

Architekturkonzept eines verteilten Data-Warehouse-Systems für das Hochschulwesen

Elmar J. Sinz, Michael Böhnlein, Markus Plaha, Achim Ulbrich-vom Ende¹

¹ Univ.-Prof. Dr. Elmar J. Sinz, Dipl.-Wirtsch.Inf. Michael Böhnlein, Dipl.-Wirtsch.Inf. Markus Plaha, Dipl.-Inf. Achim Ulbrich-vom Ende,
Otto-Friedrich-Universität Bamberg,
Lehrstuhl für Wirtschaftsinformatik, insb. Systementwicklung und Datenbankanwendung,
Feldkirchenstr. 21, 96045 Bamberg, Tel.: +49 (0)951/863-2512, Fax: +49 (0)951/9370412,
E-Mail: {elmar.sinz | michael.boehnlein | markus.plaha | achim.ulbrich}@sowi.uni-bamberg.de

Architekturkonzept eines verteilten Data-Warehouse-Systems für das Hochschulwesen

Zusammenfassung:

Der vorliegende Beitrag untersucht die Beziehung zwischen (a) den Leistungsprozessen, der Organisations- und der Führungsstruktur einer verteilten Organisation und (b) der Architektur eines zugehörigen Data-Warehouse-Systems. Dabei wird deutlich, dass zentrale DWH-Systeme zur Unterstützung verteilter Organisationen nur bedingt geeignet sind. Anhand der Konzeption eines Data-Warehouse-Systems für das Hochschulwesen wird exemplarisch ein für verteilte Organisationen geeignetes Architekturkonzept eines verteilten Data-Warehouse-Systems entwickelt.

1 Einführung

Data-Warehouse-Systeme haben sich mittlerweile zu einem wichtigen Bestandteil von Führungsinformationssystemen entwickelt. Data-Warehouse-Systeme stellen entscheidungsrelevante Daten in Form einer konsolidierten und meist historisierten Datenbasis bereit. Die Daten umfassen eine Reihe von Metriken (Maßzahlen), die nach unterschiedlichen Dimensionen flexibel auswertbar sind. Ein Beispiel sind Umsätze eines Unternehmens (Metrik), die nach den Dimensionen Region, Warengruppe, Kundengruppe usw. analysiert werden können. Eine solche Datenbasis wird als Data-Warehouse (DWH), die Datenbasis zusammen mit der zugehörigen Management-Software als Data-Warehouse-System (DWH-System) bezeichnet.

DWH-Systeme lassen sich aus einem regelungsorientierten Managementverständnis heraus begründen. Danach wird zwischen dem Lenkungssystem (Regler) und dem Leistungssystem (Regelstrecke) einer Organisation unterschieden [FeSi01]. Das Lenkungssystem plant, steuert und kontrolliert die leistungserstellenden Prozesse der Organisation, dem Leistungssystem obliegt ihre Durchführung. Führungsinformationssysteme sind Anwendungssysteme zur Unterstützung des Lenkungssystems; sie dienen der Versorgung von Entscheidungsträgern mit führungsrelevanten Informationen. Ein DWH-System als Bestandteil eines Führungsinformationssystems stellt nach diesem Verständnis eine Hilfsregelstrecke im Sinne eines Modells des Leistungssystems dar.

Bei dieser Interpretation als Hilfsregelstrecke sind insbesondere zwei Merkmale eines DWH-Systems von Interesse:

- **Reichweite** des DWH-Systems: Grad der Vollständigkeit der Hilfsregelstrecke bezüglich der Leistungsprozesse der Organisation.
- **Architektur** des DWH-Systems: Berücksichtigung der Organisationsstruktur und der darauf abgestimmten Führungsstruktur der Organisation im Aufbau des DWH-Systems.

