationsangeboten im World Wide Web am Beispiel der Branche Informationstechnik, Frankfurt am Main 1997.

Wedeking, H. (1981): Datenbanksysteme I – Eine konstruktive Einföhrung in die Datenverarbeitung in Wirtschaft und Verwaltung. 2. Aufl., Mannheim u. a. 1981.



# **Generierung von neuartigem Wissen durch die Analyse von Weblogs**

## *Vorstellung eines Forschungsvorhabens*

Andreas Hilbert, Andreas Schieber

Technische Universität Dresden – Professur für Wirtschaftsinformatik  
Business Intelligence Research

### **1 Einleitung**

In den vergangenen Jahren wurde die Entwicklung des World Wide Webs von einem Thema beherrscht: der Interaktivität (KAISER (2009), S. 379). Darunter wird das Mitgestalten von Web-Inhalten durch den Nutzer verstanden. Im Zuge dieser Entwicklung hat sich vor allem die Anzahl der Weblogs stark erhöht (WORDPRESS (2009)). Mit der Weblog-Suchmaschine Technorati beispielsweise ließen sich bereits letztes Jahr Informationen aus über 112 Mio. Weblogs finden (TECHNORATI (2008)).

Der Begriff Weblog ist eine Wortneubildung aus den Worten Web und Log(buch) (FISCHER (2007), S. 7). In einem Weblog werden – vergleichbar mit einem Tagebuch – aktuelle Ereignisse und Entwicklungen aus der realen oder virtuellen Welt auf einer Webseite im Internet dokumentiert. Inzwischen hat sich für den Begriff Weblog auch die Kurzform Blog etabliert.

In den Weblogs wird überwiegend über Erlebnisse aus dem täglichen Leben berichtet, um sie mit anderen – etwa Freunden und Bekannten – zu teilen, und so in Verbindung zu bleiben (ZERFAB & BOGOSYAN (2007), S. 7). Inzwischen schreiben die Autoren jedoch nicht mehr nur über Privates, sondern auch über allgemein interessante Themen wie Politik oder Wirtschaft und tauschen untereinander Erfahrungen mit Produkten oder Unternehmen aus. In den Diskussionen werden auf diese Weise sowohl Meinungen als auch Anregungen geäußert, mit deren Hilfe sich die Produkte oder Dienstleistungen eines Unternehmens im Sinne des Kunden weiterentwickeln und verbessern lassen.

Daher führen immer mehr Unternehmen einen eigenen Weblog als weiteren Kommunikationskanal zu ihren Kunden ein, um von ihnen etwa Ratschläge zur Produktverbesserung zu erhalten (WRIGHT (2006), S. 25). Diese Entwicklung verfolgt auch die Studie Euro-

Blog, die Marketing-Experten zu ihrem Umgang mit Weblogs befragt. Seit Beginn der Untersuchung im Jahr 2006 steigt die Anzahl der Unternehmen kontinuierlich an, die Weblogs als Informationsquelle heranziehen und selbst zur Kommunikation nutzen (ZERFAß ET AL. (2007), S. 13). Viele dieser Unternehmen geben dabei an, dass neben dem eigenen Weblog auch Einträge in anderen Weblogs beobachtet werden müssen, um zeitnah z.B. auf Beschwerden oder Produktmängel reagieren zu können (ZERFAß ET AL. (2007), S. 22). Auch DAVIS & OBERHOLTZER (2008), S. 3 weisen darauf hin, dass Blogs bei der Sammlung von Kundenfeedback eine wichtige Informationsquelle sind. Die Analyse von Weblogs und deren Einträgen ist daher eine Ergänzung zur Marktforschung, die wertvolle Ergebnisse liefert.

Die Zahl der bestehenden Weblogs und Einträge ist jedoch inzwischen unüberschaubar geworden, so dass es nur in sehr geringem Umfang möglich ist, die Einträge relevanter Weblogs auszuwerten und wichtige Inhalte zu erkennen. In der Literatur finden sich daher zahlreiche Arbeiten, die sich mit der Analyse von Weblogs beschäftigen. Mit Methoden der Business Intelligence wie Data und Text Mining lassen sich wesentliche Einträge aus der unübersichtlichen Menge separieren. Als Datenstamm für diese Methoden müssen die Blog-Einträge zunächst im World Wide Web gefunden und extrahiert werden. Dazu ist ein System zu entwickeln, das Weblogs im Web identifiziert und deren Einträge speichert. Im Anschluss sollen diese Daten analysiert werden können, um Informationen, wie beispielsweise Kundenmeinungen, herauszufiltern.

## 2 Forschungsziele

Die Ziele der Forschungsarbeit lassen sich in einen theoretischen, erkenntniszielgeleiteten Teil sowie in einen konzeptionellen, gestaltungszielorientierten Teil gliedern; als Gestaltungsziel wird ein Konzept für ein Blog-Analysesystem erstellt, das anschließend prototypisch implementiert und evaluiert wird.

Der theoretische Bereich umfasst daher als Erkenntnisziel einerseits die Aufarbeitung des aktuellen Forschungsstands zu Methoden und Vorgehensweisen bei der Analyse von Weblogs und deren Einträgen. Andererseits beinhaltet dieser Teil auch eine Bestimmung möglicher Analyseziele, von denen einige in Abschnitt 4.4 vorgestellt werden.

Die Konzeption erläutert daraufhin den Aufbau eines Systems, das – im Sinne eines Management Support Systems – die Ergebnisse der Analysen geeignet darstellen bzw. den Anwendern bereitstellen kann. Im Rahmen der Analyse von Weblogs kommen beispielsweise Vergleiche von Diskussionen aus Blog-Einträgen und -Kommentaren in Betracht, z.B. zu welcher Zeit besonders über ein bestimmtes Thema diskutiert wurde. Analog kann für Unternehmen wichtig sein, die Verkaufszahlen ihrer Produkte mit den Diskussionen

in den Weblogs in Bezug zu bringen; in diesem Zusammenhang kann das Analysesystem als Messinstrument für Marketing-Aktivitäten dienen, wenn anhand der Diskussionsintensität erkannt werden kann, dass unterschiedliche Werbemethoden verschiedene Diskussionsintensitäten in der Blogosphäre bewirken.

Nach der Konzeption befasst sich der letzte Teil der Arbeit mit der prototypischen Umsetzung eines solchen Weblog-Analyse-Systems. Dabei soll – nach Möglichkeit mit Opensource-Tools – eine Implementierung geschaffen werden, welche die Untersuchung von Weblogs sowie die Darstellung der Ergebnisse ermöglicht.

Das Forschungsprojekt befindet sich noch am Anfang. Bisher wurde vor allem aktuelle Literatur zum Thema Blog Mining untersucht. Im Fokus standen dabei Vorgehensweisen und Methoden bei der Analyse von Weblogs. Die so gewonnenen Erkenntnisse sollen im Folgenden beschrieben werden. Dazu gehören zunächst die Definition des Begriffs Weblog sowie die Beschreibung der Eigenschaften eines Blogs. Im Anschluss wird das Ergebnis der Literaturanalyse in Bezug auf aktuelle Verfahren für das Blog Mining dargestellt und systematisch aufgearbeitet. Das Vorgehen bei der Analyse von Weblogs orientiert sich dabei an einem Standardprozess für Data Mining, der in jeder Phase an die spezifischen Anforderungen beim Blog Mining angepasst wird.

### 3 Eigenschaften von Weblogs

Zur Differenzierung zwischen gewöhnlichen Webseiten und Blogs muss der Begriff Weblog eindeutig definiert werden: Nach SCHMIDT (2006), S. 13 sind Weblogs „regelmäßig aktualisierte Webseiten, die bestimmte Inhalte (...) in umgekehrt chronologischer Reihenfolge darstellen. Die Beiträge sind einzeln über URLs adressierbar und bieten in der Regel die Möglichkeit, Kommentare zu hinterlassen. Dadurch sowie durch Verweise auf andere Weblogs (...) bilden sich Netzwerke von untereinander verbundenen Texten und Webseiten heraus; die Gesamtheit aller Weblogs wird auch als ‚Blogosphäre‘ bezeichnet.“ Darüber hinaus weist ein Blog weitere, teils optionale Eigenschaften auf (FISCHER (2007), S. 43ff.). So besteht jeder Eintrag aus einem Titel, dem Veröffentlichungsdatum und dem Text. Die Leser können zu jedem Eintrag Kommentare hinterlassen. Des Weiteren kann der Autor eines Eintrags angegeben werden, und jeder Eintrag kann einer oder mehreren Kategorien zugeordnet sein. Die meisten Blogs stellen einen RSS<sup>15</sup>-Feed zur Verfügung, der den Lesern automatisch mitteilt, wenn ein neuer Eintrag im Blog er-

---

<sup>15</sup> RSS: Really Simple Syndication (RSS ADVISORY BOARD (2009)). RSS-Feeds basieren auf dem Dateiformat XML und sind in ihrem Aufbau sehr einfach gehalten (FISCHER (2007), S. 46).

stellt wurde. In der sogenannten Blogroll sind Verknüpfungen zu anderen Blogs gesammelt, die der Blog-Autor regelmäßig besucht.

## 4 Entwicklung eines Vorgehensmodells für die Analyse von Weblogs

Eine Analyse von etablierten Standardprozessen<sup>16</sup> für Data Mining hat gezeigt, dass es sinnvoll ist, als Basis für das Vorgehen bei der Blog-Analyse den KDD-Prozess nach FAYYAD (1996) heranzuziehen. Im Gegensatz zum CRISP-DM, der die Aufgaben aus der Sicht des Managements betrachtet, liegt der Fokus des KDD-Prozesses auf der technischen Sicht. Dies zeigt sich darin, dass die Phasen Datenselektion und -aufbereitung explizit herausgestellt werden (siehe Abbildung 1). Gerade diese Phasen sind bei der Blog-Analyse als kritische Punkte anzusehen, da als Datenquelle lediglich Webseiten im World Wide Web, das bedeutet semi- oder unstrukturierte Daten, zur Verfügung stehen. Diese Daten kommen in der Analyse-Phase zunächst nicht für die klassischen Data-Mining-Methoden in Frage. Stattdessen werden vermehrt verwandte Techniken wie Text Mining, aber auch einfache statistische Analysen eingesetzt.

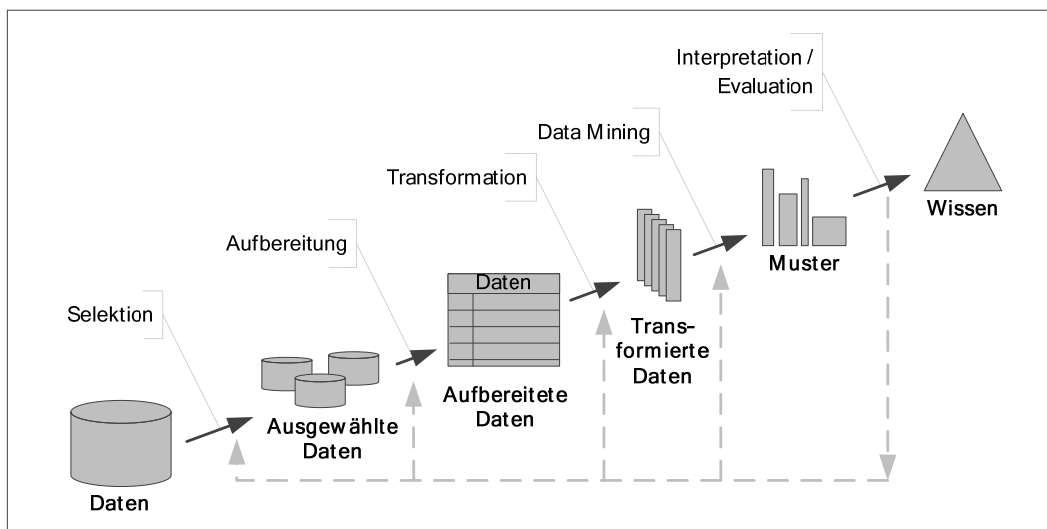


Abbildung 1: Der KDD-Prozess nach FAYYAD (1996), S. 6

Im Anschluss werden die einzelnen Phasen beschrieben und an die speziellen Anforderungen im Rahmen des Blog Mining angepasst.

<sup>16</sup> KDD nach FAYYAD (1996) und CRISP-DM nach CHAPMAN ET AL. (2000)

#### 4.1 Phase 1: Selektion von Weblogs und Einträgen

In der Selektionsphase werden die Daten ausgewählt, die später analysiert werden sollen. Im Gegensatz zu traditionellen Data-Mining-Projekten liegen diese Daten nicht in strukturierter Form vor. Die relevanten Weblogs müssen erst identifiziert werden, bevor einzelne Einträge dieser Weblogs gespeichert werden können.

##### *Identifikation von Weblogs im WWW*

Die Herausforderung bei der Identifikation von Weblogs ist, sie von gewöhnlichen Webseiten zu separieren. Dazu müssen spezifische Kriterien herangezogen werden, die eine Differenzierung zwischen Weblogs und Webseiten aufgrund des Aufbaus bzw. des Inhalts erlauben.

Eine Möglichkeit bei der Identifikation von Weblogs ist die Verwendung eines Crawlers. Ein Crawler durchsucht eine Webseite nach Verknüpfungen zu anderen Webseiten und speichert die gefundenen Links. Nachdem er eine Seite vollständig untersucht hat, wird die nächste URL aus dem Speicher abgerufen und durchsucht (LIU (2007), S. 273f.). In der Regel erhält der Crawler eine Liste mit URLs<sup>17</sup>, die er der Reihe nach abarbeitet (LIU (2007), S. 274). Damit der Crawler nicht auch URLs von gewöhnlichen Webseiten sammelt, muss er anhand von besonderen Kriterien bestimmen können, ob die Seite ein Weblog ist oder nicht. In der Literatur beschäftigen sich daher einige Autoren mit den speziellen Anforderungen, die ein Crawler für die Suche nach Weblogs erfüllen muss (HURST & MAYKOV (2009), S. 1615; SRIPHAEW ET AL. (2008); YU ET AL. (2008), S. 213).

Eine weitere Option zur Identifikation ist die Nutzung von sogenannten Ping-Diensten. Ein Ping-Dienst ist ein Online-Verzeichnis, das zuletzt erschienene Blog-Einträge auflistet. Sobald der Autor einen neuen Eintrag im Blog speichert, wird ein Signal, der Ping<sup>18</sup>, an das Online-Verzeichnis gesendet. Dieses nimmt den neuen Eintrag anschließend auf seiner Webseite auf. Da eine Auflistung eines Ping-Dienstes stets neue Einträge enthält, besteht der Vorteil dieser Methode darin, dass die gefundenen Blogs aktiv sind. Kritisch anzumerken ist jedoch, dass Ping-Dienste zunehmend dazu missbraucht werden, Spam-Blogs oder andere Webseiten bekannt zu machen, die keine Weblogs sind (HURST & MAYKOV (2009), S. 1617). Solche Webseiten müssen vor der Analyse z.B. mit regel- bzw. inhaltsbasierten Verfahren herausgefiltert werden (SRIPHAEW ET AL. (2008), S.

---

<sup>17</sup> URL: Uniform Resource Locator; eine Webseite wird über eine URL aufgerufen, z.B. <http://www.tu-dresden.de> (<http://tools.ietf.org/html/rfc3986>, Abruf am: 30.09.2009)

<sup>18</sup> SAUER (2007), S. 35 bezeichnet einen Ping als „eine lautmalerische Bezeichnung für das Senden eines Echsignals im Web.“

1617). Diese Filterung ist auch bei der Nutzung eines Crawlers erforderlich, um auszuschließen, dass Spam-Blogs analysiert werden.

Darüber hinaus existieren viele Blogs, in denen keine weiteren Einträge mehr veröffentlicht werden (ECK & PLEIL (2006), S. 91). In manchen Situationen ist es aber wichtig, nur aktive Weblogs zu identifizieren. KRAMER & RODDEN (2007) berechnen dazu für jeden Blog eine individuelle Zeitspanne, innerhalb der ein Eintrag erscheinen muss, damit der Blog als aktiv eingestuft wird.

In der Literatur finden sich Belege, dass Weblogs mit beiden Methoden identifiziert werden können. Beispielsweise haben GLANCE ET AL. (2004), S. 3 bei der Entwicklung des Blog-Analyse-Systems BlogPulse<sup>19</sup> mehr als 100.000 Weblogs mit einem Crawler gefunden. In einer weiteren Arbeit beschreiben GLANCE ET AL. (2005), S. 422, dass sie für ihren Datenstamm die Listen mehrerer Ping-Dienste zusammengelegt und so „about 300.000 updated weblogs per day“ GLANCE ET AL. (2005), S. 422 gefunden haben. BANSAL & KOUDAS (2007), S. 1412 beschreiben ein Verfahren, das beide Methoden kombiniert: Ein Crawler erhielt dabei eine Liste eines Ping-Dienstes mit aktiven Blogs und hat darüber weitere Blogs gefunden.

### *Identifikation von Einträgen*

Nach der Identifikation von Weblogs müssen die Einträge und Kommentare in den gefundenen Webseiten ebenfalls identifiziert und anschließend extrahiert werden. Hierbei besteht zunächst die Herausforderung darin, den Text des Eintrags auf der Webseite zu lokalisieren.

Da Weblogs im World Wide Web als HTML-Code gespeichert sind, lässt sich der Seitenquelltext des Blogs untersuchen, um die Einträge zu finden. Jeder Eintrag enthält per definitionem einen Titel, das Veröffentlichungsdatum und den Text; außerdem sind die Einträge eines Weblogs chronologisch umgekehrt sortiert. Daher ist der Quellcode der Webseite auf Elemente zu untersuchen, die in dieses Schema passen. In diesem Zusammenhang kann das Document Object Model<sup>20</sup> (DOM) der Webseite zu Hilfe genommen werden. Dabei werden die HTML-Elemente der Webseite als Baumstruktur dargestellt; am Ende der jeweiligen Äste finden sich Textstellen wie Einträge, Titel, Datumsangaben und Kommentare, die auf diese Weise identifiziert werden können (siehe Abbildung 2).

---

<sup>19</sup> <http://www.blogpulse.com/>

<sup>20</sup> Das Document Object Model ist eine Programmierschnittstelle, mit deren Hilfe auf HTML- oder XML-Dokumente zugegriffen werden kann; online unter <http://www.w3.org/DOM/>.



```

...
<div class="post-1932">
  <h2>UCI Road world Championships 2009&nbsp;Mendrisio</h2> } Titel
  <small class="date">
    <span class="date_day">27</span>
    <span class="date_month">09</span>
    <span class="date_year">2009</span> } Datum
  </small>
  <div class="entry">
    <p><strong>Cadel Evans neuer Straßen weltmeister</strong></p>
    <p>Cadel Evans hat zum großen Schlag ausgeholt. Der 32-jährige Australier ... </p>
  </div>
</div>
...

```

Eintrag

Abbildung 2: Quellcode eines Blog-Eintrags (in Anlehnung an Kirchgessner (2009))

Dieses Vorgehen kann zusätzlich unterstützt werden, wenn die Auszeichnungsregeln der Weblog-Programme berücksichtigt werden. Viele Blogs werden mit Hilfe von speziellen Programmen implementiert, die den HTML-Code für die Webseite des Blogs selbständig generieren. Jedes dieser Programme hinterlässt dabei gewisse Spuren im HTML-Code. Beispielsweise nutzt die Software WordPress das Quellcode-Element `<div class="post">...</div>`, um den Text des Blog-Eintrages darzustellen (ATTARDI & SIMI (2006), S. 3; siehe auch Abbildung 2). Durch diese Besonderheiten im HTML-Code können die Einträge und Kommentare auf der Webseite schneller identifiziert und gespeichert werden.

Viele Weblogs bieten ihren Lesern neben der Webseite des Blogs einen RSS-Feed an, der sie mit Informationen zu aktuellen Einträgen versorgt (siehe Abbildung 3). Diese Feeds können ebenfalls zur Identifikation von Einträgen und Kommentaren verwendet werden. Meistens beinhalten die Feeds alle wichtigen Informationen eines neuen Eintrags: Neben dem Titel und dem Veröffentlichungsdatum können im Feed auch der Name des Autors, die zugewiesenen Kategorien sowie der Inhalt des Eintrags selbst enthalten sein (RSS ADVISORY BOARD (2009)). Die Feeds liegen aufgrund des XML-Formats in maschinenlesbarer Form vor, so dass die übermittelten Informationen direkt weiterverarbeitet werden können (FISCHER (2007), S. 45). Voraussetzung dafür ist jedoch, dass die Feeds den gesamten Text des Blog-Eintrags beinhalten.

Die Literatur zeigt, dass die Analyse des HTML-Codes genutzt wird, um die Einträge aus dem Weblog zu extrahieren. CAO ET AL. (2008) beschreiben beispielsweise einen zweistufigen Prozess, der mit Hilfe des DOM Einträge und Kommentare voneinander trennt. Dabei wird zunächst der „main text“ (CAO ET AL. (2008), S. 2) isoliert. Der main text beinhaltet sowohl den Artikel des Blog-Autors als auch die Reaktionen darauf in Form der Kommentare. Wenn im DOM der Ast mit dem main text gefunden wurde, müssen im zweiten Schritt der Eintrag und die Kommentare voneinander getrennt werden. Dazu untersuchen die Autoren den identifizierten Ast weiter, indem sie die unterschiedliche

Verteilung von HTML-Code-Elementen im Eintrag und in den Kommentaren berechnen und vergleichen. Da ein Eintrag hauptsächlich Text beinhaltet, sind hier weniger Code-Elemente zu erwarten als bei Kommentaren, bei denen mindestens der Name, das Datum und der Kommentartext durch HTML-Code abgetrennt werden müssen.

GLANCE ET AL. (2005) nutzen ebenfalls die Quellcode-Informationen zur Identifikation der Einträge, haben ihr System aber nach anfänglichen Problemen erweitert: Wenn vorhanden, sollen zuerst RSS-Feeds verarbeitet werden, bevor der Quellcode analysiert wird.

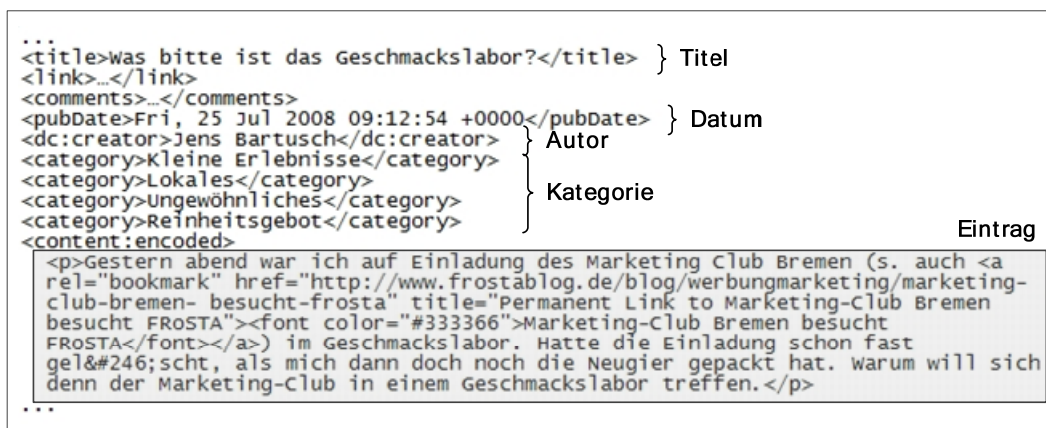


Abbildung 3: Quellcode eines RSS-Feeds mit HTML-Code-Elementen  
(in Anlehnung an FROSTA (2008))

Experimente der Autoren zeigen, dass die Extraktion der Einträge mit Hilfe der RSS-Feeds erwartungsgemäß schneller und besser möglich ist als mit der Quellcode-Analyse (CAO ET AL. (2008), S. 8; GLANCE ET AL. (2005), S. 423). Durch die überwiegend einheitliche Struktur der RSS-Feeds können die benötigten Informationen zudem mit überschaubarem Aufwand extrahiert werden. Dabei muss der Feed jedoch den gesamten Inhalt des Eintrags übermitteln. RSS-Feeds werden zwar von der Mehrheit der Blogs angeboten, jedoch nicht von allen. Daher ist ein zweistufiges System empfehlenswert, das zunächst versucht, den RSS-Feed zu analysieren und im Zweifel eine Extraktion direkt von der Webseite des Blogs vornehmen kann.

## 4.2 Phase 2: Aufbereitung der Daten

Nach der Identifikation der Weblogs und der Extraktion ihrer Einträge werden die gesammelten Informationen in einer Datenbank gespeichert; dadurch ist gewährleistet, dass die Informationen aus Blogs dauerhaft und performant untersucht werden können. Dafür muss vor der Speicherung eine strukturierte und konsistente Datenbasis geschaffen werden.

Dabei ist im ersten Schritt der HTML-Code aufzubereiten. Hierbei ist besonders darauf zu achten, dass Code-Fragmente – etwa für Umlaute oder Sonderzeichen – aus dem Quelltext entfernt werden, um spätere Analyseergebnisse nicht zu verfälschen. Im Ergebnis liegen der Titel des Eintrags, das Datum und der Text sowie optionale Angaben wie der Name des Autors oder die Bezeichnung der Kategorie separat als reiner Text vor.