Derzeit im Einsatz befindliche DWH-Systeme weisen bezüglich beider Merkmale Defizite auf. Die Reichweite ist im Allgemeinen auf relativ eng abgegrenzte Ausschnitte von Leistungsprozessen beschränkt (z.B. Absatz- und Umsatzmetriken im Vertrieb). Entsprechend wird das Management nur punktuell mit entscheidungsrelevanten Informationen unterstützt (z.B. Vertriebs-Management). Im Hinblick auf eine umfassende Entscheidungsunterstützung ist aber eine möglichst vollständige Hilfsregelstrecke anzustreben, welche führungsrelevante Informationen aller Leistungsprozesse einschließt.

Die heute gebräuchlichen Architekturkonzepte für DWH-Systeme berücksichtigen nur sehr rudimentär die Organisations- und damit die Führungsstruktur der zu unterstützenden Organisationen. In Industrie-, Handels- und Dienstleistungsunternehmen dominiert als klassische Form die Architektur eines zentralen DWH-Systems mit z.B. abteilungsspezifischen Data-Marts als Sichten auf das zentrale DWH. Diese Form eines DWH-Systems unterstützt eine weitgehend zentralisierte und statische Organisations- und Führungsstruktur. Verteilte Organisationen mit autonomen oder teilautonomen, ggf. konkurrierenden Organisationseinheiten werden dabei ebenso wenig berücksichtigt wie Veränderungen der Organisations- und Führungsstruktur im Zeitablauf. Beispiele für solche verteilten Organisationen sind Netzwerkorganisationen (z.B. Virtuelle Unternehmen) oder Wertschöpfungsnetze von Unternehmen, die mit Hilfe eines DWH-Systems das Supply Chain Management unterstützen wollen.

Im vorliegenden Beitrag wird die Architektur eines verteilten DWH-Systems vorgestellt, welches im Rahmen des Projekts CEUS² für die bayerischen Hochschulen und das bayerische Staatsministerium für Wissenschaft, Forschung und Kunst (zusammen im Folgenden als Hochschulwesen bezeichnet) entwickelt wird (siehe auch [SiBU99]). Vergleichbare Vorhaben sind im Hochschulbereich derzeit noch selten vorzufinden. Ein Projekt mit ähnlicher Zielsetzung wie CEUS wird an der Universität Osnabrück durchgeführt [RiPo00].

² Das Projekt CEUS (Computerbasiertes Entscheidungsunterstützungssystem für Bayerische Hochschulen) ist ein Kooperationsprojekt des Bayerischen Staatsinstituts für Hochschulforschung (Prof. Dr. H.-U. Küpper) und des Lehrstuhls für Wirtschaftsinformatik, insbesondere Systementwicklung und Datenbankanwendung, der Universität Bamberg (Prof. Dr. Elmar. J. Sinz). CEUS wird im Auftrag des bayerischen Staatsministeriums für Wissenschaft, Forschung und Kunst durchgeführt. <http://ceus.uni-bamberg.de>

Das Hochschulwesen eines Landes weist wichtige Merkmale einer verteilten Organisation auf. Seine Koordination beruht auf einer Mischform aus Markt und Hierarchie. Die einzelnen Hochschulen sind weitgehend autonom, sie stehen untereinander zunehmend in Wettbewerb. Dabei konkurrieren sie u.a. um Studierende (Beschaffungsmarkt) und um die Positionierung ihrer Absolventen (Absatzmarkt). Zentrale Entscheidungen auf Landesebene betreffen insbesondere strukturpolitische Vorgaben und die Verteilung von Personal- und Sachmitteln (bei Globalhaushalt nur Sachmittel). Innerhalb einer Hochschule sind die Fakultäten bezüglich der Gestaltung und Durchführung ihrer Leistungsprozesse weitgehend autonom. Sie kooperieren untereinander im Rahmen gemeinsamer Studienangebote und Forschungsprogramme. Zentrale Entscheidungen betreffen auch hier strukturpolitische Vorgaben und die intrauniversitäre Mittelverteilung.

Primäres Ziel des vorliegenden Beitrags ist es, die bislang kaum diskutierte Beziehung zwischen den Leistungsprozessen, der Organisations- und der Führungsstruktur einer verteilten Organisation und der Architektur eines zugehörigen DWH-Systems aufzuzeigen. Hierzu werden in Kapitel 2 Anforderungen an die Architektur des DWH-Systems für das Hochschulwesen als verteilte Organisation entwickelt. In Kapitel 3 wird die Makroarchitektur, in Kapitel 4 die Mikroarchitektur eines DWH-Systems vorgestellt, welches den beschriebenen Anforderungen genügt. In Kapitel 5 folgt eine kurze Zusammenfassung und ein Ausblick auf die Übertragbarkeit der Ergebnisse.

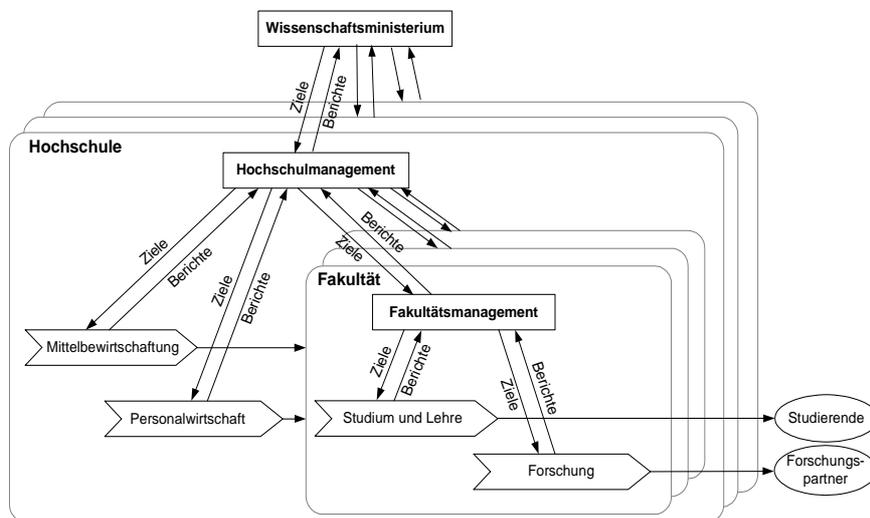


Abbildung 1: Leistungsprozesse, Organisationseinheiten und Führungsstruktur des Hochschulwesens [SiBU99]

2 Anforderungen an die Architektur eines DWH-Systems für das Hochschulwesen

Ausgangspunkt für die Entwicklung der Anforderungen an ein DWH-System für das Hochschulwesen sind die Leistungsprozesse der Hochschule, deren Zuordnung zu Organisationseinheiten und die darauf aufbauende Führungsstruktur der Hochschulen (Abbildung 1):

- **Leistungsprozesse:** Bei den Leistungsprozessen (Geschäftsprozessen) der Hochschule wird zwischen Haupt- und Serviceprozessen unterschieden [Sinz98]:

Hauptprozesse erzeugen die mit den Sachzielen der Hochschule abgestimmten Leistungen und übergeben diese an Nachfrager in der Umwelt der Hochschule. Korrespondierend zu den Sachzielen *Forschung* und *Lehre* erzeugt der Hauptprozess *Forschung* Forschungsleistungen und gibt diese an konkrete *Forschungspartner* sowie an die interessierte Öffentlichkeit ab. Der Hauptprozess *Studium und Lehre* erzeugt Ausbildungs- und Prüfungsleistungen und gibt diese an Studierende ab. Weitere Sachziele und Leistungen, wie z.B. Krankenversorgung und medizinische Dienstleistungen bei klinikführenden Hochschulen, werden im Folgenden nicht betrachtet.