Im zweiten Schritt werden die Daten vereinheitlicht. Das Ziel ist dabei, Formate und Texte in eine einheitliche Form zu bringen. Dazu zählen z.B. die Datumsangaben, die Schreibweise der Autorennamen und die Bezeichnung der Kategorie.

An dieser Stelle soll darauf hingewiesen werden, dass im Zuge der Aufbereitung auch linguistische Methoden zum Einsatz kommen, welche die Texte analysieren. Für weiterführende Darstellungen sei z.B. auf die Werke von LIU (2007) sowie FELDMAN & SANGER (2007) verwiesen.

### **4.3 Phase 3: Transformation der gespeicherten Daten**

Nach der Aufbereitung steht eine strukturierte Datenbasis für die Analysen zur Verfügung. Je nach Analysemethode kann eine weitere Transformation der Daten erforderlich sein. So können aus den bestehenden Attributen neue Variablen gebildet werden (z.B. ein Attribut, das die Schlüsselwörter eines Eintrags enthält) bzw. Fehlwerte oder Ausreißer eliminiert werden. Diese verfahrensabhängige Transformation wird in der dritten Phase des Prozesses durchgeführt und unterscheidet sich im Bezug auf die Aufgaben nicht von der Phase des KDD-Prozesses.

### **4.4 Phase 4: Analyse der Weblogs**

Für die Analyse von Weblogs stehen unterschiedliche Methoden zur Verfügung. Diese Methoden lassen sich in zeitraumbezogene Längsschnitt- und zeitpunktbezogene Querschnittanalysen unterscheiden (KUSS 2007, S. 41ff.).

Im Rahmen der Längsschnittanalyse kommen Verfahren in Betracht, die Trends in der Blogosphäre aufdecken. Dabei werden die Blog-Einträge mit Methoden des Text Mining analysiert und häufig vorkommende Wörter daraus extrahiert. So sind Themengebiete erkennbar, welche die Aufmerksamkeit der Blog-Autoren auf sich ziehen. Die Analyse-Systeme BlogScope<sup>21</sup> und BlogPulse stellen dem Nutzer z.B. anhand von Diagrammen vieldiskutierte Themen im Verlauf der Zeit dar (BANSAL & KOUDAS (2007), S. 1411; GLANCE ET AL. (2004), S. 4f.).

---

<sup>21</sup> <http://www.blogscope.net/>

Im Gegensatz dazu beschreiben Querschnittanalysen die Ereignisse in der Blogosphäre zu einem bestimmten Zeitpunkt. Dazu gehören die Verfahren des Opinion Mining und der sozialen Netzwerkanalyse. Mit Hilfe des Opinion Mining können die Meinungen der Blogger z.B. zu Produkten aufgedeckt werden. Dabei das Produkt und seine Eigenschaften in den Einträgen gesucht und mit der dazu geäußerten Meinung des Autors verbunden. Diese Bewertungen können als Marktforschungsergebnis zur Weiterentwicklung eines Produkts verwendet werden. In diesem Rahmen haben GLANCE ET AL. (2005), S. 419 Blog-Einträge analysiert und konnten so die Meinungen zu PDAs, Pocket PCs und Smartphones verschiedener Hersteller auswerten.

Die soziale Netzwerkanalyse betrachtet die Blogosphäre im Ganzen. Dabei ist das Ziel, eine Segmentierung der Blogosphäre zu erhalten, indem Verknüpfungen zwischen den Blogs aufgedeckt und z.B. als Netz dargestellt werden. Durch die Analyse dieser Blog-Communities lassen sich einflussreiche Blog-Autoren innerhalb der Gemeinschaft identifizieren. Solche Analysen und deren Ergebnisse sind in der Literatur bereits beschrieben worden (AGARWAL ET AL. (2008); ZHOU & DAVIS (2006)). GLANCE ET AL. (2005), S. 420ff. zeigen mit Hilfe einer Fallstudie eine Kombination von Opinion Mining und sozialer Netzwerkanalyse: Ein Netz von Einträgen stellt eine Diskussion über einen PDA der Firma Dell dar. Dabei sind drei Segmente von Einträgen zu sehen, die jeweils unterschiedliche Eigenschaften des Geräts diskutieren. Ein „Drill-down“ in eines der Segmente offenbart dabei die schlechte Audioqualität des PDAs (GLANCE ET AL. (2005), S. 421).

#### **4.5 Phase 5: Interpretation / Evaluation der Resultate**

In der letzten Phase des angepassten KDD-Prozesses werden die Analyseresultate interpretiert und hinsichtlich ihrer Eignung evaluiert. Für den Fall, dass die Ergebnisse unzureichend sind, müssen die vorgelagerten Schritte wiederholt werden. Nach positiver Evaluation gehen die Ergebnisse in Wissen über und können so z.B. als Entscheidungsgrundlage für Marketing-Aktivitäten dienen.

#### **4.6 Zusammenfassung in einem Vorgehensmodell für das Blog Mining**

Der für die Blog-Analyse angepasste Prozess teilt sich wie der KDD-Prozess in mehrere, aufeinander folgende Phasen auf (siehe Abbildung 4).

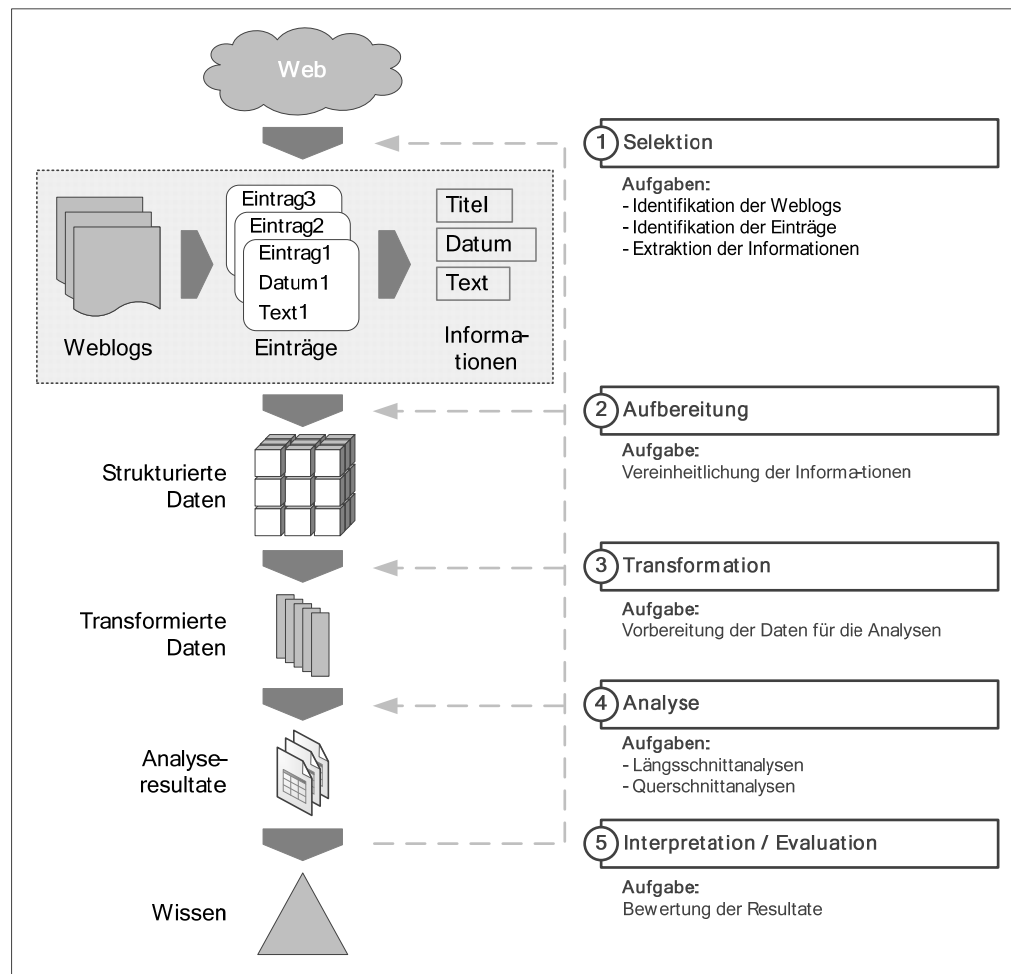


Abbildung 4: Das Vorgehensmodell für das Blog Mining  
(in Anlehnung an FAYYAD (1996), S. 6)

Das Vorgehensmodell zeigt die Abfolge der fünf Phasen zur Durchführung eines Blog-Mining-Projektes. Sie beginnen mit der Datenerhebung und enden mit der Interpretation der Resultate. Im ersten Schritt werden im World Wide Web die Weblogs und Einträge identifiziert sowie extrahiert, die für die Untersuchung relevant sind. Diese Daten sind anschließend aufzubereiten, so dass sie in einer einheitlichen Form vorliegen. In der nächsten Phase findet – teils abhängig von den eingesetzten Analysemethoden – eine Vorverarbeitung der Blog-Einträge statt. Im Anschluss können die Informationen mit Hilfe unterschiedlicher Methoden ausgewertet werden. Sind nach der Evaluation und Interpretation der erzielten Resultate weitere Analysen nötig, müssen je nach Situation die Daten neu ausgewählt, aufbereitet oder transformiert werden. Bei erfolgreicher Evaluation gehen die Ergebnisse in Wissen über.

## 5 Fazit und Ausblick

Diese Arbeit beschreibt ein aktuelles Forschungsvorhaben, bei dem neues Wissen durch die Analyse der Blogosphäre gewonnen werden soll. Der Fokus liegt auf den Methoden der Business Intelligence, welche die dabei anfallenden Aufgaben unterstützen. In diesem Zusammenhang wurde das Vorgehen bei der Sammlung und Analyse von Daten im Rahmen des Blog Mining dargestellt. Dabei wurde eine Adaption des KDD-Prozesses für traditionelle Data-Mining-Projekte auf die besonderen Anforderungen bei der Analyse von Weblogs vorgenommen. Vor allem in den Phasen Selektion, Aufbereitung und Analyse (bzw. Modelling) sind Anpassungen erforderlich, um die spezifischen Aufgaben beim Blog Mining zu integrieren. Im weiteren Verlauf wurden Methoden der Datenerhebung, -speicherung und -analyse vorgestellt und deren erfolgreiche Anwendung anhand von Literaturbelegen nachgewiesen.

Es besteht jedoch weiterer Forschungsbedarf bei den jeweils dargestellten Schritten. Vor allem bei der Datenselektion müssen potente Alternativen zur Identifizierung von Weblogs geschaffen werden. Auch bei der Speicherung der Einträge ist weitere Forschungsarbeit nötig. Um die Daten in geeigneter Form für die Auswertungen bereithalten zu können, müssen dafür analytische Datenmodelle entwickelt werden.

Generell dürfen für das Forschungsvorhaben neue – blog-ähnliche – Technologien in der schnelllebigen Welt des World Wide Web nicht übersehen werden. Dies bezieht sich vor allem auf sogenannte „micro-blogging“-Dienste wie Twitter (HONEYCUTT & HERRING (2009), S. 1). Twitter erlaubt dem Nutzer, kurze Nachrichten bis 140 Zeichen zu senden, die auf der Twitter-Webseite<sup>22</sup> aufgelistet werden. Damit sind diese Nachrichten – wie bei einem Weblog auch – öffentlich einsehbar. Aufgrund des hohen Nachrichtenaufkommens und aufgrund der Verwandtschaft mit Weblogs hat die Blog-Suchmaschine Technorati inzwischen eine Webseite (Twittorati<sup>23</sup>) entwickelt, mit deren Hilfe Twitter-Nachrichten gesucht werden können (TECHNORATI (2009)).

## 6 Literatur

Agarwal, N.; Liu, H.; Tang, L.; Yu, P. (2008): Identifying the influential bloggers in a community, in: ACM (Hrsg.): Proceedings of the International Conference on Web Search and Web Data Mining, ACM, New York, New York, S. 207-218.

Attardi, G.; Simi, M. (2006): Blog Mining through Opinionated Words, URL: <http://trec.nist.gov/pubs/trec15/papers/upisa.blog.final.pdf>, Abruf am: 10.09.2009.

---

<sup>22</sup> [www.twitter.com](http://www.twitter.com)

<sup>23</sup> [www.twittorati.com](http://www.twittorati.com)

Bansal, N.; Koudas, N. (2007): BlogScope - A System for Online Analysis of High Volume Text Streams, in: Koch, C.; Gehrke, J.; Garofalakis, M.; Srivastava, D.; Aberer, K.; Deshpande, A.; Florescu, D.; Chan, C.; Ganti, V.; Kanne, C.-C.; Klas, W.; Neuhold, E. (Hrsg.): Proceedings of the 33rd International Conference on Very Large Data Bases, ACM, New York, New York, S. 1410-1413.

Cao, D.; Liao, X.; Xu, H.; Bai, S. (2008): Blog Post and Comment Extraction Using Information Quantity of Web Format, Chinese Academy of Sciences, Institute of Computing Technology, Beijing, China.

Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. (2000): CRISP-DM 1.0 - Step-by-step data mining guide, URL: <http://www.crisp-dm.org/CRISPWP-0800.pdf>, Abruf am: 15.09.2009.

Davis, H.; Oberholtzer, M. (2008): What are they saying about us?, URL: [http://greenfield-ciaosurveys.com/assets/pdfs/Davis\\_0108-Blogmining.pdf](http://greenfield-ciaosurveys.com/assets/pdfs/Davis_0108-Blogmining.pdf), Abruf am: 06.08.2009.

Eck, K.; Pleil, T. (2006): Public Relations beginnen im vormedialen Raum, in: Picot, A.; Fischer, T. (Hrsg.): Weblogs professionell, 1. Auflage, dpunkt-Verlag, Heidelberg, S. 77-94.

Fayyad, U. (1996): Advances in Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, California.

Feldman, R.; Sanger, J. (2007): The text mining handbook, Cambridge Univ. Press, Cambridge.

Fischer, E. (2007): Weblog & Co, 1. Auflage, VDM Müller, Saarbrücken.

Frosta (2008): RSS-Feed zum Blog, URL: <http://www.frostablog.de/blog/feed>, Abruf am: 25.07.2008.

Glance, N.; Hurst, M.; Nigam, K.; Siegler, M.; Stockton, R.; Tomokiyo, T. (2005): Deriving Marketing Intelligence from Online Discussion, Intelliseek Applied Research Center, Pittsburgh, Pennsylvania.

Glance, N.; Hurst, M.; Tomokiyo, T. (2004): BlogPulse - Automated Trend Discovery for Weblogs, Intelliseek Applied Research Center, Pittsburgh, Pennsylvania.

Honeycutt, C.; Herring, S. (2009): Beyond Micro-Blogging: Conversation and Collaboration via Twitter, Abruf am: 20.07.2009.

Hurst, M.; Maykov, A. (2009): Social Streams Blog Crawler, in: IEEE (Hrsg.): Proceedings of the 25th International Conference on Data Engineering, IEEE Press, S. 1615-1618.

Kaiser, C. (2009): Analyse von Meinungen in sozialen Netzwerken des Web 2.0, in: Hansen, H.; Karagiannis, D.; Fill, H.-G. (Hrsg.): Business Services: Konzepte, Technologien, Anwendungen, 1. Auflage, Österreichische Computer Gesellschaft, Wien, S. 379-388.

Kirchgessner (2009): UCI Road World Championships 2009 Mendrisio, URL: <http://kirchgessner.wordpress.com/2009/09/27/uci-road-world-championship/>, Abruf am: 04.10.2009.

Kramer, A.; Rodden, K. (2007): Applying a User-Centered Metric to Identify Active Blogs, in: ACM (Hrsg.): Conference on Human Factors in Computing Systems, ACM, New York, New York, S. 2525-2530.

Liu, B. (2007): Web Data Mining, 1. Auflage, Springer-Verlag, Berlin, Heidelberg.

RSS Advisory Board (2009): RSS 2.0 Specification (version 2.0.11), URL: <http://www.rssboard.org/rss-specification>, Abruf am: 29.09.2009.

Sauer, M. (2007): Weblogs, Podcasting & Online-Journalismus, 1. Auflage, O'Reilly, Köln.

Schmidt, J. (2006): Weblogs, 1. Auflage, UVK Verlagsgesellschaft, Konstanz.

Sriphaew, K.; Takamura, H.; Okumura, M. (2008): Cool Blog Identification Using Topic-Based Models, in: IEEE/WIC/ACM (Hrsg.): International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Press, S. 402-406.

Technorati (2008): Technorati: About Us, URL: <http://www.technorati.com/about/>, Abruf am: 21.08.2008.

Technorati (2009): We've Launched Twittorati - Discover Where Blogs and Tweets Converge, URL: <http://technorati.com/weblog/2009/07/512.html>, Abruf am: 20.07.2009.

WordPress (2009): Stats « WordPress.com, URL: <http://en.wordpress.com/stats/>, Abruf am: 29.09.2009.

Wright, J. (2006): Blog Marketing als neuer Weg zum Kunden, 1. Auflage, Redline Wirtschaft, Heidelberg.

Yu, F.; Zheng, D.; Zhao, T.; Cheng, X. (2008): Structure and Content Based Blog Pages Identification, in: Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery, Vol. 2, S. 213-217.

Zerfaß, A.; Bogosyan, J. (2007): Blogstudie 2007, Universität Leipzig, Institut für Kommunikations- und Medienwissenschaft, Leipzig.

Zerfaß, A.; Sanhu, S.; Young, P. (2007): EuroBlog 2007: European Perspectives on Social Software in Communication Management, URL: [www.euroblog2007.org/euroblog2007-results.pdf](http://www.euroblog2007.org/euroblog2007-results.pdf), Abruf am: 23.08.2007.

Zhou, Y.; Davis, J. (2006): Community Discovery and Analysis in Blogspace, International World Wide Web Conference Committee, Edinburgh, Scotland.



# **Rahmenkonzept für das Management unternehmensindividueller BI-Architekturen**

Walid Mehanna

Universität Stuttgart, Lehrstuhl für ABWL und Wirtschaftsinformatik I

## **1 Motivation**

Business Intelligence (BI) beschreibt einen integrierten, unternehmensspezifischen, IT-basierten Gesamtansatz zur betrieblichen Entscheidungsunterstützung (Kemper, Mehanna, & Unger, 2006, S. 8). Die Entwicklung von einzelnen BI-Anwendungssystemen erfolgt anhand von speziellen BI-Werkzeugen und bildet Teilaspekte des Gesamtansatzes ab. Für moderne BI-Anwendungen kommen in der Regel mehrere spezialisierte Technologiekomponenten im Zusammenspiel zum Einsatz (Bange, 2006; Kemper et al., 2006, S. 10).

Aus Unternehmenssicht führen die Summe der Anwendungssysteme, der eingesetzten Werkzeuge sowie deren Zusammenspiel zu einer umfangreichen und hochindividuellen *BI-Architektur* (Mertens & Meier, 2009, S. 16; Kemper et al., 2006, S. 19ff.). Dabei werden vor allem in großen Unternehmen aus Gründen der Professionalisierung die Entwicklung, der Betrieb und die Nutzung von Business Intelligence auf mehrere spezialisierte Stellen in Fachbereich und IT verteilt (Finger, 2008; Miller, Bräutigam, & Gerlach, 2006; Unger & Kemper, 2008). Ein BI-Anwendungssystem wird somit zum Ergebnis einer differenzierten Wertschöpfungskette, entlang derer einzelne BI-Services zur Verfügung gestellt werden (Horakh, Baars, & Kemper, 2008). In der Gesamtperspektive entsteht ein komplexes System mit mehreren Akteuren, Services und Technologien, welches das Management vor die Herausforderung stellt, Transparenz über die Zusammenhänge zu erlangen und ein unternehmensspezifisches Gesamtoptimum zu finden.

Aktuelle Ansätze thematisieren die *BI-Governance* als Instrument zur Abgrenzung von Verantwortlichkeiten und zur Definition des Zusammenspiels (Arnott, 2006; Gutierrez, 2006; ESCC, 2009; Finger, 2008; Trost & Zirkel, 2008; Larson & Matney). Analog zu dem übergeordneten Themenfeld der IT-Governance, werden hierbei organisationsweite Richtlinien und Kontrollmechanismen festgelegt, die eine konsequente Ausrichtung an den Unternehmenszielen sicherstellen (IT Governance Institute, 2003, S. 19). Während durch die BI-Governance primär Verantwortlichkeiten und Entscheidungsbefugnisse

festgelegt werden, erfolgt die konkrete Entscheidungsfindung und -umsetzung im Rahmen eines *BI-Managements* (Weill & Ross, 2004, S. 8f.). Im Folgenden wird das BI-Management als Operationalisierung der BI-Governance zum Zwecke der Koordination, Gestaltung und Steuerung der BI-Architekturen verstanden.

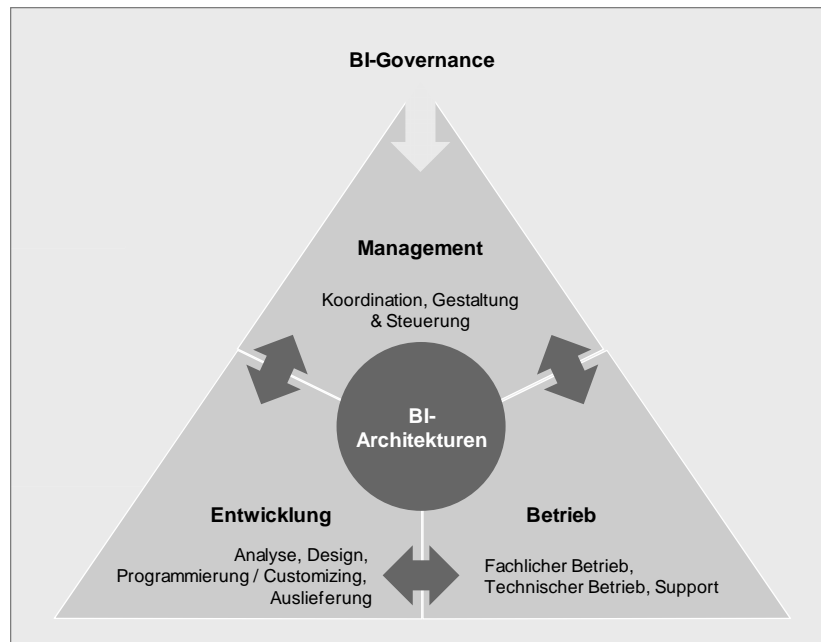


Abbildung 1: Einordnung des Managements von BI-Architekturen

Wie die Abbildung 1 zeigt, ist das Management als Fachfunktion komplementär zu Entwicklung und Betrieb positioniert. Die Governance nimmt eine Rahmenfunktion ein und zeichnet sich durch eine enge Kopplung zum Management aus.

## 2 Forschungsfrage & Zielsetzung

Die Koordination, Gestaltung und Steuerung von BI-Architekturen bedarf des *Zusammenspiels mehrerer Managementaufgaben*. Exemplarisch seien die Themenfelder der BI-Strategie (Totok, 2006; Gluchowski, 2009), des BI-Programmmanagements (Finger, 2008, S. 116; Schwarze, 2006) und der BI-Organisation (Miller et al., 2006; Unger & Kemper, 2008) genannt. Die isolierte, punktuelle Betrachtung einzelner Aufgaben greift zu kurz, da vielfältige Schnittstellen und Interdependenzen bestehen. So kann z.B. die Priorisierung einzelner BI-Initiativen im Programmmanagement nur sinnvoll unter Kenntnis der strategischen Ziele aus Unternehmens-, IT- und BI-Strategie vorgenommen werden (Kemper et al., 2006, S. 153). Dies ist eine wichtige Grundlage der *strategischen Ausrichtung* des BI-Gesamtansatzes auf die Unternehmensziele.

Die organisationsweite Koordination, Gestaltung und Steuerung von Business Intelligence in großen, arbeitsteiligen Unternehmen kann durch den Einsatz eines passenden *Anwendungssystems* zielführend unterstützt werden.

Die daraus abgeleitete Forschungsfrage lässt sich wie folgt zusammenfassen:

»?«

„Wie kann auf Basis eines **Anwendungssystems** das **Management** von unternehmensindividuellen BI-Architekturen in großen Unternehmen – unter Berücksichtigung einer **strategischen Ausrichtung** und der **Integration von Managementaufgaben** – unterstützt werden?“

Die Abbildung eines solchen Anwendungssystems kann im Rahmen eines Forschungsvorhabens nur Beispielcharakter haben, da vielfältige Unternehmensspezifika wie existierende Richtlinien, Aufbau- und Ablauforganisation und Unternehmenskultur berücksichtigt werden müssen. Als generelles Ergebnisartefakt bietet sich ein *Rahmenkonzept* an, das den Lösungsraum für die unternehmensindividuelle Gestaltung determiniert. Bewährt haben sich dabei *Referenzmodelle*, die allgemeingültige Elemente eines Systems als Empfehlung strukturiert zusammenfassen (Becker, Delfmann, & Rieke, 2007). Um die Abbildbarkeit in ein Anwendungssystem zu prüfen, ist die exemplarische Realisierung in einem *Prototyp* sinnvoll. Dabei ist die Abgrenzung zum operativen Bereich hervorzuheben: Fokus ist die Betrachtung der *übergeordneten Koordinationsebene* und nicht die vollständige Berücksichtigung aller Entwicklungs- und Betriebsprozesse. Die daraus abgeleitete Zielsetzung lässt sich wie folgt zusammenfassen:

»!