Serviceprozesse erzeugen Leistungen, die von mehreren Hauptprozessen und ggf. weiteren Serviceprozessen zu deren Durchführung benötigt werden. Sie sind nur mittelbar aus den Sachzielen der Hochschule heraus begründbar. Im Hinblick auf die Konzeption eines DWH-Systems für das Hochschulwesen sind vor allem die Serviceprozesse *Personalwirtschaft* und *Mittelbewirtschaftung* von Interesse. Diese dienen der Bereitstellung der wichtigsten Ressourcen einer Hochschule, Personal und Sachmittel, und sind daher von besonderer Entscheidungsrelevanz.

- **Organisationseinheiten:** Die Hauptprozesse *Forschung* sowie *Studium und Lehre* werden von Fakultäten (Fachbereichen) durchgeführt. Die Bildung der Fakultäten folgt einer fachlichen Differenzierung dieser Leistungen. Interdisziplinäre Forschungsvorhaben oder Studiengänge werden ggf. von mehreren Fakultäten kooperativ durchgeführt.

Die Organisationseinheit *Fakultät* umfasst mehrere fachlich zusammengehörige Hauptprozesse sowie das Fakultätsmanagement. Eine *Hochschule* umfasst mehrere Fakultäten, die nur einmal pro Hochschule vorhandenen Serviceprozesse sowie das Hochschulmanagement.

- **Führungsstruktur:** Die Führungsstruktur des Hochschulwesens eines Landes ist durch eine Mischform aus Markt und Hierarchie gekennzeichnet. Ausgehend von der grundgesetzlich garantierten Freiheit von Forschung und Lehre sind die Professoren autonom in der Bestimmung ihrer Forschungs- und

Lehrinhalte. Hierarchische Koordination von Forschung und Lehre findet innerhalb der Fakultäten lediglich in Bezug auf gemeinsame Forschungsprogramme und Studiengänge statt. Die Funktion des Fakultätsmanagements wird von Dekan, Studiendekan, Fakultätsrat sowie ggf. Kommissionen und Ausschüssen ausgeübt.

Analog dazu betrifft die hierarchische Koordination auf Hochschulebene das gesamte Forschungs- und Lehrprogramm mit dem Ziel der Differenzierung gegenüber anderen Hochschulen des Landes und der Positionierung im globalen Wettbewerb zwischen den Hochschulen. Außerdem werden die Serviceprozesse meist über hierarchische Koordination geführt. Das Hochschulmanagement wird von Rektor / Präsident, Prorektoren / Vizepräsidenten, Kanzler sowie ggf. Ausschüssen und Kommissionen ausgeübt.

Auf Ebene des Landes findet eine Gesamtkoordination des Hochschulwesens statt. Das wichtigste Instrumentarium zur hierarchischen Koordination ist die zweckgebundene Verteilung von Sach- und Personalmitteln zwischen den Hochschulen und innerhalb der einzelnen Hochschulen. Mit zunehmender Einführung von Globalhaushalten reduziert sich das Instrumentarium auf eine zweckungebundene Verteilung von Sachmitteln. Die Funktion des Landes-Hochschulmanagements wird bezüglich der Exekutive vom zuständigen Ministerium ausgeübt.

Die beschriebenen Merkmale dienen nun als Ausgangspunkt für die Entwicklung der zentralen Anforderungen an ein DWH-System für das Hochschulwesen. Zunächst machen die obigen Ausführungen deutlich, dass in der Koordination der Leistungsprozesse, d.h. in der Abstimmung von Forschungs- und Lehrprogrammen einerseits und Personal- und Sachmitteleinsatz andererseits die wichtigste Führungsaufgabe auf Fakultäts-, Hochschul- und Landesebene besteht. Um die Gesamtkoordination bestmöglich zu unterstützen, sollte das DWH-System alle entscheidungsrelevanten Leistungsprozesse umfassend in Form einer Hilfsregelstrecke abbilden. Die **Reichweite** des DWH-Systems sollte daher die Hauptprozesse *Forschung, Studium und Lehre* sowie die Serviceprozesse *Mittelbewirtschaftung* und *Personalwirtschaft* abdecken und Verknüpfungen der Metriken dieser Prozesse ermöglichen.