«

„Vorrangiges Ziel ist die **Entwicklung eines Rahmenkonzeptes** in Form eines **Referenzmodells**, das exemplarisch in einem **Prototypen** realisiert wird und als Komplement zu Entwicklung und Betrieb von BI-Architekturen die **übergeordnete Koordinationsebene** abbildet.“

### 3 Wissenschaftstheoretische Positionierung & Vorgehensweise

Das folgende Forschungsdesign in Abbildung 2 ergänzt das Forschungsziel um die grundlegende wissenschaftstheoretische Positionierung und die angedachte Forschungsmethodik (Becker, Holten, Knackstedt, & Niehaves, 2004). Aufgrund des Empfehlungscharakters eines Referenzmodells ist das Vorhaben klar dem konstruktionswissenschaftlichen

Paradigma (*design science*) zugeordnet (March & Smith, 1995; Hevner, March, Park, & Ram, 2004; Becker & Pfeiffer, 2006).

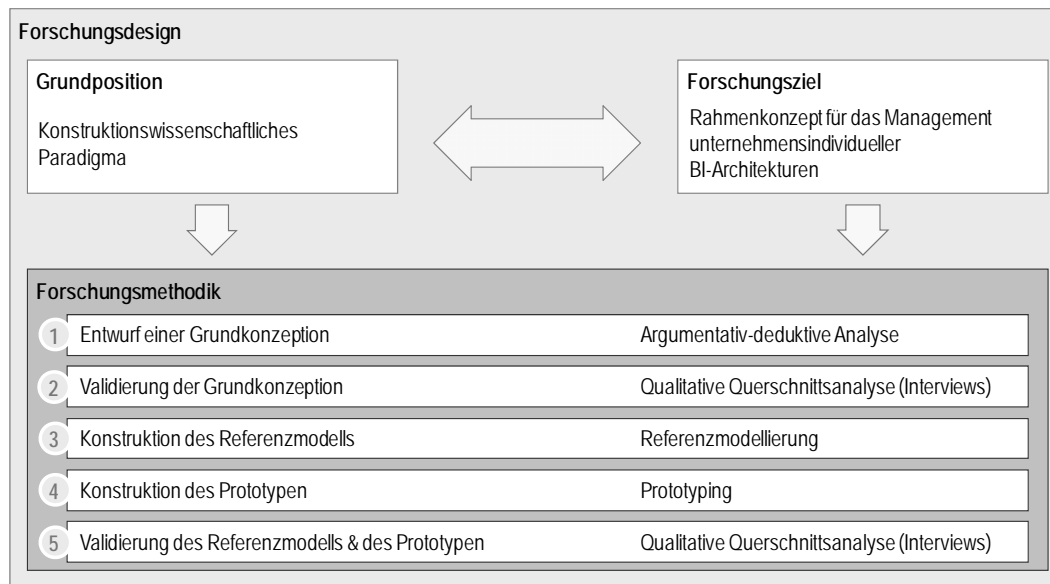


Abbildung 2: Forschungsdesign des Vorhabens

Die Wahl der Forschungsmethoden orientiert sich am etablierten Methodenprofil der Wirtschaftsinformatik und umfasst unter Berücksichtigung des Forschungsziels mehrere Methodenklassen (Wilde & Hess, 2007). Im ersten Schritt wird im Rahmen einer argumentativ-deduktiven Analyse die *Grundkonzeption* entworfen. Auf Basis einer umfangreichen Literaturanalyse werden hierbei mögliche BI-Managementaufgaben identifiziert und zusammengefasst. Im zweiten Schritt erfolgt die *Validierung* der Grundkonzeption durch Experteninterviews mit leitenden Verantwortlichen von BI-Organisationseinheiten in großen Unternehmen. Auf Basis dieser Erkenntnisse und aktueller Forschungsergebnisse aus dem Themengebiet der BI-Governance wird ein *Referenzmodell* abgeleitet. Dessen Umsetzbarkeit in ein Anwendungssystem wird grundsätzlich geprüft durch die Überführung in einen funktionalen Prototyp. Abschließend erfolgt eine erneute *Validierung* des Referenzmodells und des Prototypen im Rahmen von Experteninterviews, die bei Bedarf in einer weiteren Anpassung oder Erweiterung mündet.

## 4 Grundkonzeption zum Management von BI-Architekturen

Der Entwurf der *Grundkonzeption* zielt auf die Erarbeitung eines Ordnungsrahmens, der einen schnell erfassbaren Überblick der BI-Managementaufgaben bietet und die Zusammenhänge der nachgeordneten Detaillierungsebenen aufzeigt (Meise, 2001, S. 62f.). Wie die Abbildung 3 skizziert, spielen für die Bereitstellung von BI-Anwendungen zuneh-

mend umfangreiche Wertschöpfungsketten von BI-Services eine wichtige Rolle (Horakh, Baars, & Kemper, 2008). Für die Inbetriebnahme eines BI-Anwendungssystems können vielfältige Leistungen auf einem internen oder externen Markt in Anspruch genommen werden. Die isolierte Systementwicklung im Unternehmen wird teilweise oder vollständig durch die Komposition von Dienstleistungen und Komponenten ersetzt.

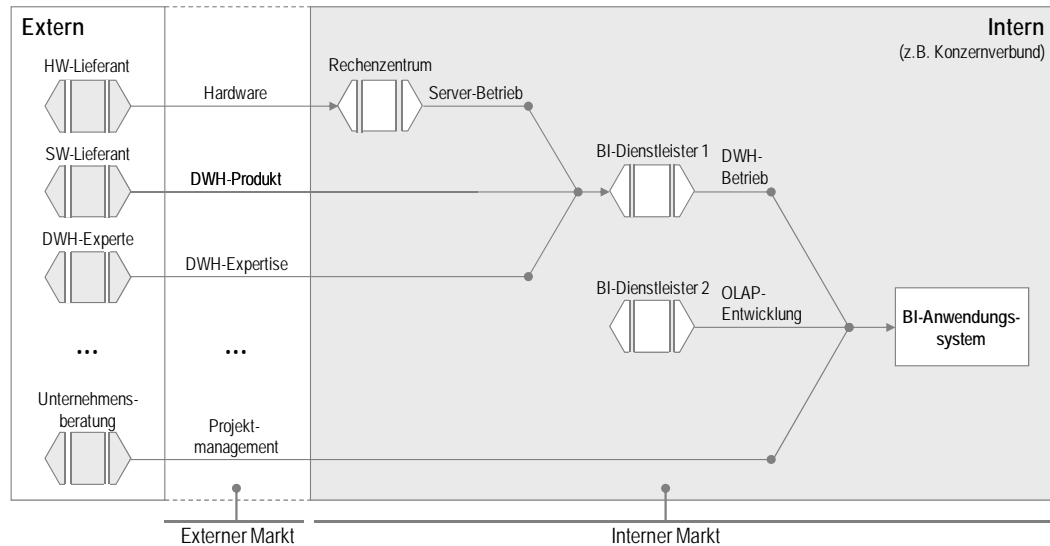


Abbildung 3: Komposition eines BI-Anwendungssystems durch BI-Services einer Wertschöpfungskette

In Anlehnung an das „Supply-Chain Operations Reference Model“ bietet sich diese wertschöpfungsorientierte Sicht der einzelnen Akteure als Rahmenstruktur für die Gliederung der Managementaufgaben an (Supply-Chain Council, 2006). Die in der argumentativ-deduktiven Analyse identifizierten und in den Interviews validierten Aufgabenbündel werden in der Abbildung 4 den Perspektiven Governance, Beschaffung, Produktion und Absatz zugeordnet.

Die Aufgaben der *BI-Governance* schaffen die Voraussetzungen für eine unternehmensweit koordinierte Managementunterstützung und eine konsequente Ausrichtung an den aktuellen Unternehmenszielen. Das Themenfeld der *Strategie* umfasst vielfältige Aspekte der strategischen Analyse, der Leitbild-Bestimmung, der Potenzialplanung und der grundlegenden Portfolio-Entscheidungen (Totok, 2006; Gluchowski, 2009; Sommer, Bayer, Rosenbaum, Wendlandt, & Overmeyer, 2008; Kemper et al., 2006, S. 149ff.).

Im Rahmen der *Richtlinien und Leitlinien* werden für das Gesamtunternehmen oder Unternehmensteile verbindliche Vorgaben und Empfehlungen für die Ausgestaltung der BI-Architekturen ausgesprochen (Kemper et al., 2006, S. 157f.; Gutierrez, 2006, S. 2). Dabei handelt es sich um Rahmenbedingungen für Entwicklung und Betrieb, wie z.B. Investiti-

onsobergrenzen für Geschäftseinheiten, die Vorgabe von einzelnen Technologiekomponenten oder die Abnahmeverpflichtung von intern angebotenen Dienstleistungen.

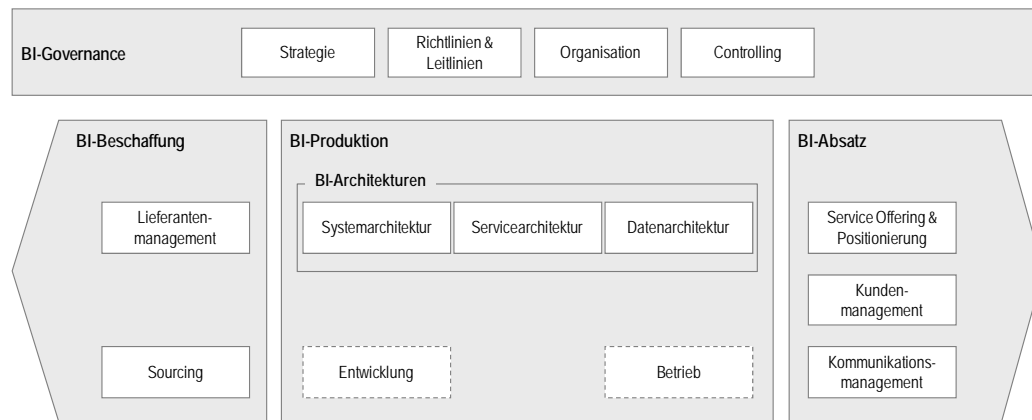


Abbildung 4: Grundkonzeption der Managementaufgaben

Unter dem Aufgabenbündel der *Organisation* werden Rollen und Verantwortlichkeiten in der Bereitstellung und Pflege von BI-Systemen thematisiert. Von besonderem Interesse ist die Aufgabenteilung zwischen Fachbereich und IT: aufgrund einer zunehmenden Endanwenderorientierung der Software eröffnen sich neue Gestaltungsoptionen in Richtung Agilität (Finger, 2008, S. 117&125ff.; Schlegel, 2007; Weill & Ross, 2004, S. 59ff.). Fachliche und technische Aspekte der BI-Architekturen werden in regelmäßigen Abständen durch feste Gremien diskutiert und entschieden (Watson, Fuller, & Ariyachandra, 2004, S. 442; Gutierrez, 2006, S. 3ff.). Weiter an Bedeutung gewinnen auch spezialisierte Unterstützungseinheiten für die Entwicklung und den Betrieb von BI-Anwendungen (Unger & Kemper, 2008). Oftmals werden sie unter der Bezeichnung „BI Competency Center“ oder dem Kürzel BICC geführt (Miller, Bräutigam, & Gerlach, 2006; Vierkorn & Friedrich, 2008).

Die kennzahlenbasierte Steuerung der BI-Architekturen erfolgt anhand eines strategischen und eines operativen *Controllings*. Als Instrument zur Strategieoperationalisierung hat sich in der Praxis die „Balanced Scorecard“ bewährt (Kaplan & Norton, 1997). Für den Einsatz im Kontext Business Intelligence sind Anpassungen in der Ausgestaltung notwendig (Vinciguerra, 2004). Ein weiterer strategischer Aspekt ist das Risikomanagement, das aufgrund der zunehmenden Verzahnung von BI mit operativen Systemen und der steigenden Anforderungen hinsichtlich Ausfallsicherheit, Zuverlässigkeit und Performance an Bedeutung zunimmt (Prokein, 2008; Junginger, 2005). Im operativen Controlling stellt ein übergreifendes Ressourcenmanagement die Verfügbarkeit und Priorisierung von Personal und technischer Infrastruktur sicher (IT Governance Institute, 2007, S. 12). Die parallele Entwicklung mehrerer BI-Anwendungssysteme wird durch ein

Multiprojektmanagement gesteuert (Kesten, Müller, & Schröder, 2007, S. 81ff.). Analog dazu dienen die Kennzahlen des Betriebsmanagement, um eine kontinuierliche und effiziente Bereitstellung der Anwendungen zu gewährleisten (Kesten et al., 2007, S. 177ff.).

In der *BI-Produktion* findet die operative Wertschöpfung der BI-Organisationseinheiten statt. In der *Entwicklung* werden Analyse, Design, Programmierung, Customizing und Auslieferung neuer BI-Anwendungssysteme vorgenommen (Baars, Horakh, & Kemper, 2007, S. 1162). Nach erstmaliger Verfügbarkeit übernimmt der *Betrieb* die fachliche und technische Bereitstellung sowie den Support für die Endanwender (Unger & Kemper, 2008, S. 148).

Der Oberbegriff der *BI-Architekturen* bezeichnet das Zusammenspiel von Systemen, Services und Daten zur Managementunterstützung. Die *Systemarchitektur* dient der Beschreibung aller Anwendungssysteme hinsichtlich ihrer Art, ihrer Eigenschaften und ihres Zusammenwirkens (Kemper et al., 2006, S. 19ff.; Mertens & Meier, 2009, S. 16ff.; Bauer, Günzel, Vaduva, & Zeh, 2009). In der *Servicearchitektur* werden die Bestandteile der Anwendungssysteme als separate Services im Sinne von „Bausteinen“ differenziert. Motivation ist die Möglichkeit, einzelne Bestandteile von einem internen oder externen Leistungserbringer zu beziehen. Als Gliederungskriterien dienen die technischen Komponenten (z.B. ETL, DWH und ODS), die Lebenszyklus-Phase (Entwicklung oder Betrieb) und die Geschäftsspezifität (Horakh et al., 2008, S. 204). Letztere bezeichnet die Nähe zu den zu unterstützenden Geschäftsprozessen und kann in die Ausprägungen Hardware, Werkzeuge, Templates und Inhalte detailliert werden. Unternehmensweit einheitliche Inhalte können allerdings nur erreicht werden, wenn für die grundlegenden Informationsobjekte und deren Beziehungen zueinander ein globaler Bauplan auf hoher Abstraktionsebene vorliegt. Im Kontext der Datenmodellierung spricht man von einer dispositiven *Datenarchitektur* (Kemper et al., 2006, S. 153ff.).

Die wertschöpfungsorientierte Sicht wird vervollständigt durch die Managementaufgaben der *BI-Beschaffung*, die sich mit Fragen des Outsourcings, des Offshorings und der Koordination der Dienstleistungs- und Technologie-Lieferanten beschäftigen (Kemper, Baars, & Horakh, 2007; Philippi, 2005). Durch die Bestimmung des *Service Offerings* wird auf Seiten des *BI-Absatzes* die *Positionierung* einer oder mehrerer BI-Organisationseinheiten vorgenommen. Diese können je nach Leistungsspektrum als Assistenz-Center, Volldienstleister, Betriebs-Center, Hosting-Center oder Entwicklungs-Center unterschieden werden (Unger & Kemper, 2008, S. 149). Das *Kundenmanagement* bildet die Schnittstelle zum Leistungsabnehmer und dient der Beziehungspflege, der Verhandlung von Abnahme-Konditionen sowie der Erfassung neuer oder veränderter Geschäftsanforderungen (Zarnekow, Brenner, & Pilgram, 2005, S. 85ff.). Das *Kommunikationsmanagement* stellt sicher, dass Informationen zur Zielsetzung und zur

Leistungsfähigkeit der BI-Organisationseinheiten die Zielgruppen erreicht (Zarnekow et al., 2005, S. 89ff.).

## 5 Angestrebte Ergebnisse und exemplarische Artefakte

Ein Problemlösungsansatz kann in Form eines *Referenzmodells* strukturiert beschrieben werden und somit als Ausgangslösung für eine unternehmensspezifische Ausprägung dienen (Becker & Delfmann, 2004; Fettke & Loos, 2004). Referenzmodelle sind konzeptionelle Informationsmodelle, die Teile von realen betrieblichen Systemen abbilden und sich durch Allgemeingültigkeit, Wiederverwendbarkeit und Anpassbarkeit auszeichnen (Krallmann, Frank, & Gronau, 1999; Becker & Schütte, 2004, S. 65ff.; Schütte, 1998, S. 69ff.). Unter Verwendung von Modellierungsmethoden werden die relevanten Elemente des betrachteten Anwendungs- oder Organisationssystems (z.B. Prozesse, Daten und Akteure) strukturiert abgebildet (vom Brocke & Buddendick, 2004). Der inhärente Empfehlungscharakter des Referenzmodells spiegelt sich wider in Common-Practice-Modellen, die bekannte und etablierte Branchenstandards abbilden und in Best-Practice-Modellen, die zusätzlich innovative theoriebasierte Erkenntnisse mit einbeziehen (Becker, 2004, S. 325; Becker, Algermissen, Delfmann, & Knackstedt, 2002, S. 1392f.). Die Motivation für den Einsatz von Referenzmodellen ist sowohl wirtschaftlicher als auch inhaltlicher Natur: durch die Nutzung einer existierenden Ausgangslösung können Kosteneinsparungen realisiert werden, während auf bewährten betriebswirtschaftlichen Konzepten aufgesetzt wird (Schütte, 1998, S. 75ff.).

Primäres Ergebnisartefakt ist hier ein *Referenzmodell* für das Management von BI-Architekturen. Ergänzt wird es durch das sekundäre Ergebnis eines *Prototypen*, durch das die Umsetzbarkeit des Modells in ein Anwendungssystem geprüft wird. Im Folgenden werden die einzelnen Bestandteile beschrieben und zur Illustration erste Ergebnisse exemplarisch wiedergegeben. Die Abbildung 5 bietet einen Überblick des Abschnitts.



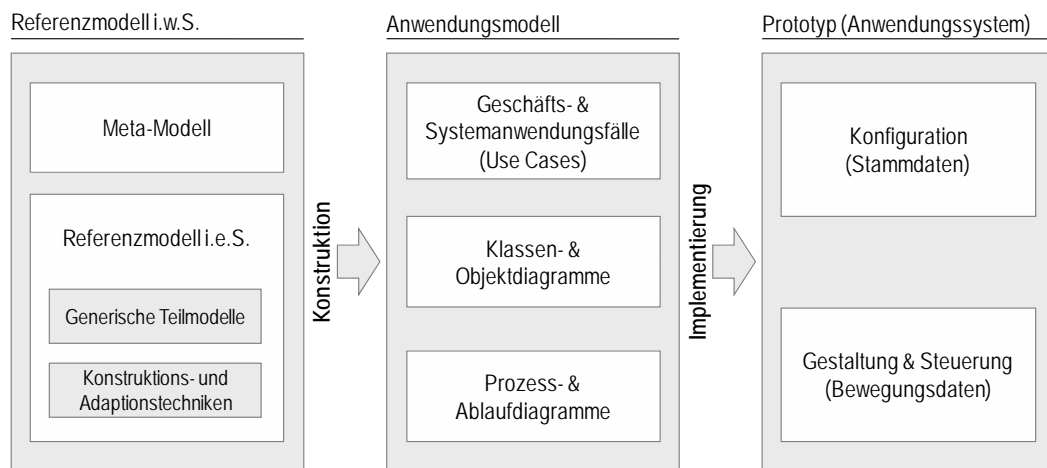


Abbildung 5: Überblick des Abschnitts

Als Modellierungssprache kommt die aktuelle Version 2 der „Unified Modeling Language“ (UML) zum Einsatz (Object Management Group, 2009). Aus Platzgründen wird im Folgenden keine Beschreibung der Notation vorgenommen und auf die vielfältige Literatur verwiesen (Balzert, 2005; Oestereich, 2006; Seemann & von Gudenberg, 2006). Die Grundkonzeption aus dem vorherigen Kapitel dient als Struktur für das Referenzmodell und findet sich im Paketdiagramm in Abbildung 6 wieder.

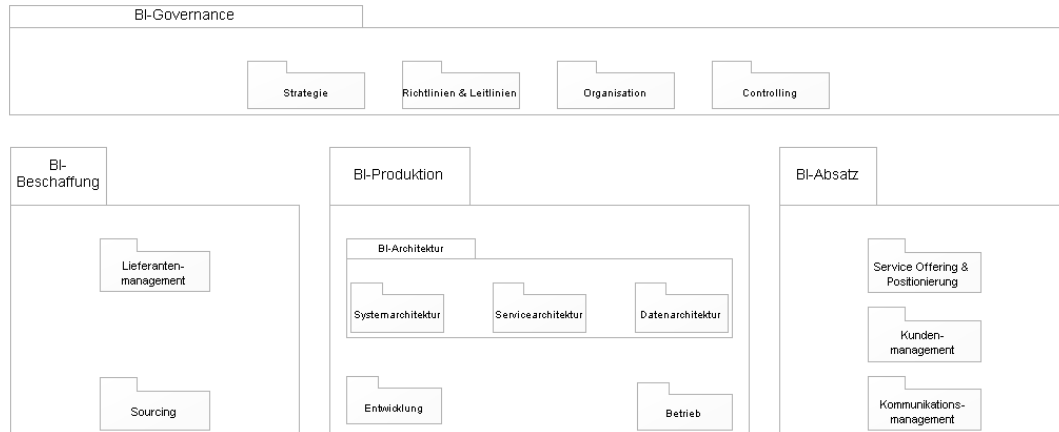


Abbildung 6: Paketdiagramm "Management der BI-Architekturen"

Das Referenzmodell beinhaltet ein *Meta-Modell*, das die Syntax des Modellsystems auf einer höheren Abstraktionsstufe formal beschreibt (Schütte, 1998, S. 73). Die folgende Abbildung 7 zeigt exemplarisch das Meta-Modell für eine Systemarchitektur und ist im gleichnamigen Paket beheimatet.

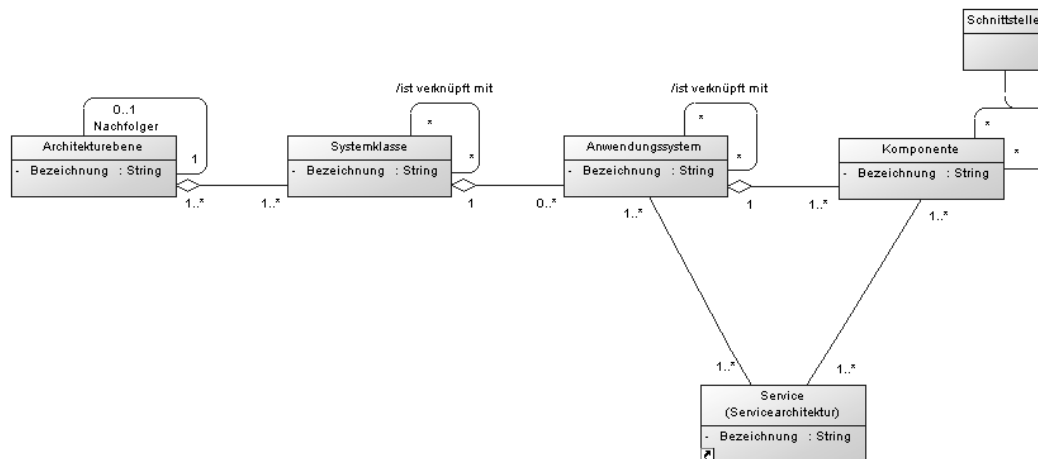


Abbildung 7: Meta-Modell Klassendiagramm BI-Systemarchitektur

Wie ersichtlich werden vier primäre Klassen der „Architekturebene“, der „Systemklasse“, des „Anwendungssystems“ und der „Komponente“ unterschieden. Die Klasse „Service“ ist eine Verknüpfung aus dem Paket „Servicearchitektur“ und daher grafisch markiert. Die Klasse „Schnittstelle“ ist eine Assoziationsklasse, die nur bei Verknüpfungen zwischen Komponenten zum Tragen kommt.

Die Architekturebene ist die grundlegendste Strukturierung, z.B. im Sinne einer „Datenbereitstellung“. Eine Architekturebene kann einen Nachfolger haben, der dann die nächsthöhere Architekturebene in der Hierarchie abbildet. Weiterhin besteht die Architekturebene aus einer Aggregation von mindestens einer Systemklasse, wie z.B. „Data Mart“. Systemklassen können untereinander verknüpft sein. Allerdings ist dies eine abgeleitete Assoziation, die von den Schnittstellen zwischen den Komponenten determiniert wird. Einzelne Anwendungssysteme können den Systemklassen zugeordnet werden. Ein Anwendungssystem besteht wiederum aus mindestens einer Komponente und wird über Services zur Verfügung gestellt.