Aus der Organisationsstruktur und der Führungsstruktur des Hochschulwesens werden nun im Weiteren wesentliche Anforderungen an die **Architektur** des DWH-Systems abgeleitet.

Hinsichtlich seiner Organisationsstruktur stellt das Hochschulwesen eines Landes ein mehrstufig verteiltes System dar. Die einzelnen Teilsysteme (Fakultäten, Hochschulen) weisen differenzierte Leistungsprofile auf und unterliegen zeitlichen Veränderungen. Bezüglich der Architektur des DWH-Systems folgt daraus:

- Das DWH-System sollte ebenfalls in Form eines verteilten Systems realisiert werden, dessen Verteilung der Organisationsstruktur folgt (Architekturmerkmal Schalenarchitektur: Abschnitt 3).
- Operative Datenquellen für das DWH-System sind in erster Linie die in den einzelnen Teilsystemen des Hochschulwesens eingesetzten Anwendungssysteme. Im Allgemeinen weist die Palette der eingesetzten Anwendungssysteme und ihrer Nutzungsformen eine erhebliche Heterogenität auf (Architekturmerkmal Interface-Layer: Abschnitt 4.1).
- Das DWH-System muss die Systemevolution des Hochschulwesens, d.h. individuelle Strukturveränderungen der einzelnen Teilsysteme im Zeitablauf, berücksichtigen. Dies erfordert ein geeignetes Konzept der Historisierung von Daten (Architekturmerkmal Base Layer: Abschnitt 4.4).

Die Führungsstruktur des Hochschulwesens stellt eine Kombination von verhandlungsbasierter (Markt) und hierarchischer Koordination (Hierarchie) dar. Die einzelnen Teilsysteme sind im Rahmen der gesetzlichen Regelungen weitgehend autonom in der Definition und Ausgestaltung ihrer Leistungsprozesse. Andererseits erfolgt insbesondere über die Zuteilung von Personal- und Sachmitteln eine hierarchische Koordination. Bezüglich der Architektur des DWH-Systems folgt daraus:

- Wegen der unterschiedlichen Informationsbedarfe auf den einzelnen Managementebenen (Land, Hochschulen, Fakultäten) muss die Architektur des DWH-Systems die Führungsstruktur des Hochschulwesens nachbilden (Architekturmerkmal Schalenarchitektur: Abschnitt 3).
- Aus der Teil-Autonomie von Fakultäten und Hochschulen folgt die Notwendigkeit, die „Privatsphäre“ der zugehörigen DWH-Teilsysteme durch Vertrauensbereiche zu schützen. Eine Übergabe von Daten von einem DWH-Teilsystem an ein anderes erfolgt ausschließlich über vertrauensbasierte Schnittstellen, deren Spezifikation beiden Beteiligten bekannt ist, und auf Basis einer zugehörigen Vereinbarung. Keinesfalls kann das DWH-Teilsystem einer höheren Managementebene ohne Zustimmung der nachgelagerten Managementebene auf deren DWH-Teilsystem zugreifen (Abschnitt 4.2).
- Die hierarchische Führungsstruktur in Verbindung mit der Heterogenität der zu führenden Teilsysteme bedingt, dass Daten bei Übernahme von einem DWH-Teilsystem in das der nächsthöheren Managementebene zu konsolidieren sind. Dieses Problem tritt insbesondere bei der Übernahme von Daten der Hochschulen in das DWH-Teilsystem des Landes auf und wird dort durch Schnittstellen-Standardisierung auf Basis der amtlichen Statistik gelöst (Architekturmerkmal Interface-Layer: Abschnitt 4).
- In den Leistungsprozessen des Hochschulwesens werden personenbezogene Daten (Studierende, Personal) verwendet, die dem Datenschutz unterliegen. Personenbezogene Daten werden ausschließlich in faktisch anonymisierter

Form in das DWH übernommen. Die Verschlüsselung der Daten erfolgt bereits innerhalb des Vertrauensbereiches des Betreibers der jeweiligen Datenquelle (Abschnitt 4.3).