Diese im ersten Schritt abstrakte Sichtweise gilt es weiter zu detaillieren. Das *Referenzmodell im engeren Sinne* besteht dabei aus mehreren *Teilmodellen* (Schütte, 1998, S. 71). Die Abbildung 8 zeigt einen konzeptionellen Ordnungsrahmen für BI-Anwendungssysteme.

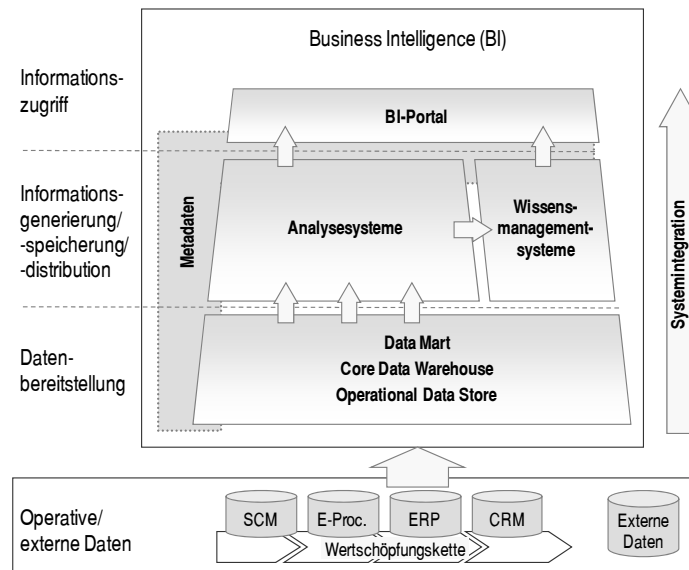


Abbildung 8: BI-Ordnungsrahmen nach Kemper et al. 2006

Unter Anwendung des Meta-Modells zur BI-Systemarchitektur, wird dieser Ordnungsrahmen in ein Objektmodell überführt. Die Abbildung 9 zeigt die Umsetzung. Alternative Systemarchitekturen wie z.B. Mertens & Meier, 2009, S. 16ff. oder Bauer et al., 2009 können im gleichen Sinne übertragen werden und stehen dann dem Modellierer als Alternativen zur Verfügung.

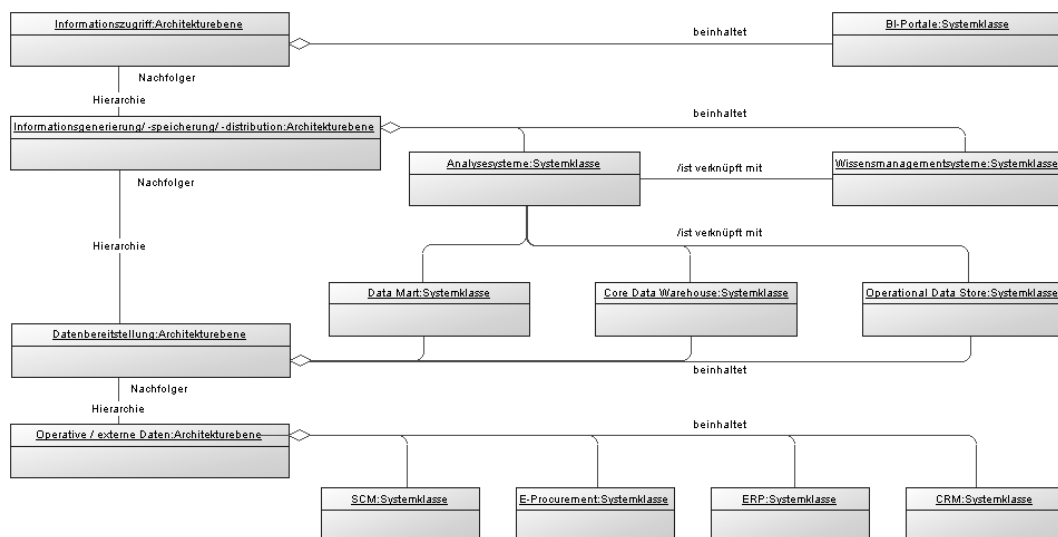


Abbildung 9: Objektdiagramm BI-Architektur nach Kemper et al. 2006

Für die Überführung in ein unternehmensspezifisches Datenmodell werden *Konstruktions- und Adaptionstechniken* benötigt (Becker, Delfmann, Knackstedt, & Kuroпка, 2002; vom Brocke & Buddendick, 2004). Unter ihrer Anwendung wird auf der

gleichen semantischen Stufe das *Anwendungsmodell* erstellt. Die unternehmensindividuellen Anforderungen an die Interaktion mit dem Anwendungssystem werden in Form von *Geschäfts- und Systemanwendungsfällen* – sogenannten Use-Cases – beschrieben (Oestereich, 2006, S. 101ff. & 116ff.). Für jede Managementaufgabe beschreibt ein statisches Modell in Form von *Klassen- und Objektdiagrammen* die Aufgabenelemente in der unternehmensspezifischen Terminologie. Ergänzend beschreibt das dynamische Modell die Funktionsabläufe bei der Nutzung des Anwendungssystems und innerhalb des Anwendungssystems. Hierzu kommen Prozess- und Ablaufdiagramme wie Aktivitätsdiagramme, Szenariodiagramme und Zustandsautomaten zum Einsatz (Balzert, 2005, S. 11).

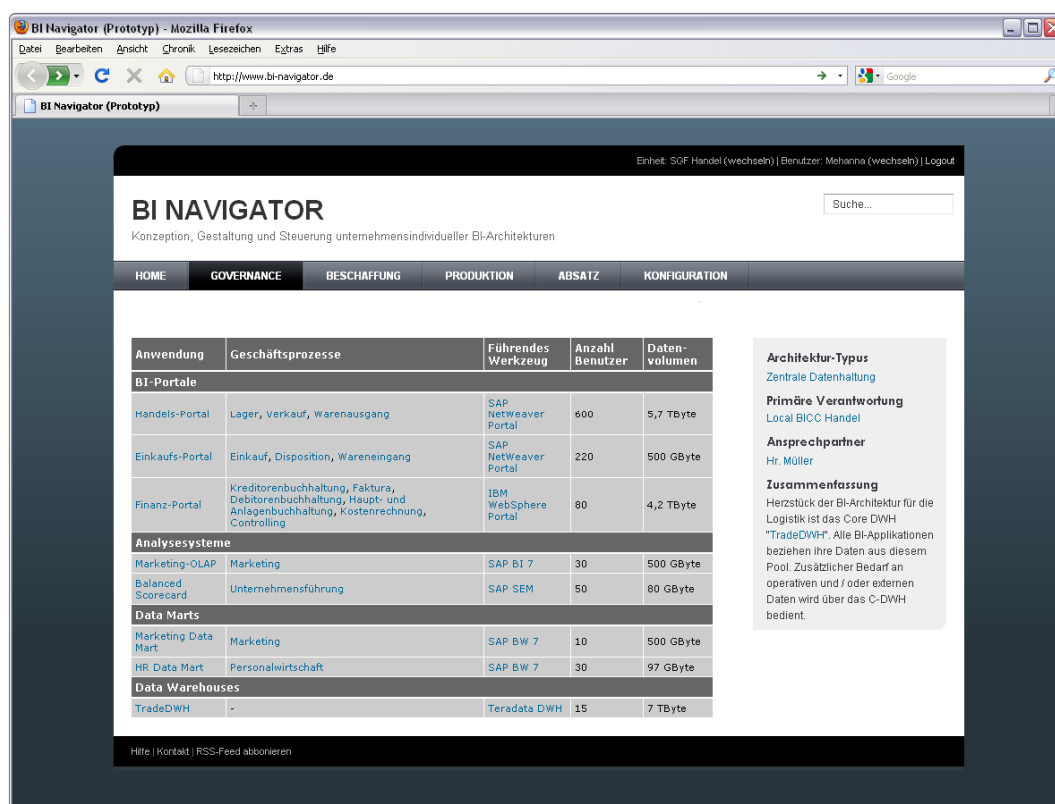


Abbildung 10: Prototyp "Management von BI-Architekturen"

Die Übertragung des Anwendungsmodells in ein lauffähiges *Anwendungssystem* wird durch die prototypische Implementierung sichergestellt. Hierzu werden zwei grundlegende Komponenten mit unterschiedlichen Zielgruppen realisiert. Die *Konfigurationskomponente* ermöglicht das Erfassen von Stammdaten, wie z.B. die Bezeichnung der BI-Organisationseinheiten oder die Kondition eines Service Level Agreements. Die eigentliche Anwendung bietet Funktionalität zur *Gestaltung und Steuerung*, indem z.B. die BI-Strategie durch die Hinterlegung einer Balanced Scorecard

unterstützt wird oder Berichte über die monatliche Erfüllung von Service Levels abrufbar sind. Als abschließenden Ausblick zeigt die Abbildung 10 einen Screen-Prototypen, der die vorangehenden Beispiele in einem webbasierten Anwendungssystem abbildet.

## 6 Literatur

- Arnott, D. (2006). Data warehouse and business intelligence governance: An empirical study. In F. Adam, P. Brezillon, S. Carlsson, & P. Humphreys (Hrsg.), *Creativity and innovation in decision making and decision support* (S. 711–730). London: Ludic Publishing.
- Baars, H., Horakh, T., & Kemper, H.-G. (2007). Business Intelligence Outsourcing: A Framework. In H. Österle, J. Schelp, & R. Winter (Hrsg.), *Proceedings of the 15th European Conference on Information Systems* (S. 1155–1166). St. Gallen: University of St. Gallen.
- Balzert, H. (2005). *Lehrbuch der Objektmodellierung: Analyse und Entwurf mit der UML 2* (2. Auflage). Heidelberg: Spektrum Akademischer Verlag.
- Bange, C. (2006). Werkzeuge für analytische Informationssysteme. In P. Chamoni & P. Gluchowski (Hrsg.), *Analytische Informationssysteme. Business-Intelligence-Technologien und -Anwendungen* (S. 89–110). Berlin: Springer.
- Bauer, A., Günzel, H., Vaduva, A., & Zeh, T. (2009). Aspekte einer Referenzarchitektur. In A. Bauer & H. Günzel (Hrsg.), *Data-Warehouse-Systeme. Architektur Entwicklung Anwendung* (3. Auflage, S. 33–38). Heidelberg: dpunkt-Verlag.
- Becker, J. (2004). Referenzmodellierung: Aktuelle Methoden und Modelle. *WIRTSCHAFTSINFORMATIK*, 46(5), S. 325–326.
- Becker, J., & Delfmann, P. (Hrsg.) (2004). *Referenzmodellierung: Grundlagen, Techniken und domänenbezogene Anwendung*. Heidelberg: Physica.
- Becker, J., & Pfeiffer, D. (2006). Beziehungen zwischen behavioristischer und konstruktionsorientierter Forschung in der Wirtschaftsinformatik. In S. Zelewski & N. Akca (Hrsg.), *Fortschritt in den Wirtschaftswissenschaften. Wissenschaftstheoretische Grundlagen und exemplarische Anwendungen* (S. 39–57). Wiesbaden: Deutscher Universitäts Verlag.

- Becker, J., & Schütte, R. (2004). *Handelsinformationssysteme* (2. Auflage). Frankfurt am Main: Moderne Industrie.
- Becker, J., Algermissen, L., Delfmann, P., & Knackstedt, R. (2002). Referenzmodellierung. *WISU- Das Wirtschaftsstudium*, 31(11), S. 1392–1395.
- Becker, J., Delfmann, P., & Rieke, T. (2007). Referenzmodellierung – Perspektiven für die effiziente Gestaltung von Softwaresystemen. In J. Becker, P. Delfmann, & T. Rieke (Hrsg.), *Effiziente Softwareentwicklung mit Referenzmodellen* (S. 1–9). Heidelberg: Physica.
- Becker, J., Delfmann, P., Knackstedt, R., & Kuropka, R. (2002). Konfigurative Referenzmodellierung. In J. Becker & R. Knackstedt (Hrsg.), *Wissensmanagement mit Referenzmodellen. Konzepte für die Anwendungssystem- und Organisationsgestaltung* (S. 25–144). Heidelberg: Physica.
- Becker, J., Holten, R., Knackstedt, R., & Niehaves, B. (2004). Epistemologische Positionierungen in der Wirtschaftsinformatik am Beispiel einer konsensorientierten Informationsmodellierung. In U. Frank (Hrsg.), *Wirtschaftswissenschaft. Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik. Theoriebildung und -bewertung, Ontologien, Wissensmanagement* (S. 335–366). Wiesbaden: Deutscher Universitäts Verlag.
- ESCC - U.S. Army Enterprise Solutions Competency Center (2009). *Business Intelligence (BI): Policy & Governance*. Letzter Aufruf 5. Oktober 2009 unter <http://escc.army.mil/BI/PolicyGovernance.htm>.
- Fettke, P., & Loos, S. (2004). Referenzmodellierungsforschung. *WIRTSCHAFTSINFORMATIK*, 46(5), S. 331–340.
- Finger, R. (2008). Agile Business Intelligence und klassisches Data Warehousing: Einordnung in ein BI Governance Framework. In T.-L. Mayer, A. Wald, & R. Gleich (Hrsg.), *Advanced Project Management. Advanced Project Management. Herausforderungen Praxiserfahrungen Perspektiven* (S. 113–130). Berlin: LIT.
- Gluchowski, P. (2009). Ansatzpunkte zur Gestaltung einer Business Intelligence-Strategie. In U. Götze & R. Lang (Hrsg.), *Strategisches Management zwischen Globalisierung und Regionalisierung* (S. 387–402). Wiesbaden: Gabler.
- Gutierrez (2006). *Business Intelligence (BI) Governance*. Letzter Aufruf 5. Oktober 2009 unter <http://www.infosys.com/industries/consumer-packaged-goods/white-papers/bi-governance.pdf>.

- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science In Information Systems Research. *MIS Quarterly*, 28(1), S. 75–105.
- Horakh, T., Baars, H., & Kemper, H.-G. (2008). Management von Business Intelligence Services. In B. Dinter, R. Winter, P. Chamoni, N. Gronau, & K. Turowski (Hrsg.): *Band P-138. GI-Edition - Lecture Notes in Informatics (LNI), DW 2008. Synergien durch Integration und Informationslogistik. 27.-28. Oktober 2008* (S. 199–216). St. Gallen, Schweiz: Gesellschaft für Informatik.
- IT Governance Institute (2003). *Board Briefing on IT Governance*. Letzter Aufruf 5. Oktober 2009 unter [http://www.itgi.org/AMTemplate.cfm?Section=Board\\_Briefing\\_on\\_IT\\_Governance&Template=/ContentManagement/ContentDisplay.cfm&ContentID=39649](http://www.itgi.org/AMTemplate.cfm?Section=Board_Briefing_on_IT_Governance&Template=/ContentManagement/ContentDisplay.cfm&ContentID=39649).
- IT Governance Institute (2007). *COBIT 4.1*. Rolling Meadows, Illinois: ISACA.
- Junginger, M. (2005). *Wertorientierte Steuerung von Risiken im Informationsmanagement*. Wiesbaden: Deutscher Universitäts Verlag.
- Kaplan, R. S., & Norton, D. P. (1997). *Balanced Scorecard: Strategien erfolgreich umsetzen*. Stuttgart: Schäffer-Poeschel.
- Kemper, H.-G., Baars, H., & Horakh, T. (2007). Business Intelligence Outsourcing: A Framework. In H. Österle, J. Schelp, & R. Winter (Hrsg.), *Proceedings of the Fifteenth European Conference on Information Systems* (S. 1155–1166). St. Gallen: University of St. Gallen.
- Kemper, H.-G., Mehanna, W., & Unger, C. (2006). *Business Intelligence - Grundlagen und praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung* (2. Auflage). Wiesbaden: Vieweg.
- Kesten, R., Müller, A., & Schröder, H. (2007). *IT-Controlling: Messung und Steuerung des Wertbeitrags der IT*. München: Vahlen.
- Knackstedt, R. (2006). *Fachkonzeptionelle Referenzmodellierung einer Managementunterstützung mit quantitativen und qualitativen Daten: Methodische Konzepte zur Konstruktion und Anwendung*. Advances in Information Systems and Management Science: Band 24. Berlin: Logos.
- Krallmann, H., Frank, H., Gronau, N. (Hrsg.) (1999). *Systemanalyse im Unternehmen: Partizipative Vorgehensmodelle, objekt- und prozeßorientierte Analysen, flexible Organisationsarchitekturen* (3. Auflage). München, Wien: Oldenbourg.

- Larson, D. & Matney, D. (o.J.). *The Four Components of BI Governance*. Letzter Aufruf 5. Oktober 2009 unter <http://www.bi-bestpractices.com/view-articles/4681>.
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), S. 251–266.
- Meise, V. (2001). *Ordnungsrahmen zur prozessorientierten Organisationsgestaltung: Modelle für das Management komplexer Reorganisationsprojekte*. Studien zur Wirtschaftsinformatik. Hamburg: Kovač.
- Mertens, P., & Meier, M. C. (2009). *Integrierte Informationsverarbeitung 2: Planungs- und Kontrollsysteme in der Industrie* (10. Auflage). Wiesbaden: Gabler.
- Miller, G. J., Bräutigam, D., & Gerlach, S. V. (2006). *Business intelligence competency centers: A team approach to maximizing competitive advantage*. Indianapolis: John Wiley & Sons.
- Object Management Group (2009). *OMG Unified Modeling Language™ (OMG UML), Infrastructure: Version 2.2*. Letzter Aufruf 5. Oktober 2009 unter <http://www.omg.org/spec/UML/2.2/Infrastructure/PDF/>.
- Oestereich, B. (2006). *Analyse und Design mit UML 2.1: Objektorientierte Softwareentwicklung* (8. Auflage). München, Wien: Oldenbourg.
- Philippi, J. (2005). Outsourcing und Offshoring von Business Intelligence-Lösungen: Empirische Studien und Praxiserfahrung. In J. Schelp & R. Winter (Hrsg.), *Auf dem Weg zur Integration factory. Proceedings der DW 2004 - Data warehousing und EAI* (S. 73–106). Heidelberg: Physica.
- Prokein, O. (2008). *IT-Risikomanagement: Identifikation Quantifizierung und wirtschaftliche Steuerung*. Wiesbaden: Gabler.
- Schlegel, K. (2007). *Toolkit Case Study: Unicredit Leverages BI to Create a New Business Model*. Gartner Report.
- Schütte, R. (1998). *Grundsätze ordnungsmäßiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsorientierter Modelle*. Neue betriebswirtschaftliche Forschung; Band 233. Wiesbaden: Gabler.
- Schwarze, L. (2006). Ausrichtung des IT-Projektportfolios an der Unternehmensstrategie. In H.-P. Fröschle & S. Strahinger (Hrsg.), *IT-Governance* (S. 49–58). Heidelberg: dpunkt.



- Seemann, J., & Gudenberg, J. W. von (2006). *Software-Entwurf mit UML 2: Objektorientierte Modellierung mit Beispielen in Java* (2. Auflage). Xpert.press. Berlin: Springer.
- Sommer, T., Bayer, W., Rosenbaum, S., Wendlandt, B., & Overmeyer, A. (2008). BI-Bebauungsplanung im Rahmen der BI-Strategie des Volkswagen-Konzerns. In B. Dinter, R. Winter, P. Chamoni, N. Gronau, & K. Turowski (Hrsg.): *Vol. P-138. GI-Edition - Lecture Notes in Informatics (LNI), DW 2008. Synergien durch Integration und Informationslogistik. 27.-28. Oktober 2008* (S. 267–286). St. Gallen, Schweiz: Gesellschaft für Informatik.
- Supply-Chain Council (Hrsg.) (2006). *Supply-Chain Operations Reference-Model Version 8.0*. Pittsburgh, Pennsylvania.
- Totok, A. (2006). Entwicklung einer Business-Intelligence-Strategie. In P. Chamoni & P. Gluchowski (Hrsg.), *Analytische Informationssysteme. Business-Intelligence-Technologien und -Anwendungen* (S. 51–70). Berlin: Springer.
- Trost, U., & Zirkel, M. (2008). BI-Governance - Königsdisziplin strategischer Business Intelligence. In B. Dinter, R. Winter, P. Chamoni, N. Gronau, & K. Turowski (Hrsg.): *Band P-138. GI-Edition - Lecture Notes in Informatics (LNI), DW 2008. Synergien durch Integration und Informationslogistik. 27.-28. Oktober 2008* (S. 39–52). St. Gallen, Schweiz: Gesellschaft für Informatik.
- Unger, C., & Kemper, H.-G. (2008). Organisatorische Rahmenbedingungen der Entwicklung und des Betriebs von Business Intelligence: Ergebnisse einer empirischen Studie. In M. Bichler, T. Hess, H. Krcmar, U. Lechner, F. Matthes, A. Picot, et al. (Hrsg.), *Multikonferenz Wirtschaftsinformatik 2008. München, 26.2.2008 - 28.2.2008, Proceedings* (S. 141–153). Berlin: GITO-Verlag.
- Vierkorn & Friedrich (2008). *Organization of Business Intelligence: The advantage gained by companies using competency centers to coordinate their business intelligence initiatives*. Letzter Aufruf 5. Oktober 2009 unter <http://www.barc.de/de/marktforschung/research-ergebnisse/organisation-von-business-intelligence.html> (Anmeldung erforderlich).
- Vinciguerra, B. M. (2004). Evaluating Business Intelligence: A Balanced Scorecard Approach. In M. Anandarajan (Hrsg.), *Business Intelligence Techniques. A Perspective from Accounting and Finance* (S. 213–226). Berlin: Springer.
- vom Brocke, J., & Buddendick, C. (2004). Konstruktionstechniken für die Referenzmodellierung: Systematisierung, Sprachgestaltung und Werkzeugunterstützung.

- In J. Becker & P. Delfmann (Hrsg.), *Referenzmodellierung. Grundlagen, Techniken und domänenbezogene Anwendung* (S. 19–49). Heidelberg: Physica.
- Watson, H. J., Fuller, C., & Ariyachandra, T. (2004). Data warehouse governance: best practices at Blue Cross and Blue Shield of North Carolina. *Decision Support Systems*, 38(3), S. 435-450.
- Weill, P., & Ross, J. W. (2004). *IT Governance: How top performers manage IT decision rights for superior results*. Boston, Mass: Harvard Business School Press.
- Wilde, T., & Hess, T. (2007). Forschungsmethoden der Wirtschaftsinformatik: Eine empirische Untersuchung. *WIRTSCHAFTSINFORMATIK*, 49(4), S. 280–287.
- Zarnekow, R., Brenner, W., & Pilgram, U. (2005). *Integriertes Informationsmanagement: Strategien und Lösungen für das Management von IT-Dienstleistungen*. Berlin: Springer.

# **Führungsinformationssysteme für Kontrollorgane - Paradigmenwechsel in der MIS-Lehre?**

Peter Mertens

Universität Erlangen-Nürnberg – Wirtschaftsinformatik I

## **1 Die neuen Probleme**

Zu den Kontrollgremien beziehungsweise Kontrollorganen (KOG) zählen unter anderem Aufsichtsräte von Aktiengesellschaften und GmbHs, Verwaltungsräte von öffentlich-rechtlichen Banken und Aufsichtsbehörden im Finanzsektor. Die Situation ist durch folgende Auffälligkeiten gekennzeichnet:

1. Die Komplexität des Überwachungsgegenstands hat stark zugenommen (Tabelle 1). Symptomatisch ist z. B., dass die Bewertung von Optionsscheinen mit einer Kombination aus Black-Scholes-Verfahren und multidimensionaler Monte-Carlo-Simulation zu den schwierigsten Aufgaben beim Einsatz von Supercomputern neuer Generation zählt oder dass man Finanzprodukte von hoch spezialisierten Mathematikern entwickeln und analysieren lässt [Bung 09], [Gerk 09]. Ein anderes Indiz ist die Vielzahl der neuen Akronyme für strukturierte Kredite (ABCP, ABS, MBS, CDO) [Post 09].
2. Viele Mitglieder dieser Gremien sind nicht aufgrund ihrer Fachkenntnisse, sondern nach politischen Gesichtspunkten ausgewählt (Tabelle 2). Zwar sind Bemühungen eingeleitet, die fachliche Eignung der Aufsichtsräte durch Maßnahmen des Gesetzgebers zu erhöhen, jedoch dürften sich dann beträchtliche Probleme einstellen, z. B. was den Eingriff in die Wahl der Arbeitnehmer-Vertreter angeht.
3. Vor allem nebenamtliche Mitglieder von KOG, aber auch solche mit vielen Mandaten, leiden unter den zunehmenden Anforderungen.
4. Die KOG sind in besonderem Maß auf die Vorstände oder ähnliche Leitungsgremien als Informationsquelle angewiesen [Gros 02, S. 243]. Stefanie Beckmann schreibt [Beck 09, S. 450]: „Die Informationsbeschaffung ist in vielen Mandatsunternehmen zu einseitig auf den Vorstand als Auskunftsperson gerichtet ... Die Vorteile der anderen Kommunikationsmedien werden zu wenig genutzt“.
5. Das früher fast selbstverständliche kooperative Verhältnis zwischen den Leitungs- und den Aufsichtsgremien, vor allem von Vorständen und Aufsichtsräten [Gros 02, S. 244], kann nicht mehr unesehen unterstellt werden, nachdem sich Meldungen des Inhalts häufen, dass Vorstände den Aufsichtsräten ungünstige Informationen, z. B. von im Rahmen des Risikomanagements von Stabsstellen geschriebene Berich-

te mit Warnhinweisen, vorenthalten oder bedenkliche Fakten zu Verlusten oder Verlustrisiken nur dosiert („scheibchenweise“) weiterleiten. Dies hat auch damit zu tun, dass manche Vorstandsmitglieder erhebliche charakterliche Defizite, wie z. B. Geldgier, erkennen lassen oder das nötige Bewusstsein für ihre gesellschaftliche Verantwortung nicht entwickelt haben [Anne 09], [Gerk 09].

6. Soweit spezielle Systeme zur Risikoberichterstattung vorhanden sind, wird die ungenügende Integration von Einzelsystemen bemängelt [Long 09]. Dadurch wird die Sicht auf das Gesamtrisiko des zu überwachenden Unternehmens („total risk exposure“), auf die es den KOG meist ankommen muss, sehr erschwert. Zwar wurde in den letzten Jahren die Value-at-Risk-Methode mathematisch-statistisch sehr verfeinert. Damit soll es z. B. gelingen, Aussagen der Art „Mit 99% Wahrscheinlichkeit verliert eine Institution in einem definierten Zeitraum höchstens 50 Mio. €“ (VaR = 50) zu treffen. Das Konzept ist neuerdings aber stark in die Kritik geraten, weil es außerordentliche Ereignisse zu wenig berücksichtigt [OV 09].

	Meldung / Zitat	Quelle
1.	„Der Aufbau der CDO's (Collateralized Debt Obligation, P. M.) wurde fortwährend verschachtelt, bald gab es die CDO der CDO der CDO der CDO. Immer weniger wussten die Banken, mit wem sie eigentlich Handel trieben.“	Bauer, W., Das Schwarze Loch, FOCUS Online vom 29.12.2008
2.	In der Citygroup wurden schon vor rund 15 Jahren Chemiker, Mathematiker und Physiker mit der Entwicklung von Finanzprodukten beauftragt, die so kompliziert waren, dass außerhalb dieser mathematisch besonders geschulten Berufsgruppen kaum jemand die Wirkungsketten verstand.	O.V., Die Zukunft des Investmentbanking, FAZ vom 05.03.2009

*Tabelle 1: Indizien der Komplexität*

	Meldung / Zitat	Quelle
1.	„Wir haben denen (Verwaltungsrat der BayernLB, P. M.) das alles vorgelegt, aber die Jungs haben das nicht verstanden. ... In den Protokollen des Verwaltungsrats heißt es über die Zusammenkünfte, auf denen zum Teil über Milliarden-Investitionen entschieden wurde: „Ohne Wortmeldung, ohne Gegenstimme“.“	Bauer, W., Das Schwarze Loch, FOCUS Online vom 29.12.2008
2.	„Ich hatte es mit Politikern zu tun, die sich scheuten, Zahlen zur Kenntnis zu nehmen ... Frei nach dem Motto: Wer sich gründlich mit Zahlen beschäftigt, wird zum Mitwisser und kann als solcher haftbar gemacht werden ... Da werden die Weltfinanzmärkte von einem Beben bislang unbekannter Stärke erschüttert (Lehman-Zusammenbruch, P. M.), und Berger (ehemaliger HSH-Vorstandsvorsitzender, P. M.) erzählt dem Kabinett, dass bei ihm ... alles in Ordnung sei. Der hatte noch nicht einmal einen Zettel dabei.	O.V., Das ist ein Wahnsystem, SPIEGEL-GESPRÄCH mit Werner Marnette, ehemaliger Wirtschaftsminister in Schleswig-Holstein, Der Spiegel Nr. 15/2009, S. 48-52

	Meldung / Zitat	Quelle
3.	Der Präsident Sanio der Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) beklagte, dass sich die Aufsichtsbehörden „nicht ausreichend mit dem Phänomen der Gier als einem der stärksten Risikotreiber überhaupt auseinandergesetzt“ hätten.	FAZ vom 29.05.2009

Tabelle 2: Qualität der Kontrollgremien

	Meldung / Zitat	Quelle
1.	„Georg Funke, Vorstandsvorsitzender der Hypo Real Estate AG, hatte aus der HRE binnen weniger Jahre ein weit verzweigtes Imperium geformt, das er nicht mehr unter Kontrolle bekam ... Mit etlichen IT-Systemen, die nicht aufeinander abgestimmt waren ... Funkes Bank, das zeigt sich nun mit aller Härte, verfügt weder über ein ausgefeiltes Risikomanagement noch über ein funktionierendes Berichtswesen.“	Papendick, U., Die Schlachtbank, Manager Magazin, Nr. 2/2009, S. 48-52
2.	Jens Nonnenmacher, Vorstand der HSH Nordbank, sagte aus: „Aus heutiger Sicht ist klar, dass keine adäquaten Risikokontrollsysteme für das ... Kreditersatzgeschäft installiert wurden.“	Im Gespräch: Dirk Jens Nonnenmacher, der Vorstandsvorsitzende der HSH Nordbank AG "Die Risikokontrolle der HSH Nordbank war unterentwickelt", FAZ vom 02.06.2009
3.	„Die Gewinne (der Bayerischen Landesbank, P. M.) sprudelten scheinbar, bis im Februar 2008 Finanzminister Huber angebliche Milliarden-Belastungen als „reine Spekulation“ bezeichnete. ... LB-Chef Werner Schmidt tritt im Februar 2008 zurück, Finanzminister Huber im Oktober.“	O.V., Geschichte vom Absturz, www.abendzeitung-nuernberg.de vom 03.03.2009
4.	„Niemand im Umfeld der Tochter (IKB, P. M.) der staatseigenen Förderbank KfW hatte geahnt, dass sich das Haus mit Investments am amerikanischen Immobilienmarkt verspekulieren würde, auch Steinbrücks Staatssekretär Jörg Asmussen als Aufsichtsrat nicht“.	Dams, J., Raufbold aus Kalkül, Welt am Sonntag vom 12.04.2009
5.	Wolfgang Porsche (AR-Vorsitzender der Porsche AG, P. M.) wurde von Deutsche-Bank-Chef Ackermann (nicht vom Vorstand, P. M.) in einem vertraulichen Gespräch auf die schwierige finanzielle Lage hingewiesen.	O. V., Wiedeking verteidigt seine Bilanz, SPIEGEL ONLINE vom 26.07.2009
6.	„Offenbar fühlte sich die Führungsspitze der Gewerkschaften über das wahre Ausmaß der AHBR-Schieflage (Allgemeine Hypothekenbank Rheinboden, gehört zur gewerkschaftlichen Beteiligungsgesellschaft BGAG, P. M.) unzureichend informiert.“	FAZ vom 23.04.2009

Tabelle 3: Informationsmängel in den KOG

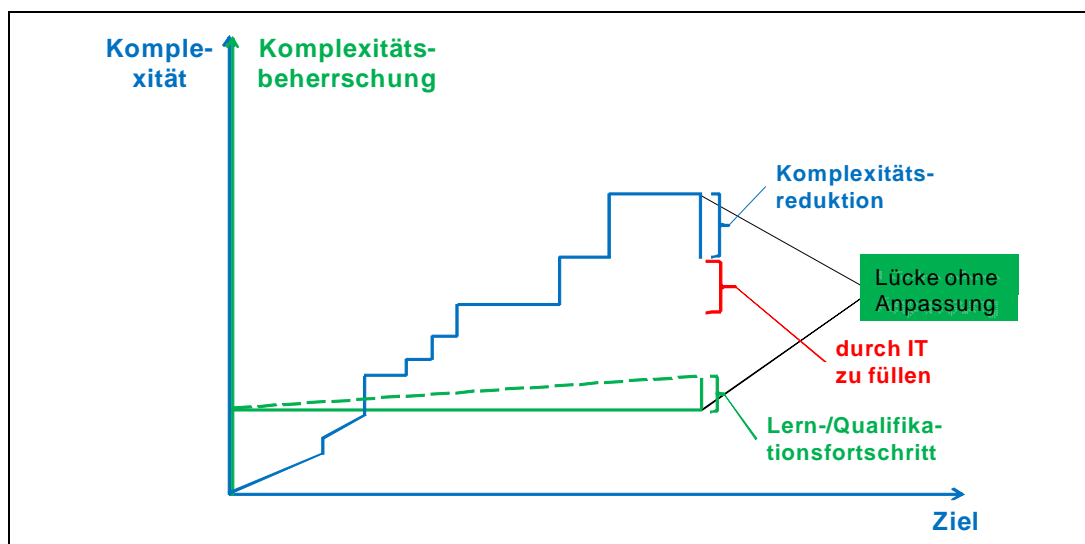
	Meldung / Zitat	Quelle
1	Nikolaus von Bomhard, Vorstandsvorsitzender der Münchener Rückversicherung AG, antwortete auf die Frage, ob viele Manager gierige Menschen seien: „Menschen jedenfalls, deren Handeln ungleich schwerer zu kalkulieren ist als etwa die Wahrscheinlichkeit eines Tsunamis in einer gewissen Region. ... Es ist bedauerlich, aber wohl auch menschlich, dass mit der Übernahme herausgehobener Positionen eine gewisse Entfernung vom normalen Leben einhergeht.“	„Es gab klare Exzesse“, SPIEGEL-GESPRÄCH, Der Spiegel Nr. 8/2009, S. 73-75

*Tabelle 4: Ethik der Mitglieder von Vorständen und KOG*

Zusammengefasst ist also eine sich öffnende Schere zwischen der wachsenden Komplexität und der Kapazität (verfügbare Zeit, Kompetenz) durch IT zu schließen (Abbildung 1).

## 2 Der notwendige Paradigmenwechsel

Die herkömmliche Lehre von den Führungsinformationssystemen geht davon aus, dass der Löwenanteil der Berichte an die obere Führungsschicht im Wege der vertikalen Integration aus den Daten der operativen Ebene verdichtet wird (siehe Abbildung 2). Dieser Prozess endet aber in der Regel bei dem obersten Führungsgremium, in Aktiengesellschaften also beim Vorstand, in GmbH's bei der Geschäftsführung usw. Nur in Ausnahmefällen verdichtet man weiter bis zum Aufsichtsrat oder zu anderen KOG. Nach einer empirischen Studie über börsennotierte Aktiengesellschaften im Jahr 2005/06 wird für die Informationsversorgung von Aufsichtsräten bisher noch wenig IT benutzt [FiBe 07, S. 53].



*Abbildung 1: Gesamtkonzeption der Integrierten Informationsverarbeitung*

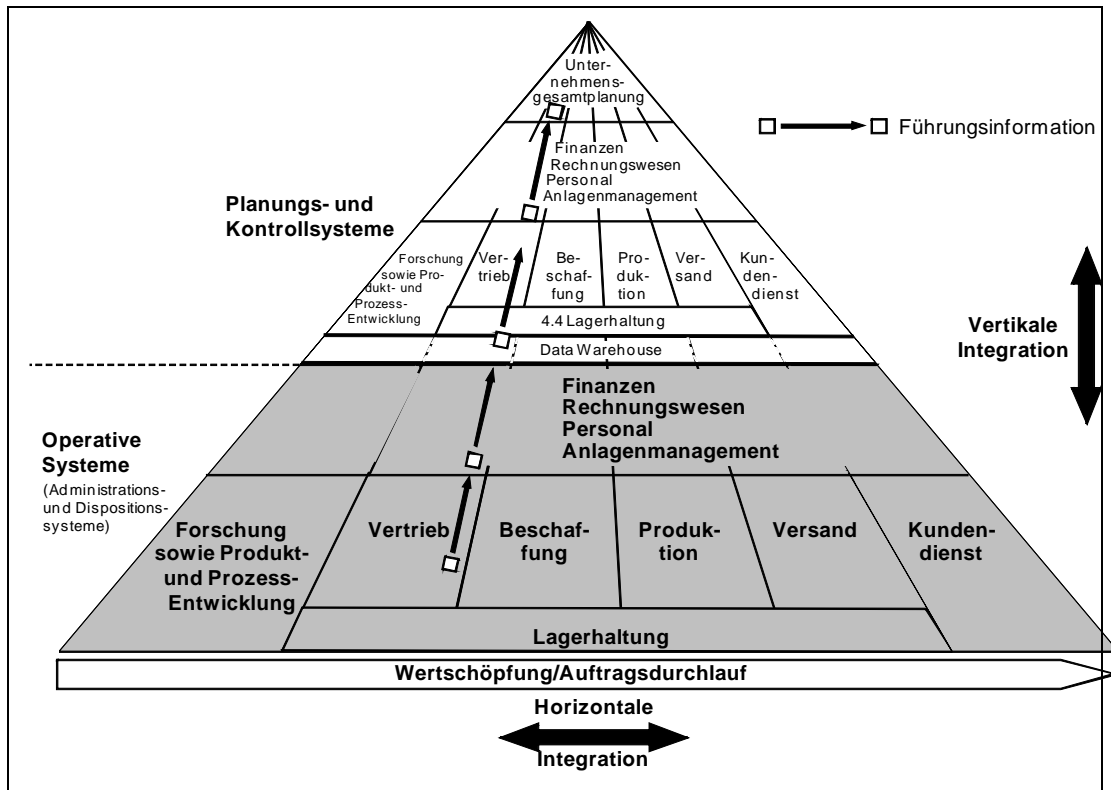


Abbildung 2: Gesamtkonzeption der Integrierten Informationsverarbeitung  
[Mert 09, S. 6]

Über ein ausgereiftes System dieser Art, welches unter anderem die Ergebnisse von Kapitalfluss-, Liquiditäts-, Ergebnis- sowie und Gewinn- und Verlustrechnungen mit der Bilanz verknüpft, verfügt der Pharmakonzern Boehringer Ingelheim GmbH. Der Abbildung 3 entnimmt man, dass die Anteilseigner an der Spitze der Pyramide stehen [MeMe 09, S. 252].

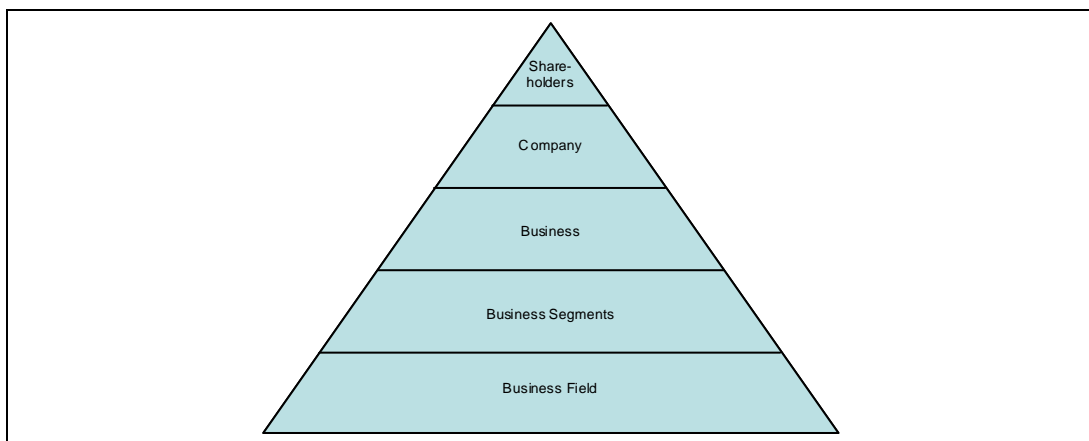


Abbildung 3: Verdichtungspyramide bei der Boehringer Ingelheim GmbH

Die tradierte Lehre von den IT-gestützten Führungsinformationssystemen beinhaltet – in aller Regel unausgesprochen – zwei Prämissen:

1. Die Empfänger der maschinell erzeugten Berichte sind in Bezug auf den Gegenstand kompetent.
2. Die Berichtslieferanten sind „ehrbare Kaufleute“.

Zumindest für Informationssysteme für Kontrollorgane (KOGIS) gelten diese beiden Voraussetzungen nicht mehr ohne Weiteres. Daher ist anzustreben:

1. Das Berichtswesen muss so weit automatisiert werden, wie es sinnhaft ist, vor allem hinsichtlich der Verdichtung. Damit wären Fach- und vor allem Führungskräften mit betrügerischen Absichten Grenzen gesetzt, weil sie in den Automatismus nicht eingreifen können. Man mag das als eine Fassade der konkreten Utopie „Sinnhafte Vollautomation“ begreifen, die als Fernziel der Wirtschaftsinformatik dienen kann [Mert 95].
2. Die von der Informatik und Wirtschaftsinformatik erarbeiteten Methoden, betriebswirtschaftliches Zahlenmaterial nicht nur „mechanisch“ zu verdichten, sondern auch die Ergebnisse automatisch zu interpretieren, sind für KOGIS zu nutzen und für diesen speziellen Zweck weiterzuentwickeln.
3. Die Psychologische Ökonomik konnte zeigen, dass in Situationen großer Unsicherheit Experten die Wahrscheinlichkeit zukünftiger Entwicklungen weniger gut einschätzen als Fachfremde [GrTv 92/Oste 09]. Daraus ist zu folgern, dass die Mitglieder der KOG, soweit sie nicht unmittelbar fachnahe sind, geeignete Kerninformationen bekommen sollten, die etwas mit allgemeinem Wirtschaftswissen oder sogar nur dem „gesunden Menschenverstand“ zu tun haben, nicht aber sehr spezielle Informationen, z. B. Resultate extrem komplexer Risikoanalysen.

### **3 Lösungsansätze im Einzelnen**

Die Wirtschaftsinformatik kann folgende Hilfe stellen:

1. Das Basismaterial auf niedriger Verdichtungsstufe lässt sich nach den Prinzipien moderner Datenverwaltung so ordnen, dass sehr flexible Auswertungen möglich sind. Insbesondere steht zu vermuten, dass sich strukturierte (hierarchische, tief gegliederte) Wertpapiere wie z. B. Dachfonds oder CDO's als Stücklisten nach dem Vorbild der Informationsverarbeitung (IV) in Fertigungsbetrieben [Mert 09] modellieren und



speichern lassen. Dann können maschinell Risikoballungen isoliert werden, z. B. eine starke und gleichzeitig stark wachsende Abhängigkeit eines Wertpapierbestands vom US-Immobilienmarkt, die den einzelnen Papieren nicht ohne Weiteres anzusehen ist. Abbildung 4 zeigt andeutungsweise, wie die Struktur eines Dachfonds dargestellt werden könnte. Wiederum in Analogie zur industriellen IV könnte man die einzelnen „Stücklistenpositionen“ um eine Art Teilestammsatz ergänzen, in dem vor allem solche Merkmale („Deskriptoren“) abgespeichert sind, welche zu Analysen beitragen. Beispielsweise geht es um Risikoklumpen in einzelnen Ländern (L) wie Argentinien, in Regionen (z. B. Golfstaaten), bei Anlagentypen (T), etwa Calls, oder in Wirtschaftszweigen bzw. Branchen (W), so z. B. in der Agrar- oder in der Immobilienwirtschaft. Eventuell wären im Datenmodell die Deskriptoren um Metainformationen zu ergänzen. Ein Beispiel sind nähere Angaben zu Länderrisiken, wo man neuerdings Kennzahlen zu nationalen Banken ins Verhältnis zum Bruttosozialprodukt setzt, um abzuschätzen, ob beziehungsweise inwieweit ein kleiner Staat (Island, Schweiz, ...) in der Lage wäre, ein großes Kreditinstitut wie die Kaupthing Bank oder die UBS zu stützen.

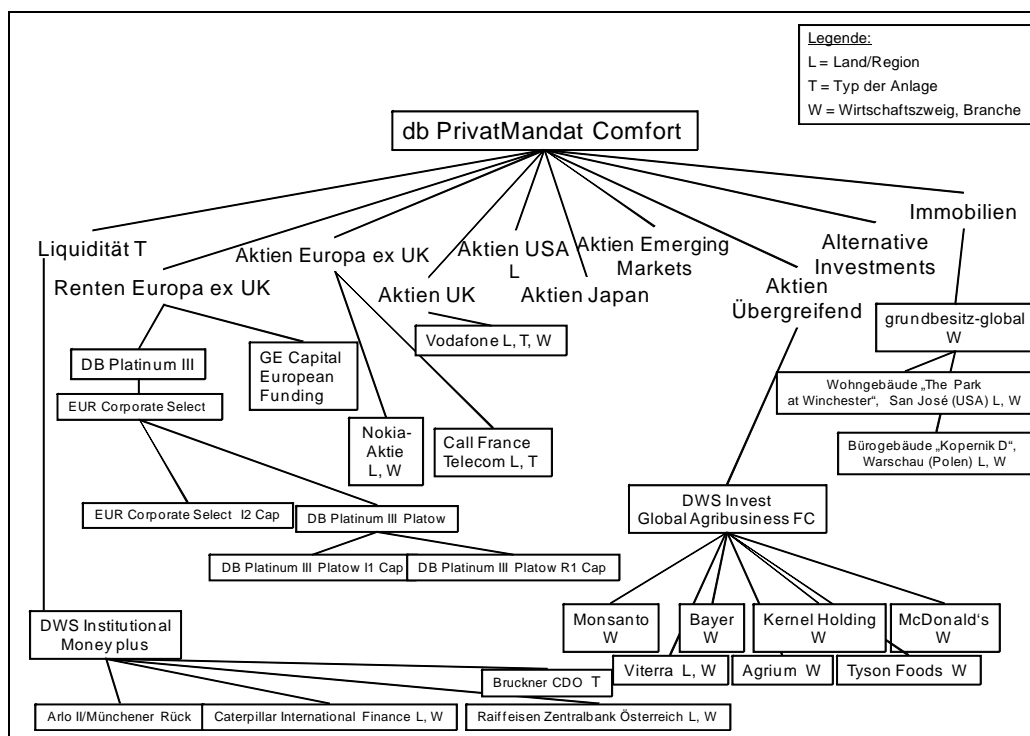


Abbildung 4: Stückliste (angedeutet)

Soweit CDO's Gegenstand der Datenmodellierung sind, ist an die Beschreibung der Tranchen, die hinsichtlich der Ausfallwahrscheinlichkeiten gemäß Ratings und der Absicherung gebildet werden (Senior (S) – Mezzanine (M) – Equity (E)) zu denken

(vgl. Abbildung 5) [BuHe 05, S. 34].

Auch Verknüpfungen zu anderen Finanzinstrumenten zur Modellierung von starken indirekten Abhängigkeiten sind angedacht [Müll 09]. Die Vermutung, dass solche Zusammenhänge zwischen zwei Finanzprodukten bestehen, mag z. B. dann auftreten, wenn - etwa durch Schrittweise Regressionsanalysen oder Adaptive Einflussgrößenkombinationen [Matt 05] - gezeigt werden kann, dass beide auf externe Einflüsse (z. B. Änderungen von Rohstoffpreisen) in der Vergangenheit gleich reagierten.

In entsprechender Weise werden die Merkmale D1, ..., D7 von Kreditverbriefungen A, B, C, D aus deren „Teilestammsätzen“ in die Tranchen von CDO's „vererbt“, die die betreffenden Kreditverbriefungen enthalten (Abbildung 5). Münden die CDO's erneut in CDO's ein, so werden die Deskriptoren nochmals aggregiert usw., sodass man jede Verbriefung nach Merkmalen beziehungsweise Merkmalsklumpen analysieren kann. Beispiele sind Ansammlungen von Krediten in Zusammenhang mit Exporten in bestimmte Länder (Risiko z. B. Änderungen von Doppelbesteuerungsabkommen) oder an Betriebe einer Branche (Risiko z. B. Überkapazitäten) oder Klumpen von Kredittypen (z. B. Konsumentenkredite). Freilich handelt es sich auch hier um eine Näherungslösung, weil sich bei mehrfacher Strukturierung Risiken, etwa durch Kompensationen, verändern können [BaBH 09, S. 135/Rudo 08, S. 727].

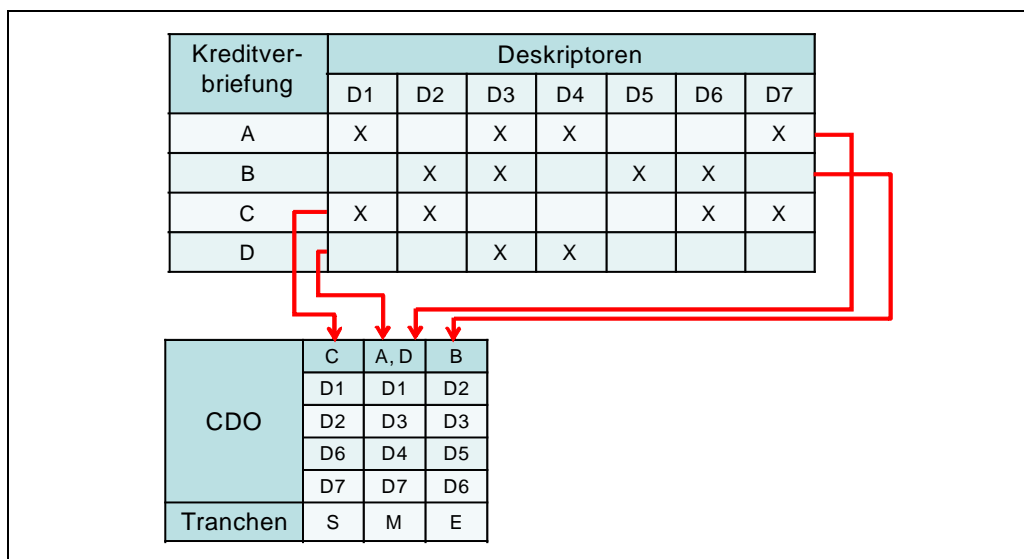


Abbildung 5: Vererbung von Deskriptoren aus Einzelkrediten in CDO-Tranchen  
(vereinfacht, Mengenbeziehungen sind nicht eingezeichnet)

Während uns die Konstruktion der Abbildung 4, also der „Stückliste“ des Dachfonds, allein auf der Grundlage weniger Unterlagen aus der Deutsche Bank AG und von Recherchen im Internet gelang, ist der Aufbau der Referenzketten von Immobilienkredi-

ten, die über mehrere Stufen aggregiert werden, nicht so einfach. Es ist jedoch vorstellbar, dass im Zuge der Bemühungen, Finanzkatastrophen in Zukunft besser vorzubeugen, den Erzeugern solcher komplizierter Finanzprodukte striktere Auflagen gemacht werden, was die Produktbeschreibungen („Beipackzettel“) betrifft, sodass die Modellierung leichter fällt. Vorbilder findet man in der Nahrungsmittel- oder Pharmabranche, etwa in Gestalt des Nachweises von gentechnischen Veränderungen. Jüngst hat die Direktbank ING-DiBa AG beschlossen, ihren Kunden solche kompakten Beschreibungen zu liefern. Ebenso beginnen die Institutionen, welche MBS ausgeben, die in den Papieren enthaltenen Positionen in „Freewriting prospects“ detailliert aufzuführen; diese sind freilich nicht so, dass sie unmittelbar maschinell abgespeichert werden könnten.

Auch die Normungsbemühungen in der Finanzdaten-Kommunikation in Gestalt der eXtensible Business Reporting Language (XBRL) könnten für KOGIS-Zwecke fruchtbar gemacht werden [Debr 09]. Insgesamt handelt es sich also um ein globales Standardisierungsproblem, auf das mit Bemühungen zu KOGIS eine Art Bedarfssog ausgeübt werden dürfte.

Ein gradueller Unterschied zur Stückliste in der industriellen Fertigungswirtschaft liegt darin, dass sich im Prinzip nach jedem Kauf oder Verkauf einer Position eines Fonds die Stückliste ändert, mit anderen Worten die Änderungsrate sehr hoch ist. Daher mag es sich empfehlen, die Darstellung für KOGIS-Zwecke nur zu bestimmten Stichtagen beziehungsweise in bestimmten Rhythmen oder bei Überschreiten von parametrisierten Schwellen anzupassen.

2. Die Informationen für die KOG werden aus den elementaren Datenspeichern im Wege der vertikalen Integration verdichtet. Es ist erwägenswert, die Berichte direkt an die KOG zu adressieren, ohne den Stabsstellen und Leitungsgremien des zu überwachenden Unternehmens die Möglichkeit zu lassen, die Berichte zu redigieren oder gar zu verfälschen. (Als Vorbild dienen die „Blackbox“ und die „Redbox“ in Flugzeugen, die nach Zwischenfällen versiegelt an die Kontrollbehörden gelangen, ohne dass die Piloten oder Fluggesellschaften Einfluss nehmen können.) So lassen sich Irreführungen ausschalten, wie sie z. B. Ronald Weckesser in seiner Eigenschaft als Verwaltungsrat der Sachsen-Landesbank beklagt: „Man kriegt ja regelmäßig den Risikobericht, da kann man nachgucken. Und da hat man so eine kleine Ampel drin mit den drei Farben rot, gelb, grün. Und wenn es grün ist, und der überwiegende Teil war eben im grünen Bereich, ... ‚Triple A‘ heißt das, sagt man, na, dann scheint es so zu gehen“ [Äußerung in den „Tagesthemen“ vom 27.08.2007]. Als Lösung des mittleren Weges bietet es sich an, in den Berichten zu kennzeichnen, welche Informationen originär-automatisch gewonnen und welche durch Zwischenstationen auf dem Weg

vom Buchungssatz zum KOG manipuliert wurden (z. B. durch Auflösung einer Rückstellung).

3. Die Fakten in Berichten können maschinell auf Auffälligkeiten (z. B. starke Veränderung in jüngerer Zeit, deutliche Veränderungen in Rangfolgen beziehungsweise „Hitlisten“, dicke und zugleich rasch wachsende Risikoklumpen) geprüft werden. Hierzu sind leistungsfähige Verfahren des Data Mining, die „Verdachtsmomente generieren“, vorstellbar.
4. Durch Methoden zur maschinellen Kommentierung (Expertisesysteme [MeMe 09], Erklärungskomponenten von Expertensystemen, Interpretationsmodule von Methodenbanken) mögen die KOG sehr gezielt auf jene Befunde hingewiesen werden, die besondere Aufmerksamkeit erfordern, dies wiederum ohne Einflussnahme durch Vorstände oder andere Instanzen des Unternehmens. Solche Analysen erlauben kompakte Aussagen, die gleichwohl gründlicher sind als die oben erwähnten Ampeldarstellungen.
5. Der sich im Rahmen der Informatik-Teildisziplin „Visualisierung“ rasch entwickelnde Wissenskörper zur automatischen Gewinnung von nicht-verfälschenden Darstellungen von Zahlenmaterial, mit dem beabsichtigte und unbeabsichtigte Suggestionen vermieden werden, trägt weiter zur sichereren Beurteilung der Unternehmenssituation durch KOG bei [BMBC 09]. Abbildung 6 zeigt eine für Mitglieder von KOG, die Frühwarnsignale in grober Form benötigen, geeignete Darstellung. Das KOGIS habe per Data Mining gefunden, dass ein Risikoklumpen entstanden ist, in dem sich ein Branchen- und ein Länderrisiko vereinen: Immobilienkredite in Großbritannien. Daraufhin hat das System die zeitliche Entwicklung des Klumpens analysiert und gefunden, dass er in den letzten zwei Jahren stark gewachsen ist.
6. In Grenzfällen ist an die Information von Kontrollorganen in Echtzeit zu denken. Es geht also um Frühwarnsysteme und um Elemente des „Echtzeit-Unternehmens“ [ScAB 03].
7. Verfahren zur maschinellen Textanalyse von Geschäftsberichten, speziell von Lageberichten über mehrere Jahre und/oder Unternehmen hinweg, werden zunehmend erforscht [Grot 09].

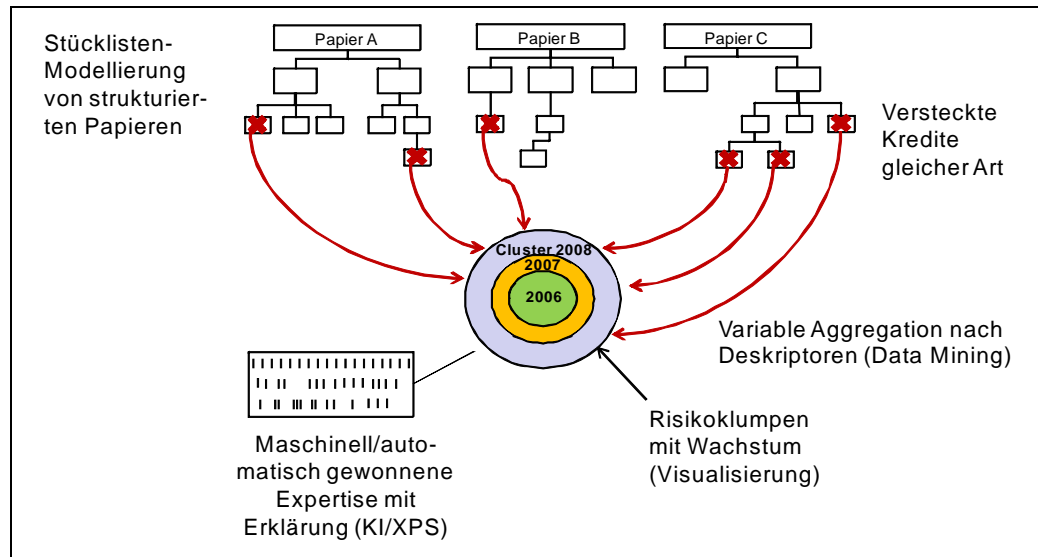


Abbildung 6: Übertragung von Ideengut aus der Fertigungsindustrie in die Finanzwelt

Abbildung 7 fasst die vorgeschlagenen Datenflüsse und Methoden(banken) zusammen.

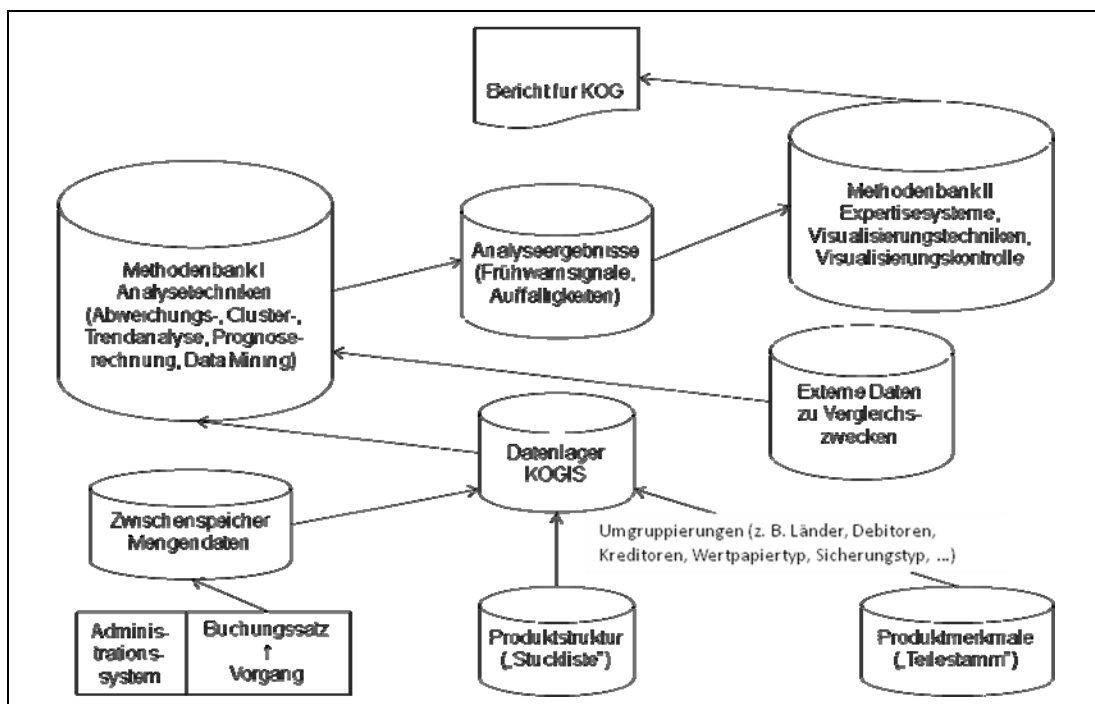


Abbildung 7: Datenflüsse und Methodenbanken

## 4 Zur rechtlichen Beurteilung

2003 schrieb der Jurist Eberhard Scheffler: „Vorstand und Aufsichtsrat haben unmittelbar miteinander zu kommunizieren. Der gesetzlich geforderte direkte Dialog ... erlaubt keinen Mittler in Form eines institutionalisierten Controlling. ... Der Aufsichtsrat darf grundsätzlich davon ausgehen, dass die vom Vorstand gegebenen Informationen zutreffend sind“ [Sche 03]. Der Aktienrechtler Marcus Lutter, Mitbegründer und Mitglied des Deutschen Corporate Government Kodex<sup>4</sup> (DCGK), sprach sich noch 2006 dafür aus, dass der Aufsichtsrat zwar die Information inhaltlich, keinesfalls aber auch in der formalen Darstellung verbindlich dem Vorstand vorgeben kann [Lutt 06, S. 41]. Dies würde der Gestaltungsfreiheit des Wirtschaftsinformatikers, die Berichte so zu visualisieren, dass Fehleinschätzungen minimiert werden (vgl. Abschnitt 3), entgegen stehen.

Wie M. R. Theisen in einer vergleichenden Untersuchung dreier rechtswissenschaftlicher Schriften darlegt [Thei 09], ist eine Entwicklung zu beobachten, die von einer stark juristisch geprägten, vor allem am § 90 Aktiengesetz festgemachten Sichtweise zu einer mehr an betriebswirtschaftlichen Notwendigkeiten orientierten führt. Die Anforderungen haben sich erhöht, was sich „sicherlich durch die verschiedensten Unternehmens- und Marktkrisen...und das dadurch auch gewandelte Menschenbild erklären lässt“ [Thei 09, S. 533]. Die ausreichende Informationsversorgung des Aufsichtsrats ist nicht mehr allein durch den Vorstand sicherzustellen („formale Berichtsautonomie des Vorstands“), vielmehr ist sie inzwischen gemeinsame Aufgabe von Vorstand und Aufsichtsrat, und der Vorstand hat nur eingeschränkte Möglichkeiten, dem Aufsichtsrat Informationen zu verweigern, d. h. auch solche, die dieser sich mit IT-Unterstützung aus unternehmensinternen und -externen Quellen beschafft. Das ist auch die Position des DCGK [Thei 09, S. 534]. Damit stimmt die Interpretation der Mission des Aufsichtsrats durch Sven H. Schneider überein: „Der moderne Aufsichtsrat ist ein Gremium institutionalisierten Misstrauens“, das folglich das Recht hat, sich „ein unabhängiges Bild von der Lage der Gesellschaft machen zu können“ [Schn 06, S. 105/107]. Auch der novellierte §107 Abs. 4 Aktiengesetz i. d. F. 2009, der die erhöhte und erweiterte Prüfungskompetenz des Aufsichtsrats fordert, spricht dafür, dass er in Eigenverantwortung Art und Umfang der rechnergestützten Kontrollsysteme bestimmen kann.

Im Ergebnis steht einem KOGIS-Konzept wie hier vorgeschlagen die Rechtslage in Deutschland wohl nicht mehr im Weg. Es darf in Anbetracht der vielen nationalen und internationalen Gesetzesinitiativen vermutet werden, dass verbleibende Zweifel von Gesetzgeber und/oder Rechtsprechung ausgeräumt würden.

## 5 Literatur

- [Anne 09] Anne, T., Die Gier war grenzenlos, Berlin 2009.
- [BaBH 09/  
Rudo 08] Bartmann, P., Buhl, H.-U. und Hertel, M., Ursachen und Auswirkungen der Subprime-Krise, Informatik Spektrum 32 (2009) 2, S. 127-144; Rudolph, B., Lehren aus den Ursachen und den Verlauf der internationalen Finanzkrise, Zeitschrift für Betriebswirtschaftliche Forschung 60 (2008) 11, S. 713-741.
- [Beck 09] Beckmann, S., Die Informationsversorgung von Mitgliedern des Aufsichtsrats börsennotierter Aktiengesellschaften, Theoretische Grundlagen und empirische Erkenntnisse, Wiesbaden 2009.
- [BMBC 09] Bissantz, N., Mertens, P., Butterwegge, G. und Christ, V., Visualisierung betriebswirtschaftlicher Daten, in: Chamoni, P. und Gluchowski, P., Analytische Informationssysteme, Wiesbaden 2009, in Druck.
- [BrEn 09] Broy, M. und Endres, A., Verantwortung der Informatikerinnen und Informatiker, Informatik Spektrum 32 (2009) 2, S. 153-162.
- [BuHe 05] Burghof, H.-P. und Henke, S., Entwicklungslinien des Marktes für Kreditderivate, in: Burghof, H.-P. u. a. (Hrsg.), Kreditderivate. Handbuch für die Bank- und Anlagepraxis, 2. Aufl., Stuttgart 2005, S. 31-52.
- [Bung 09] Persönliche Auskunft von H.-J. Bungartz.
- [Debr 09] Debrecey, R. u.a., XBRL for Interactive Data, Berlin-Heidelberg 2009.
- [FiBe 07] Fischer, T. M. und Beckmann, S., Die Informationsversorgung der Mitglieder des Aufsichtsrats – Ergebnisse einer empirischen Studie deutscher börsennotierter Aktiengesellschaften, Arbeitsbericht, Lehrstuhl für BWL, insb. Rechnungswesen und Controlling, Universität Erlangen-Nürnberg, Erlangen-Nürnberg 2007.
- [Gerk 09] Gerke, W., Reform der Bankenaufsicht, Stellungnahme für den Finanzausschuss des Deutschen Bundestages, München 2009.
- [Glas 09] Die Rolle der Informatik im gesellschaftlichen Diskurs, Informatik-Spektrum 32 (2009) 3, S. 223-227, hier S. 224.
- [Gros 02] Gros, A., Risiko Reporting an Aufsichtsräte, Osnabrück 2002.
- [Grot 09] Grottke, M., Die strukturelle Textanalyse als bilanzanalytisches Instrument zur Auswertung von Lageberichten, Die Betriebswirtschaft 69 (2009) 4, S. 463-477.
- [GrTv 92/  
[Oste 09] Griffin, D. und Tversky, A., The Weighing of Evidence and the Determinants of Confidence, Cognitive Psychology 24 (1992) 3, S. 411-435; Osterloh, M., Was hat die Finanzmarktkrise mit Hochschul-Rankings zu tun?, Zeitschrift für Organisation 78 (2009) 4, S. 231-232.

- 
- [Long 09] Longerich, D., Vertrauen zurückgewinnen (Gespräch mit Norbert Turba, Senior Executive Manager und Head of Sales Banking bei Steria Mummert Consulting), IT Director Nr. 1-2/2009, S. 23-29.
- [Lutt 06] Lutter, M., Information und Vertraulichkeit im Aufsichtsrat, 3. Aufl., Köln u.a. 2006.
- [MeMe 09] Mertens, P. und Meier, M., Integrierte Informationsverarbeitung, Band 2: Planungs- und Kontrollsysteme in der Industrie, 10. Aufl., Wiesbaden 2009.
- [Mert 95] Mertens, P., Wirtschaftsinformatik - Von den Moden zum Trend, in: König, W. (Hrsg.), Wirtschaftsinformatik '95, Wettbewerbsfähigkeit - Innovation - Wirtschaftlichkeit, Heidelberg 1995, S. 25-64.
- [Mert 09] Mertens, P., Integrierte Informationsverarbeitung, Band 1: Operative Systeme in der Industrie, 17. Aufl., Wiesbaden 2009.
- [Müll 09] Müller, G. u.a., Sichere Nutzungskontrolle: Stufenkonzept zur Steuerung der Finanzmärkte, unveröffentlichtes Manuskript, Freiburg i. Br. 2009.
- [OV 09] O.V., Trügerische Sicherheit, Wirtschaftswoche Nr. 28 vom 06.07.2009, S. 67.
- [Post 09] Posthaus, A., Strukturierte Kredite müssen aktiv verwaltet werden, FAZ vom 06.04.2009, S. 15.
- [ScAB 03] Scheer, A.-W., Abolhassan, F. und Bosch, W. (Hrsg.), Real-Time Enterprise, Mit beschleunigten Managementprozessen Zeit und Kosten sparen, Berlin-Heidelberg-New York 2003.
- [Sche 03] Scheffler, E., Controlling als Bindeglied zwischen Vorstand und Aufsichtsrat, in: Zeitschrift für Planung und Unternehmenssteuerung 14 (2003) 4, S. 399-413, hier S. 400 und 408.
- [Schn 06] Schneider, S. H., Informationspflichten und Informationseinrichtungspflichten im Aktienkonzern, Berlin 2006.
- [Thei 09] Theisen, M. R., Funktionsgerechte Informationsversorgung des Aufsichtsrats – Von der Negation über die Integration zur Affirmation betriebswirtschaftlicher Ansätze, Zeitschrift für betriebswirtschaftliche Forschung 61 (2009) 8, S. 530-549.



# **Multi-Channel-Beziehungsmanagement für den elektronischen Kaufprozess**

Andreas Hilbert, Stefan Sommer

TU Dresden – Fakultät Wirtschaftswissenschaften, Professur für Wirtschaftsinformatik,,  
insb. Informationssysteme im Dienstleistungsbereich

## **1 Einleitung**

### **1.1 Motivation**

Der Gedanke des Multi-Channel-Managements ist nicht neu und die Kundenansprache über mehrere Kanäle ist in vielen Branchen bereits Normalität. Heutzutage verlagern zunehmend die Unternehmen des stationären Handels aufgrund von stagnierenden Umsätzen im Einzelhandel und des hohen Erfolgspotentials des elektronischen Handels ihre Geschäftsmodelle in die elektronische Welt (vgl. Heinemann (2008), S. 1 ff.). Unter der Maßgabe einer ganzheitlichen Kundenorientierung sorgt das Kundenbeziehungsmanagement für eine Zusammenführung des stationären und elektronischen Handels (vgl. von der Oelsnitz (2006), S. 44; Lantzsich et al. (2002), S. 223 ff.; Schneider (2002), S. 37). Aus analytischen Gesichtspunkten werden jedoch die verfügbaren Kundeninformationen des elektronischen Handels nicht im Rahmen einer ganzheitlichen Untersuchung in Verbindung mit dem stationären Handel genutzt. Laut einer Studie der Computerwoche aus dem Jahre 2005 sehen die deutschen Unternehmen die größte Herausforderung in der unternehmensweiten Integration aller Kundenbeziehungsmanagement-Aktivitäten (vgl. Computerwoche (2005), S. 65). Weitere Studien bestätigen diesen Trend und lassen erkennen, dass für viele Unternehmen eine kanalübergreifende Einbindung des Kundenbeziehungsmanagements im selbigen eine wichtige strategische Aufgabe für die Zukunft ist (vgl. Heinemann (2008), S. 5 ff.; Goy et al. (2007), S. 488 f.; Christ et al. (2002), S. 41).

### **1.2 Forschungsansatz**

Für die technische Umsetzung des Multi-Channel-Ansatzes empfehlen Hippner et al. (2007) Customer-Relationship-Management-Systeme (CRM), die eine effiziente Konsolidierung der heterogenen Front- und Back-Office Systemlandschaft ermöglichen. Basis dieser Systeme ist eine kundenzentrierte und homogene Data-Warehouse-Architektur, die eine Analyse der operative Kunden- und Transaktionsdaten über alle Absatzkanäle hin-

weg ermöglicht und letztendlich zur Wissensgenerierung über den Kunden beiträgt (vgl. Hippner et al. (2007), S. 47; Kollmann (2007), S. 296). Vor der Konzeption eines derartigen Systems, sollten die Unterschiede des Kundenbeziehungsmanagements über verschiedene Absatzkanäle deutlich gemacht werden. Die zielgerichtete Konzeption eines CRM-Systems erfordert eine detaillierte Auseinandersetzung mit den verschiedenen Phasen der Kundenbeziehung und den Absatzkanälen im Zeitverlauf.

Im Rahmen des Multi-Channel-Managements wird in diesem Artikel eine Systematisierung für den elektronischen Kaufprozess vorgestellt, welche die Unterschiede des Kundenbeziehungsmanagement kanal- und phasenübergreifend im Zeitverlauf verdeutlicht. Wie eingangs erwähnt, kann ein ganzheitlicher Blick auf den Kunden im Sinne des analytischen Kundenbeziehungsmanagements nicht vernachlässigt werden. Vor diesem Hintergrund zeigt die Systematisierung den Übergang eines Kunden zwischen unterschiedlichen Beziehungsphasen im stationären und elektronischen Handel auf. Es werden sowohl die Möglichkeiten genannt, den Kunden in den unterschiedlichen Phasen seiner Kundenbeziehung zu analysieren. Gleichsam bildet sie aber auch die Basis für die weitere Forschung im Bereich der Methoden des analytischen Kundenbeziehungsmanagements in einem Multi-Channel-System. Dieser Artikel dient demnach der Erkenntnisgewinnung für das Kundenbeziehungsmanagement in elektronischen Kaufprozessen und liefert eine Systematisierung, die entsprechend der Konsentheorie eine Basis für die weitere Forschung in Bereich des Multi-Channel-Managements bildet (vgl. Wilde (2007), S. 281; Becker et al. (2004), S. 340 ff.).

Der erste Teil des Artikels arbeitet den aktuellen Forschungsstand auf und liefert eine Definition zum Multi-Channel-Management unter besonderer Berücksichtigung des stationären und elektronischen Handels. Ausgehend von der Klassifizierung des Kundenbeziehungsmanagements wird der Kundenlebenszyklus als ein elementarer Bestandteil für die folgende wissenschaftliche Diskussion eingeführt. Auf Basis deduktiver Herleitung wird im zweiten Teil der Kundenlebenszyklus für den elektronischen Handel in einem Multi-Channel-System vorgestellt. Der zweite Teil des Artikel schließt mit einer Systematisierung, die als Ausgangspunkt für weitere Forschung in diesem Bereich gewertet werden kann. Im dritten und letzten Teil wird die erarbeitete Systematisierung im Bereich der Personalisierung von Angebotselementen im elektronischen Kaufprozess angewendet. Es wird diskutiert, wie die Personalisierungsalgorithmen in den einzelnen Phasen des Kundenbeziehungsmanagements in einem Multi-Channel-System gestaltet werden sollen (vgl. Wilde (2007), S. 282; Becker et al. (2004), S. 347 ff.).

## 2 Grundlagen

### 2.1 Multi-Channel-Management

*Multi-Channel-Management* bezeichnet gemäß Heinemann (2008) eine Kombination von Absatzkanälen, „die ein Kunde wahlweise nutzen kann, um Leistungen eines Anbieters nachzufragen“ (Heinemann (2008), S. 15). Hierbei muss im Gegensatz zu traditionellen Mehrkanalsystemen mindestens ein Kanal des Unternehmens das *Residenzprinzip* (den stationären Einzelhandel, bspw. eine Filiale) und ein zweiter Kanal das *Distanzprinzip* (den elektronischen Handel (bspw. E-Shop) repräsentieren, wie in Abbildung 1 dargestellt.

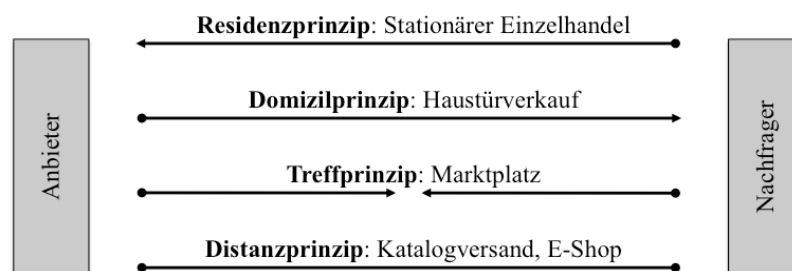


Abbildung 1: Grundlegende Kontaktprinzipien zwischen Anbieter und Nachfrager im *Multi-Channel-Management* (vgl. Wegener (2007), S. 204)

Des Weiteren muss über beide Kanäle ein klassischer bzw. elektronischer Kaufabschluss in Anlehnung an den Geschäftsmodelltyp „Commerce“ nach Wirtz (2001) möglich sein (vgl. Heinemann (2008), S. 15; Kollmann (2007), S. 40 ff.; Wirtz (2001), S. 218 ff.). Eine zusätzliche Integration weiterer Kanäle, wie bspw. des Katalogversands, ist nach Heinemann (2008) auch zulässig, wird aber mit dem Ziel der Komplexreduktion aus den folgenden Betrachtungen ausgeschlossen.

### 2.2 Kundenbeziehungsmanagement

Das *Beziehungsmanagement* bildet, als ganzheitliches Konzept nach Diller (1995), den Definitionsrahmen, in den sich weitere artverwandte Begriffe einordnen lassen, wie in Abbildung 2 dargestellt (vgl. Hippner et al. (2007), S. 19 f.). Der Begriff des *Kundenbeziehungsmanagements* leitet sich direkt aus dem Beziehungsmarketing ab und konzentriert sich ausschließlich auf die Gestaltung der Kundenbeziehungen. Wirtz (2001) definiert Kundenbeziehungsmanagement als „institutionale Analyse, Steuerung, Gestaltung und Controlling von Geschäftsbeziehungen zu den Kunden mit dem Ziel, einen unternehmerischen Erfolgsbeitrag zu leisten“ (Wirtz (2001), S. 496).

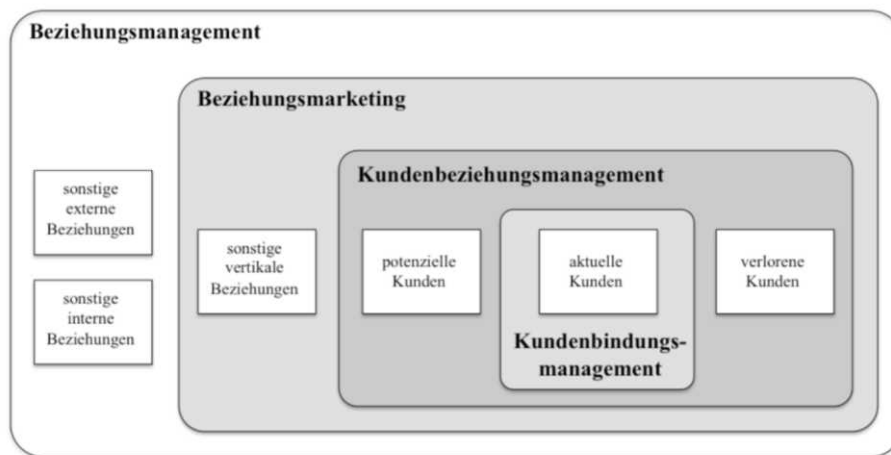


Abbildung 2: Abgrenzung des Kundenbeziehungsmanagements von verwandten Begriffen  
(vgl. Hippner et al. (2007), S. 20)

Müller (2004) erweitert die Definition von Wirtz (2001), indem er die „(...) Beeinflussung [sie!] und anschließende Kontrolle der Kundengewinnung, Kundenbindung und ggf. Beendigung der Kundenbeziehung (...)“ (Müller (2004), S. 81) explizit nennt und dadurch einen Zusammenhang mit dem Konzept des *Kundenlebenszyklus* herstellt.

### 2.3 Kundenlebenszyklus nach Stauss

Der Begriff des *Kundenlebenszyklus* basiert auf dem Begriff des *Produktlebenszyklus* und wurde maßgeblich durch Stauss (2000) geprägt. Der Kundenlebenszyklus beschreibt vor dem Hintergrund des *Paradigmenwechsels im Marketing* (vgl. Müller (2004), S. 119 ff.; Wirtz (2001), S. 497 f.) die Beziehung zwischen Kunden und Unternehmen im Verlauf der Zeit. Gemäß Muller (2004) wird das Ziel verfolgt, eine Systematisierung der Aufgaben des Kundenbeziehungsmanagements in den unterschiedlichen Phasen der Kundenbeziehung zu erreichen (vgl. Georgi (2008), S. 252; Müller (2004), S. 46).

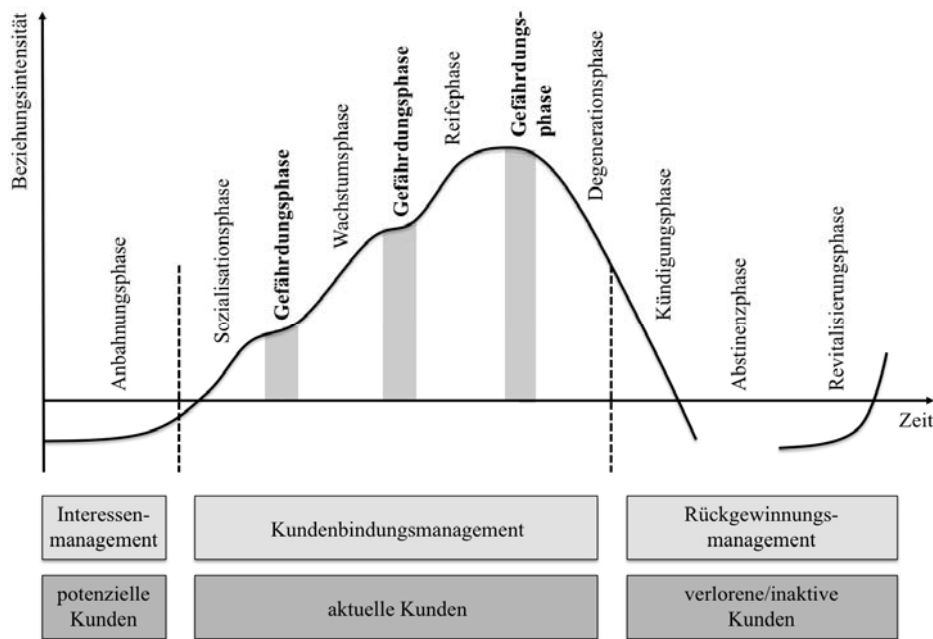


Abbildung 3: Der Kundenlebenszyklus (vgl. Stauss (2007), S. 434)

Der Kundenlebenszyklus, wie in Abbildung 3 dargestellt, umfasst drei Phasen, wobei die erste Phase, das *Interessenmanagement*, den Erstkontakt mit potenziellen Kunden kennzeichnet. Es ist das Ziel der so genannten *Anbahnungsphase*, bei potenziellen Kunden Aufmerksamkeit, sowie Interesse zu wecken und diese zum Erstkauf zu bewegen (vgl. Haas (2007), S. 446; Hippner et al. (2007), S. 37 f.; Stauss (2007), S. 435). Das *Kundenbindungsmanagement* hat die grundsätzliche Aufgabe, die Beziehungen mit aktuellen Kunden aktiv zu gestalten und zu intensivieren. Ziel ist es, den Übergang von der Kundenzufriedenheit bis zur Kundenbindung zu vollziehen, um schließlich einen signifikanten Beitrag zum ökonomischen Erfolg des Unternehmens zu leisten (vgl. Gouthier (2007), S. 487 ff.; Hippner et al. (2007), S. 37 f.; Stauss (2007), S. 433 ff.). Primäre Funktion des *Rückgewinnungsmanagements* ist es, abwanderungsgeneigte, inaktive Kunden in der *Kündigungsphase* wieder an das Unternehmen zu binden oder bereits abgewanderte, verlorene Kunden der *Abstinenzphase* zurückzugewinnen (vgl. Stauss (2006), S. 318 ff.).

### 3 Systematisierung der Kundenbeziehungen im elektronischen Kaufprozess

#### 3.1 Phasen der Kundenbeziehung

Der Kundenlebenszyklus nach Stauss (2000) strukturiert die Aufgaben des Beziehungsmanagements zwischen Kunden und Unternehmen im Zeitverlauf. Unter Bezie-

hungsgesichtspunkten können dabei die Kunden in drei unterschiedliche Kundentypen differenziert werden: *potenzielle*, *aktuelle* und *inaktive/verlorene Kunden*, wie in Abbildung 2 dargestellt. Darüber hinaus ist gemäß Stauss (2007) und weiteren Autoren eine Zuordnung der Kundentypen zu den Phasen des Kundenbeziehungsmanagements möglich (vgl. Stauss (2007), S. 435 ff.; Wimmer und Göb (2007), S. 402). Im Zuge dieser Verbindung ergeben sich für den elektronischen Kaufprozess in einem Multi-Channel-System, bestehend aus stationärem und elektronischem Handel, fünf grundlegende Phasen der Kundenbeziehung, die im Folgenden vorgestellt werden.

#### *Entstehungsphase*

Eine Person, die noch nicht Kunde eines Unternehmens über den elektronischen Absatzkanal ist oder nicht als solche identifiziert werden kann, wird in der Systematisierung entsprechend des allgemeinen Vorgehens in der Literatur als potenzieller Kunde bezeichnet. Sie befindet sich in der sogenannten *Entstehungsphase*, dessen Äquivalent im Kundenbeziehungsmanagement das Interessenmanagement bildet. In dieser Phase kann keine gesicherte Aussage getroffen werden, ob diese Person über einen der beiden Absatzkanäle bereits ein Kunde des Unternehmens ist. Des Weiteren liegen dem Unternehmen noch keine Informationen aus vorherigen Kommunikations- oder Transaktionsbeziehungen des stationären und elektronischen Handels über den Kunden vor bzw. können ihm aufgrund mangelnder aktiver oder passiver Identifikation nicht zugeordnet werden (vgl. Gaul und Schmidt-Thieme (2002), S. 46; Thurner (2002), S. 58 f.). Sobald sich der potenzielle Kunde dem Unternehmen entweder über eine erstmalige Registrierung im Zuge des elektronischen Kaufprozesses oder über andere Verfahren der Kundenidentifikation zu erkennen gibt, wird er in eine nächste Phase des Kundenlebenszyklus für den elektronischen Absatzkanal überführt, wie in Abbildung 4 gezeigt (vgl. Linder (2005), S. 140 f.; Müller (2005), S. 39.). Nutzt der potenzielle Kunde die Möglichkeit einer erstmaligen Registrierung im Zuge des elektronischen Kaufprozesses, wird er zum Neukunden des elektronischen Handels und in die *Entwicklungsphase* des Lebenszyklus überführt. Ist der potenzielle Kunde bereits Bestandskunde des elektronischen Handels und identifiziert er sich erfolgreich gegenüber dem Unternehmen, so tritt dieser den Transfer in die *Durchdringungsphase* an. Des Weiteren besteht auch die Möglichkeit, dass ein potenzieller Kunde ohne vollzogene Transaktionen den elektronische Absatzkanal wieder verlässt. In diesem Fall verbleibt dieser im Status des potenziellen Kunden, auch bei wiederholtem Besuch.

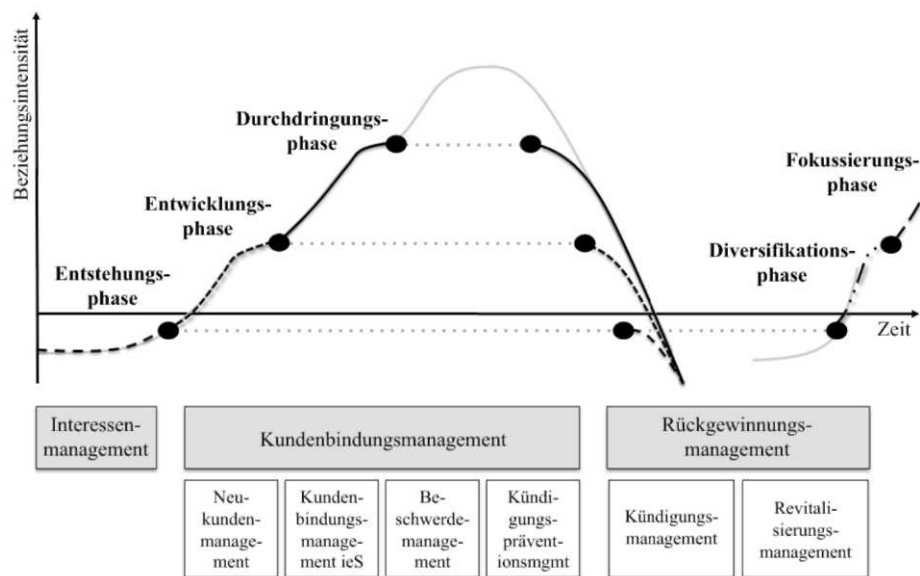


Abbildung 4: Kundenlebenszyklus für den elektronischen Kaufprozess (eigene Darstellung)

#### Entwicklungs- und Durchdringungsphase

Die Entwicklungs- und Durchdringungsphase lassen sich dem Kundenbindungsmanagement im Kundenbeziehungsmanagement zuordnen. In der Entwicklungsphase wird der Übergang eines potenziellen Kunden zum Neukunden für den elektronischen Handel vollzogen. Die im elektronischen Kaufprozess erfassten Informationen dieses Neukunden können ihm erstmalig auf Basis seiner eindeutigen Identifikation zugeordnet werden. Weiterhin ermöglicht die Identifizierung eine Verknüpfung von Informationen des stationären Handels mit seinem Kundenprofil, wie bspw. der Kaufhistorie (vgl. Blattberg et al. (2008), S. 186 ff.; Homburg und Sieben (2008), S. 505; Buxel (2001), S. 85). In der Phase der Durchdringung findet eine Weiterentwicklung des Kunden statt. Es können die anfallenden Informationen aus dem stationären wie auch dem elektronischen Handel dem Profil des Bestandskunden zugeordnet werden. Durch seine bisherigen Interaktionen im elektronischen Handel stehen weitaus mehr Informationen für Analysen zur Verfügung als in der Durchdringungsphase. Durch wiederholte Transaktionen über den elektronischen Absatzkanal vollzieht der Neukunde den Wechsel von der Entwicklungs- zur Durchdringungsphase. In beiden Phasen kann, wie auch in der Entstehungsphase, die Beziehung zwischen dem Kunden und dem Unternehmen beendet werden. Dies kann zum einen das Resultat einer lang anhaltenden Inaktivität des Kunden oder zum anderen das Ergebnis einer aktiven Kündigungshandlung einer der beiden Parteien sein. In jedem Fall wird der Kunde in den Status eines verlorenen/ inaktiven Kunden des elektronischen Handels überführt.

### *Diversifikations- und Fokussierungsphase*

Schließlich lassen sich Personen, die als Neukunden für den elektronischen Handel gewonnen werden konnten und im stationären Handel bereits als inaktive/verlorene Kunden geführt werden, in die *Diversifikationsphase* einordnen. Die Diversifikationsphase gehört ähnlich wie die Fokussierungsphase zum Rückgewinnungsmanagement des Kundenlebenszyklus. Durch die erstmalige Registrierung während des elektronischen Kaufprozesses kann dieser Kunde eindeutig identifiziert werden, und eine Zuordnung der neu erfassten Informationen des elektronischen Handels, sowie der historisierten Informationen des stationären Handels sind möglich (vgl. Müller (2005), S. 38 f.; Wirtz (2005), S. 179 f.). Durch eine zunehmende Intensivierung der elektronischen Geschäftsbeziehung des Neukunden der Diversifikationsphase mit dem Unternehmen, kann dieser analog zum Kundenbindungsmanagement als Bestandskunde in die *Fokussierungsphase* überführt werden. Die Interaktionen des Kunden beschränken sich auf elektronische Transaktionen mit dem Unternehmen. Der Übergang eines potenziellen Kunden in die Diversifikations- oder Fokussierungsphase ist nach seiner erfolgreichen Identifikation ebenfalls möglich.

## **3.2 Systematisierung**

Die fünf Phasen der Kundenbeziehungen des elektronischen Kaufprozesses zeigen auf, dass im Sinne der Hauptaufgaben des Multi-Channel-Managements eine einheitliche Sicht auf die Kunden über alle Kanäle nicht vernachlässigt werden darf. Informationen über die Kunden, die in anderen Vertriebskanälen neben dem elektronischen Handel anfallen, sollen ebenso für das analytische Kundenbeziehungsmanagement zur Verfügung stehen. Ziel ist es, im elektronischen Kaufprozess ein ganzheitliches Bild über die Kunden auf Basis sämtlicher zur Verfügung stehenden Informationen aus dem stationären und elektronischen Handel zu gewinnen.



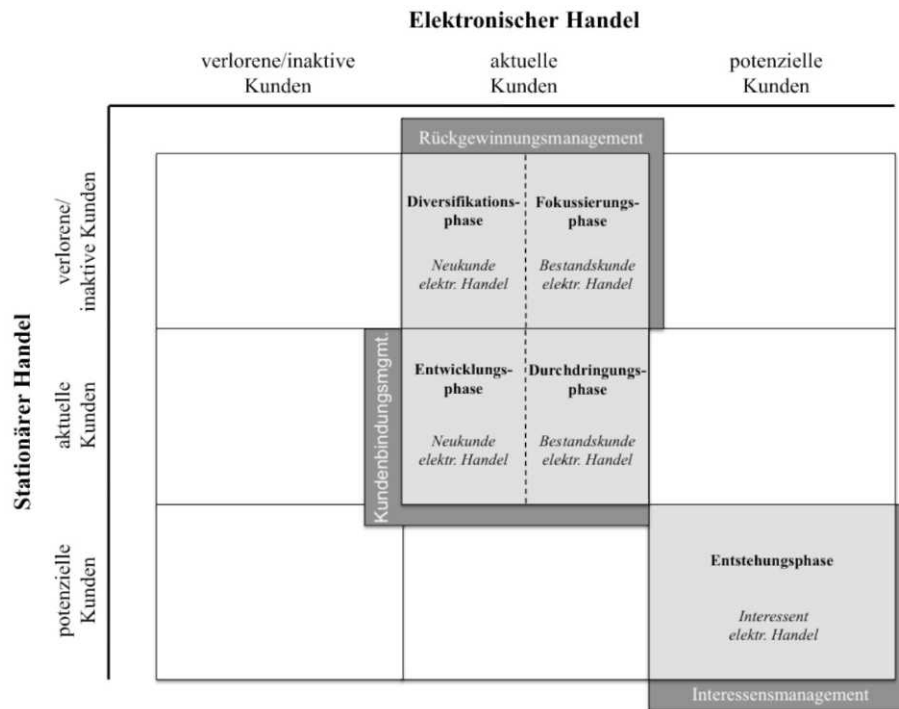


Abbildung 5: Multi-Channel-Beziehungsmanagement für den elektronischen Kaufprozess (eigene Darstellung)

Eine Zusammenfassung sämtlicher Phasen der Kundebeziehungen für den stationären und elektronischen Handel führt schließlich zu einer zweidimensionalen Systematisierung, wie in Abbildung 5 dargestellt. Die Dimensionen entsprechen den Absatzkanälen eines Unternehmens, wobei jeder Kanal in die drei grundlegende Kundentypen nach Hippner et al. (2007) unterteilt wird. Die Systematisierung beschreibt den Übergang zwischen den unterschiedlichen Beziehungsphasen für den elektronischen Kaufprozess, wie im Vorfeld erläutert. Unter Berücksichtigung der inhaltlichen Fokussierung des Artikels finden die verlorenen/inaktiven Kunden des elektronischen Handels keine Berücksichtigung in der Systematisierung und sind aus diesem Grund nicht in der Darstellung beschrieben. Die anderen nicht erfassten Felder der Systematisierung werden durch die zuvor beschriebenen Beziehungsphasen des elektronischen Kaufprozesses abgedeckt.

## 4 Personalisierte Angebote im elektronischen Kaufprozess

### 4.1 Verfahren der Personalisierung

Ziel der Personalisierung ist es, durch Kenntnis der Kundenwünsche die spezifischen Präferenzen und Bedürfnisse des Kunden kanalübergreifend im elektronischen Kaufprozess zu berücksichtigen und ein auf ihn zugeschnittenes Angebotsselement zu offerieren.

Die Literatur unterscheidet in hybride, regelbasierte, wissensbasierte, inhaltsbasierte und kollaborative Personalisierungsverfahren (vgl. Burke (2007), S. 377 ff.; Mobasher (2007), S. 93 ff.; Smith et al. (2007), S. 282 f.; Huang et al. (2004), S. 261 f.) Nach Müller (2005) und Huang et al. (2004) kommen heutzutage primär die kollaborative Verfahren im Rahmen der Personalisierung zum Einsatz, da diese Verfahren in der Regel hochwertige und Genre-übergreifende Personalisierungsmöglichkeiten gewährleisten (vgl. Müller (2005), S. 73; Huang et al. (2004), S. 261; Linden et al. (2003), S. 76 ff.). Im Praxiseinsatz hat sich gemäß Schafer et al. (2007) und weiterer Autoren bei den kollaborativen Verfahren vorwiegend der Nearest Neighbor Algorithmus durchgesetzt. Dieser unterscheidet sich in den User-based Nearest Neighbor und Item-based Nearest Neighbor Algorithmus (vgl. Schafer et al. (2007), S. 301 f.; Hwang und Tsai (2005), S. 462 ff.; Huang et al. (2004), S. 261; Linden et al. (2003), S. 76 ff.)

#### *User-based Nearest Neighbor Algorithmus*

Der nutzerbasierte Algorithmus prognostiziert gemäß Schafer et al. (2007) ein Angebotselement  $pred(u, i)$  für einen Zielkunden auf Basis einer Analyse der Ratings<sup>24</sup>  $r_{ni}$  der benachbarten Referenzkunden, wie Formel 1 zeigt. Dabei werden die Ratings der Referenzkunden, die dem Zielkunden am nächsten sind, durch das Maß der Ähnlichkeit  $userSim(u, n)$  gewichtet und auf Basis der Summe aller Ähnlichkeiten normalisiert (vgl. Schafer et al. (2007), S. 302). Um eine mögliche Streuung der Ratings zu berücksichtigen, die bspw. durch explizite Bewertungen von Angebotselementen erfolgt, fließt auch das arithmetische Mittel der Ratings der Referenzkunden  $\bar{r}_n$  in die Prognose ein (vgl. Schafer et al. (2007), S. 302).

$$pred(u, i) = \bar{r}_u + \frac{\sum_{n \in neighbors(u)} userSim(u, n) * (r_{ni} - \bar{r}_n)}{\sum_{n \in neighbors(u)} userSim(u, n)}$$

Das Maß für die Ähnlichkeit von Kunden wird nach beim User-based Nearest Neighbor Algorithmus in der Regel auf Basis des Pearson Korrelationskoeffizienten berechnet. Ergibt die Berechnung der Korrelation einen Wert von eins, so besteht eine ideale Übereinstimmung beider Kunden, hingegen liefert ein Wert von minus eins eine schlechte Korrelation (vgl. Lee und Olafsson (2009), S. 5354; Schafer et al. (2007), S. 302).

---

<sup>24</sup> Ein Rating bezeichnet eine implizit oder explizit erfasste Kundenaktion in Verbindung mit einem Angebotselement, bspw. den Kauf eines Produktes (vgl. Buxel (2001), S. 58 ff.).

*Item-based Nearest Neighbor Algorithmus*

Der Item-based Nearest Neighbor Algorithmus ist die transponierte Form der nutzerbasierten Vorgehensweise und beruht auf der Korrelationen zwischen Angebotsselementen. Die Prognose dieser für einen Zielkunden gründet sich hierbei auf die Ratings des Zielkunden selbst zu ähnlichen Angebotsselementen, wie in Formel 3 dargestellt (vgl. Schafer et al. (2007), S. 304).

$$pred(u,i) = \frac{\sum_{j \in ratedItems(u)} itemSim(i,j) * r_{ui}}{\sum_{j \in ratedItems(u)} itemSim(i,j)}$$

Die Ähnlichkeit der Angebotsselemente berechnet sich auf Basis des Kosinus-Maßes analog zum Pearson Korrelationskoeffizienten für den User-based Nearest Neighbor Algorithmus (vgl. Schafer et al. (2007), S. 304). Einen Unterschied weist lediglich die Bezugsgröße auf, die in diesem Fall nicht der Kunde, sondern das Angebotsselement ist (vgl. Schafer et al. (2007), S. 304).

*Gegenüberstellung der Algorithmen*

Sarwar et al. (2001) konstatieren, dass der User-based Nearest Neighbour Algorithmus in der Vergangenheit sehr erfolgreich war, in der Praxis aber zunehmend durch den Item-based Nearest Neighbour Algorithmus abgelöst wird. Dies liegt mit unter daran, dass der objektbasierte Algorithmus genauer die spezifischen Angebotsselemente für einen Zielkunden prognostiziert. So setzt bspw. auch Amazon gemäß Linden et al. (2003) auf den Item-based Nearest Neighbor Algorithmus (vgl. Schafer et al. (2007), S. 303 ff.; Linden et al. (2003), S. 76; Sarwar et al. (2001), S. 287 ff.).

## 4.2 Personalisierung der Angebotsselemente in verschiedenen Phasen der Kundenbeziehung

*Entstehungsphase*

Aufgrund der Tatsache, dass über den potenziellen Kunden als Referenzkunden noch keine historischen Interaktions- und Transaktionsdaten verfügbar sind bzw. diesem nicht zugeordnet werden können, erweist sich die Anwendung des Item-based Nearest Neighbor Algorithmus für die Personalisierung als schwierig. Historisierte Ratings  $r_{ui}$  des Referenzkunden sind notwendig, um eine Personalisierung der Angebotsselemente mit diesem Algorithmus zu erreichen (vgl. Schafer et al. (2007), S. 304; Huang et al. (2004), S. 261). Für eine Prognose der Angebotsselemente soll in der Entstehungsphase daher der User-based Nearest Neighbor Algorithmus eingesetzt werden.

In der Entstehungsphase fallen über den potenziellen Kunden des elektronischen Handels Weblog-Daten an, die ihm auf Basis einer Session-ID und ggf. Cookies zugeordnet werden können. Diese Weblog-Daten enthalten Informationen zu den betrachteten Angebotselementen des Interessenten im elektronischen Kaufprozess (vgl. Markov und Larose (2007), S. 148 ff.; Wiedmann und Buxel (2003), S. 13; Thurner (2002), S. 61 ff.; Harding et al. (2001), S. 17 ff.). Mittels des nutzerbasierten Algorithmus können nun diejenigen Referenzkunden  $n$  bestimmt werden, die ähnliche Angebotselemente  $r_{ni}$  in der Vergangenheit präferiert haben oder sich durch ein ähnliches Surfverhalten ausgezeichnet haben. Dabei kann auf eine konsolidierte Datenbasis des Kundenstamms im Sinne von Hippner et al. (2007) zurückgegriffen werden, die Informationen des stationären und elektronischen Handels beinhaltet (vgl. Hippner et al. (2007), S. 47 ff.). Dem potenziellen Kunden können die entsprechend bewerteten Angebotselemente der benachbarten Referenzkunden offeriert werden. Inwiefern eine Gewichtung der Informationen des stationären und elektronischen Handels erfolgen soll, kann an dieser Stelle nicht geklärt werden.

#### *Entwicklungs- und Durchdringungsphase*

Die Entwicklungs- und Durchdringungsphase sind Teil des Kundenbindungsmanagement im Bereich des Kundenbeziehungsmanagements. In beiden Phasen liegen historische Kommunikations- und Transaktionsdaten des Referenzkunden vor, aus denen Ratings über die Angebotselemente abgeleitet werden können. Folglich ist es möglich, den Item-based Nearest Neighbor Algorithmus im Rahmen der Personalisierung einzusetzen.

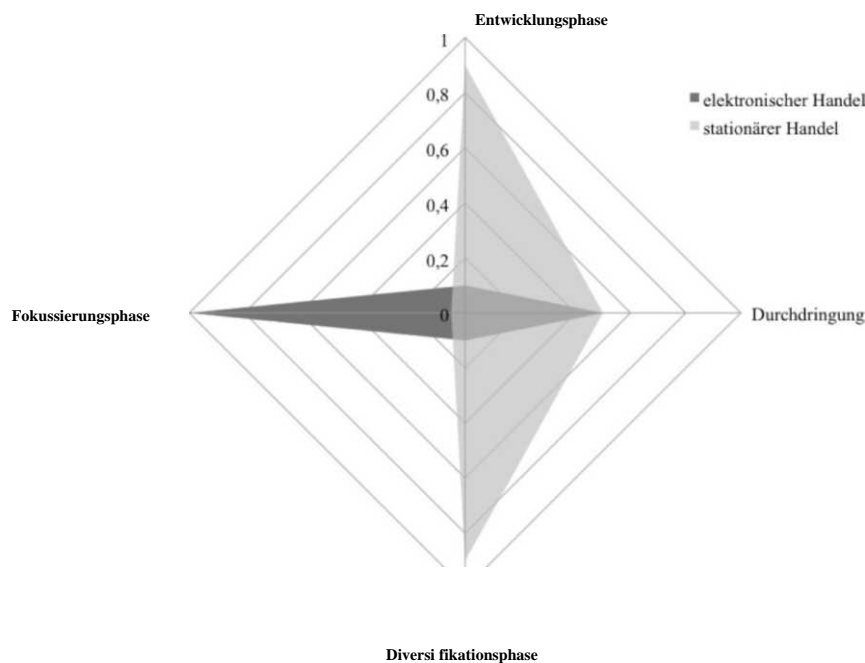


Abbildung 6: Gewichtung der Informationen aus stationärem und elektronischem Handel (eigene Darstellung)

Die Entwicklungsphase ist vom Übergang eines potenziellen Kunden zum Neukunden des elektronischen Handels gekennzeichnet. Es können dem Kunden somit die Transaktionsdaten des stationären Handels zugeordnet und in Ratings zu den Angebotsselementen  $r_{ui}$  überführt werden. Die im elektronischen Kaufprozess anfallenden Weblog-Daten geben Aufschluss darüber, welche Angebotsselemente für den Neukunden von Interesse sind, wobei die Güte der Informationen im Verlauf des Kaufprozesses zunimmt (vgl. Wiedmann und Bu-xel (2003), S. 13). Die Prognose der zu personalisierenden Angebotsselemente gründet sich demnach auf die Korrelation zwischen Angebotsselementen, an denen der Kunden im elektronischen Handel Interesse zeigt, und seinen Ratings der Angebotsselemente aus dem stationären Handel. Es empfiehlt sich nach Ansicht der Autoren, bei der Prognose der Angebotsselemente vorwiegend die historischen Informationen aus dem stationären Handel mit einzubeziehen, wie in Abbildung 6 dargestellt. Die Informationen des elektronischen Handels fallen erstmalig in Form von Weblog-Daten an, und es ist fraglich, inwieweit sie die Interessen sowie Präferenzen des Kunden ganzheitlich abbilden können. Zudem benötigt der nutzerbasierte Algorithmus des Collaborative Filtering mehrere Ratings zu Angebotsselementen, um präzise Prognosen erstellen zu können (vgl. Höhfeld und Kwiatkowski (2007), S. 267 ff.; Burke (2007), S. 377 ff.).

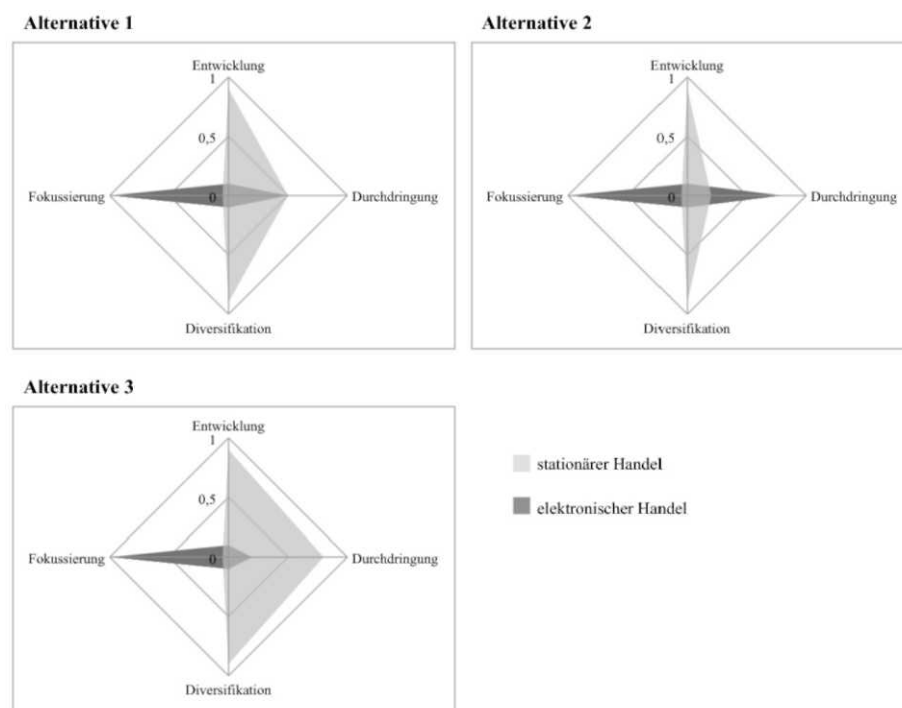


Abbildung 7: Alternativen der Gewichtung für die Durchdringungsphase (eigene Darstellung)

In der Durchdringungsphase wird der Neukunde in den Status des Bestandskunden für den elektronischen Handel überführt. Im Vergleich zur Entwicklungsphase basieren die Ratings des Zielkunden zu ähnlichen Angebotselementen auf historischen Kundeninformationen des stationären und elektronischen Handels. Bei der Gewichtung der Informationen aus beiden Absatzkanälen existieren nach Meinung der Autoren grundsätzlich drei Alternativen, wie in Abbildung 7 dargestellt. Ist das Kaufverhalten des Kunden ausgewogen bzgl. des Absatzkanals so sollten die kanalspezifischen Ratings der Angebotselemente zu gleichen Teilen in die Prognose einfließen (*Alternative 1*). Frequenziert der Kunde einen Absatzkanal deutlich stärker, so sollten die Ratings der Angebotselemente dieses Absatzkanals für die Prognose stärker gewichtet werden (*Alternative 2 und 3*).

#### *Diversifikations- und Fokussierungsphase*

Die Diversifikationsphase ist von denjenigen Kundenbeziehungen gekennzeichnet, in denen ein inaktiver/verlorener Kunden des stationären Handels aber den elektronischen Absatzkanal zurückgewonnen werden konnte. Es handelt sich um einen Neukunden des elektronischen Handels analog zur Entwicklungsphase. Jedoch können im Zuge der Personalisierung der Angebotselemente zusätzliche Informationen des stationären Handels mit einfließen, die im Rahmen der Abwanderung des Kunden erhoben wurden. Es sollten unter anderem bestimmte Abneigungen gegenüber Angebotselementen eines Zielkunden berücksichtigt werden, die bspw. aus der Unzufriedenheit über die Qualität eines Produktes zur Abwanderung beigetragen haben (vgl. Stauss (2007), S. 434 f.; Müller (2004), S. 47 f.). Es ist zu klären, inwieweit sich dadurch die Gewichtung der Ratings zu den Angebotselementen des stationären und elektronischen Handels gegenüber der Entwicklungsphase ändert. Grundsätzlich sollen nach Meinung der Autoren die Ratings in gleicher Weise gewichtet werden.

In der Fokussierungsphase intensiviert der Neukunde die Geschäftsbeziehung mit dem Unternehmen, und wird gemäß der Systematisierung als Bestandskunde des elektronischen Handels bezeichnet. Für die Personalisierung der Angebotselemente gelten die zuvor genannten Sachverhalte der Durchdringungsphase. Jedoch sollten nach Meinung der Autoren die Ratings zu den Angebotselementen aus dem elektronischen Handel deutlich stärker gewichtet werden. Im weiteren Verlauf der Kundenbeziehung sollten die Prognosen ausschließlich auf Basis der Informationen des stationären Handels erfolgen.

## **5 Fazit**

Die vorgestellte Systematisierung verdeutlicht für den elektronischen Kaufprozess eines Multi-Channel-Systems den Übergangsprozess eines Kunden zwischen unterschiedlichen

Beziehungsphasen. Sie ermöglicht dadurch der Wissenschaft und der Praxis, die Unterschiede im Umgang mit dem Kunden im Zeitverlauf zu verstehen und bildet demzufolge die Basis für die weitere Forschung zu Methoden des analytischen Kundenbeziehungsmanagements. Die Systematisierung zeigt des Weiteren, dass die Verknüpfung des stationären und elektronischen Handels insbesondere in frühen Stadien einer Kundenbeziehung für Unternehmen ein nicht zu unterschätzendes Potential birgt. Analysen über den Kunden des elektronischen Handels gründen so bereits in frühen Beziehungsphasen auf detaillierten historischen Informationen und ermöglichen dadurch bspw. genauere Aussagen und Prognosen über sein Kaufverhalten. Durch detaillierte Kenntnis der Präferenzen und Interessen eines Kunden des stationären Handels, kann eine genaue und kundenindividuelle Angebotsgestaltung sowie -vermarktung über den elektronischen Absatzkanal erfolgen, ohne dass der Kunden bereits über diesen Kanal mit dem Unternehmen interagiert haben muss (vgl. Mobasher (2007), S. 92 f.; Gohr (2006), S. 8; Müller (2005), S. 91). Die Folge ist eine erhöhte Kundenzufriedenheit, die sich wiederum positiv auf die Kundenbindung und dem damit verbundenen ökonomischen Erfolg eines Unternehmens auswirkt (vgl. Bruhn und Homburg (2008), S. 9 ff.).

Als problematisch gestaltet sich jedoch nach Einschätzung der Autoren die Gewichtung der historischen Informationen des stationären und elektronischen Handels in späteren Kundenbeziehungsphasen. Die absatzkanal-spezifische Gewichtung der Angebots Elemente des letzten Kapitels gilt es zu präzisieren und anhand realer Daten zu verifizieren. Die zukünftige Forschung hat zu klären, inwieweit bspw. die Transaktionsdaten des stationären Handels für eine personalisierte Angebotssteuerung im elektronischen Kaufprozess in späteren Kundenbindungsphasen genutzt werden können. Zudem erscheint es von besonderem Interesse zu untersuchen, wie sich die Angebotsvermarktung verändert und angepasst werden muss, wenn sich das Kaufverhalten des Kunden im elektronischen Handel maßgeblich von jenem des stationären Handels unterscheidet. Die Systematisierung sollte darüber hinaus auch auf den Katalogversand als einen weiteren Distributionskanal eines Unternehmens ausgeweitet werden.

## 6 Literatur

Becker, J., Knackstedt, R., und Niehaves, B. (2004). Epistemologische Positionierungen in der wirtschaftsinformatik am beispiel einer konsensorientierten informationsmodellierung. In Frank, U. (Hrsg.): *Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik*. 335-366. Wiesbaden. Deutscher Universitäts-Verlag.

Blattberg, R. C., Byung-Do, K., und Scott, N. A. (2008). *Database Marketing: Analyzing and Managing Customers*. 1. Auflage. New York: Springer Verlag.

Bruhn, M. und Homburg, C. (2008). Kundenbindungsmanagement - Eine Einführung in die theoretischen und praktischen Problemstellungen. In Bruhn, M. und Homburg, C. (Hrsg.): *Handbuch Kundenbindungsmanagement*. 3-37. Wiesbaden. Gabler Verlag.

- Burke, R. (2007). Hybrid Web Recommender Systems. In Brusilovsky, P., Kob-sa, A., und Nejdl, W. (Hrsg.): *The Adaptive Web*. 377-408. Berlin, Heidelberg. Springer Verlag.
- Buxel, H. (2001). *Customer Profiling im Electronic Commerce*. 1. Auflage. Aachen: Shaker Verlag.
- Christ, O., Bach, V., und Österle, H. (2002). Profitable Portale durch prozessorientierte Content- und Kanalintegration. *Thexis*. 1 (1), 41-44.
- Computerwoche (2005). CRM - Status quo und Perspektiven. *Computerwoche*. 10(1), 64.
- Gaul, W. und Schmidt-Thieme, L. (2002). Aufzeichnung des Nutzerverhaltens - Erhebungstechniken und Datenformate. In Hippner, H., Merzenich, M., und Wilde, K. D. (Hrsg.): *Handbuch Web Mining im Marketing*. 34-52. Braunschweig. Vieweg Verlag.
- Georgi, D. (2008). Kundenbindungsmanagement im Kundenbeziehungslebenszyklus. In Bruhn, M. und Homburg, C. (Hrsg.): *Handbuch Kundenbindungsmanagement*. 249-270. Wiesbaden. Gabler Verlag.
- Gohr, S. (2006). Von der Theorie zur CRM-Praxis: Augen auf und durch! *Database Marketing*. 4 (1), 4-11.
- Gouthier, M. H. J. (2007). Neukundenmanagement. In Hippner, H. und Wilde, K. D. (Hrsg.): *Grundlagen des CRM*. 473-507. Wiesbaden. Gabler Verlag.
- Goy, A., Ardissono, L., und Petrone, G. (2007). Personalization in E-Commerce Applications. In Brusilovsky, P., Kobsa, A., und Nejdl, W. (Hrsg.): *The Adaptive Web*. 485-520. Berlin, Heidelberg. Springer Verlag.
- Haas, A. (2007). Interessenmanagement. In Hippner, H. und Wilde, K. D. (Hrsg.): *Grundlagen des CRM*. 443-471. Wiesbaden. Gabler Verlag.
- Harding, W. T., Reed, A. J., und Gray, R. L. (2001). Cookies and web bugs: What they are and how they work together. *Information Systems Management*. 18 (3), 17-24.
- Heinemann, G. (2008). *Multi-Channel-Handel: Erfolgsfaktoren und Best Practices*. 2., überarb. u. erw. Aufl. Wiesbaden: Gabler Verlag.
- Hippner, H., Rentzmann, R., und Wilde, K. D. (2007). Aufbau und Funktionalitäten von CRM-Systemen. In Hippner, H. und Wilde, K. D. (Hrsg.): *Grundlagen des CRM*. 45-74. Wiesbaden. Gabler Verlag.
- Höhfeld, S. und Kwiatkowski, M. (2007). Empfehlungssysteme aus informationswissenschaftlicher Sicht - State of the Art. *Information: Wissenschaft und Praxis*. 55 (5), 265-276.
- Homburg, C. und Sieben, F. G. (2008). Customer Relationship Management (CRM) - Strategische Ausrichtung statt IT-getriebenem Aktivismus. In Bruhn, M. und Homburg, C. (Hrsg.): *Handbuch Kundenbindungsmanagement*. 501-528. Wiesbaden. Gabler Verlag.
- Huang, Z., Chung, W., und Chen, H. (2004). A Graph Model for E-Commerce Recommender Systems. *Journal of the American Society for Information Science and Technology*. 55(3), 259-274.
- Hwang, C.-S. und Tsai, P.-J. (2005). A Collaborative Recommender System Based on User Association Clusters. In Ngu, A. H. H., Kitsuregawa, M., Neuhold, E. J., Chung, J.-Y., und Sheng, Q. Z. (Hrsg.): *Web Information Systems Engineering - WISE 2005*. 462-469. Berlin, Heidelberg. Springer Verlag.



- Kollmann, T. (2007). *E-Business: Grundlagen elektronischer Geschäftsprozesse in der Net Economy*. 1. Auflage. Wiesbaden: Gabler Verlag.
- Lantzsch, K., Will, A., und Altmeyen, K.-D. (2002). Multichannel-Strategien aus Kundensicht. In Weinhardt, C. und Holtmann, C. (Hrsg.): *E-Commerce: Netze, Märkte, Technologien*. 223-238. Heidelberg. Physica Verlag.
- Lee, J.-S. und Olafsson, S. (2009). Two-way cooperative prediction for collaborative filtering recommendations. *Expert Systems with Applications*. 36(1), 5353-5361.
- Linden, G., Smith, B., und York, J. (2003). Amazon.com Recommendations. *IEEE Internet Computing*. 4 (1), 76-80.
- Linder, A. (2005). *Web Mining - Die Fallstudie Swarovski*. 1. Auflage. Wiesbaden: Deutscher Universitäts-Verlag.
- Markov, Z. und Larose, D. T. (2007). *Data Mining the Web*. 1. Auflage. New Jersey: Wiley Verlag.
- Mobasher, B. (2007). Data Mining for Web Personalization. In Brusilovsky, P., Kobsa, A., und Nejdl, W. (Hrsg.): *The Adaptive Web*. 90-135. Berlin, Heidelberg. Springer Verlag.
- Müller, H. D. (2004). *Einsatz von Customer Relationship Management-Systemen*. 1. Auflage. Wiesbaden: Deutscher Universitäts-Verlag.
- Müller, U. (2005). *Kundenbindung im E-Commerce*. 1. Auflage. Wiesbaden: Deutscher Universitäts-Verlag.
- Sarwar, B., Karypis, G., Konstan, J., und Reidl, J. (2001). Item-based collaborative filtering recommendation algorithmus. In ACM (Hrsg.): *Proceeding of the tenth international conference on World Wide Web*. 285-295. ACM Press.
- Schafer, B. J., Frankowski, D., Herlocker, J., und Shilad, S. (2007). Collaborative Filtering Recommender Systems. In Brusilovsky, P., Kobsa, A., und Nejdl, W. (Hrsg.): *The Adaptive Web*. 291-324. Berlin, Heidelberg. Springer Verlag.
- Schneider, D. (2002). Multi-Kanal-Management. In Ahlert, D., Becker, J., Knackstedt, R., und Wunderlich, M. (Hrsg.): *Customer Relationship Management im Handel*. 31-44. Heidelberg. Springer Verlag.
- Smith, M., Wenerstorm, B., Giraud-Carrier, C., Lawyer, S., und Liu, W. (2007). Personalizing E-Commerce with Data Mining. *Studies in Computational Intelligence*. 37(1), 273-286.
- Stauss, B. (2006). Beschwerdemanagement als Instrument der Kundenbindung. In Hinterhuber, H. H. und Matzler, K. (Hrsg.): *Kundenorientierte Unternehmensführung*. 315-334. Wiesbaden. Gabler Verlag.
- Stauss, B. (2007). Grundlagen und Phasen der Kundenbeziehung: Der Kundenbeziehungs-Lebenszyklus. In Hippner, H. und Wilde, K. D. (Hrsg.): *Grundlagen des CRM*. 421-442. Wiesbaden. Gabler Verlag.
- Turner, B. (2002). Einbindung von Zusatzinformationen - Nutzerregistrierung und Online-Umfragen. In Hippner, H., Merzenich, M., und Wilde, K. D. (Hrsg.): *Handbuch Web Mining im Marketing*. 54-74. Braunschweig. Vieweg Verlag.
- von der Oelsnitz, D. (2006). Der strategische Mehrkanal-Vertrieb und seine Anforderungen an das betriebliche Informationsmanagement. *Information Management und Consulting*. 21 (2), 43-48.

Wegener, M. (2007). Erfolg durch kundenorientiertes Multichannel-Management. In Reikhs, H.-C. (Hrsg.): *Retail Business in Deutschland*. 201-222.

Wiedmann, K.-P. und Buxel, H. (2003). Methodik des Customer Profiling im E-Commerce. *Marketing ZFP*. 25(1), 7-24.

Wilde, Thomas; Hess, T. (2007). Forschungsmethoden der Wirtschaftsinformatik - Eine empirische Untersuchung. *Wirtschaftsinformatik*. 49(4), 280-287.

Wimmer, F. und Göb, J. (2007). Customer Intelligence: Marktforschung und Kundenanalyse als Informationsgrundlagen im CRM. In Hippner, H. und Wilde, K. D. (Hrsg.): *Grundlagen des CRM*. 399-418. Wiesbaden. Gabler Verlag.

Wirtz, B. W. (2001). *Electronic Business*. 2., vollst. überarb. und erw. Aufl. Wiesbaden: Gabler Verlag.

Wirtz, B. W. (2005). *Integriertes Direktmarketing*. 1. Auflage. Wiesbaden: Gabler Verlag.