Im folgenden Kapitel 3 wird nun die Gesamtarchitektur (Makroarchitektur) des verteilten DWH-Systems aus funktionaler Sicht vorgestellt, in Kapitel 4 folgt die Architektur der DWH-Teilsysteme (Mikroarchitektur) aus systemtechnischer Sicht.

3 Konzeption der Gesamtarchitektur eines verteilten DWH-Systems

Mit Bezug zu den im vorherigen Abschnitt entwickelten Anforderungen wird nun die Grobarchitektur des verteilten DWH-Systems vorgestellt. Zum besseren Verständnis erfolgt zunächst eine kurze Einführung in die grundlegende Architektur eines zentralen DWH-Systems.

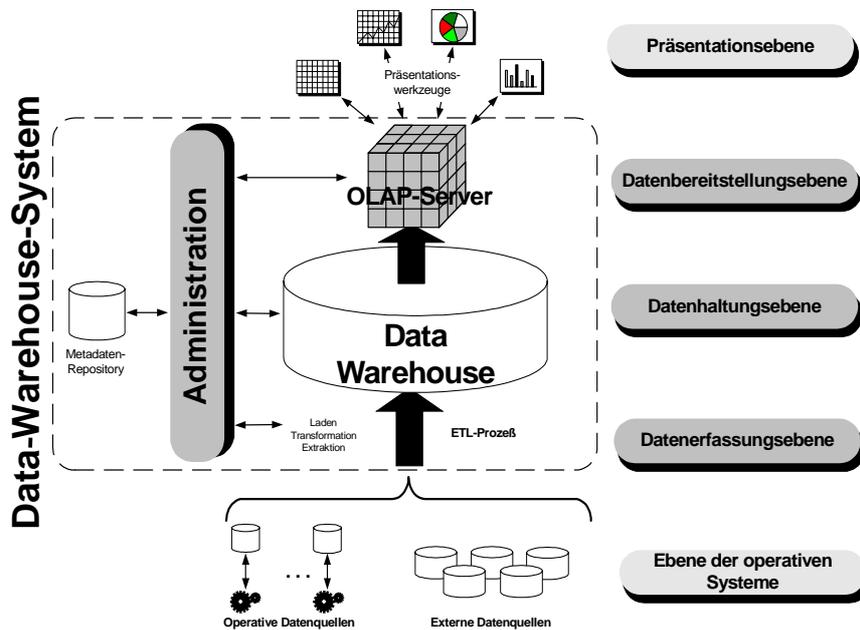


Abbildung 2: DWH-Systemarchitektur [SiBU99]

In der Literatur finden sich sehr unterschiedliche Definitionen des Begriffs Data-Warehouse. In diesem Beitrag wird analog zur Definition von *Datenbank* und *Datenbanksystem* [LoDi87] zwischen einem *Data-Warehouse* und einem *Data-Warehouse-System* unterschieden. Das *Data-Warehouse-System* umfasst sowohl das *Data-Warehouse-Management-System* mit Extraktions-, Bereinigungs-, Datenbereitstellungs- und Administrationsfunktionen als auch das eigentliche *Data-Warehouse*, die gegenüber operativen Systemen redundant gehaltene Datenbasis.

Die Architektur eines zentralen DWH-Systems ist in Abbildung 2 dargestellt. Das System läßt sich in drei Ebenen untergliedern: eine *Datenerfassungsebene* mit der Schnittstelle zu den operativen Systemen, eine *Datenhaltungsebene* mit dem eigentlichen DWH und eine *Datenbereitstellungsebene* mit den Schnittstellen zu den Präsentationswerkzeugen und anderen Anwendungssystemen. Allen Ebenen sind Administrationsfunktionen zugeordnet, die durch eine *Metadatenbank* unterstützt werden.

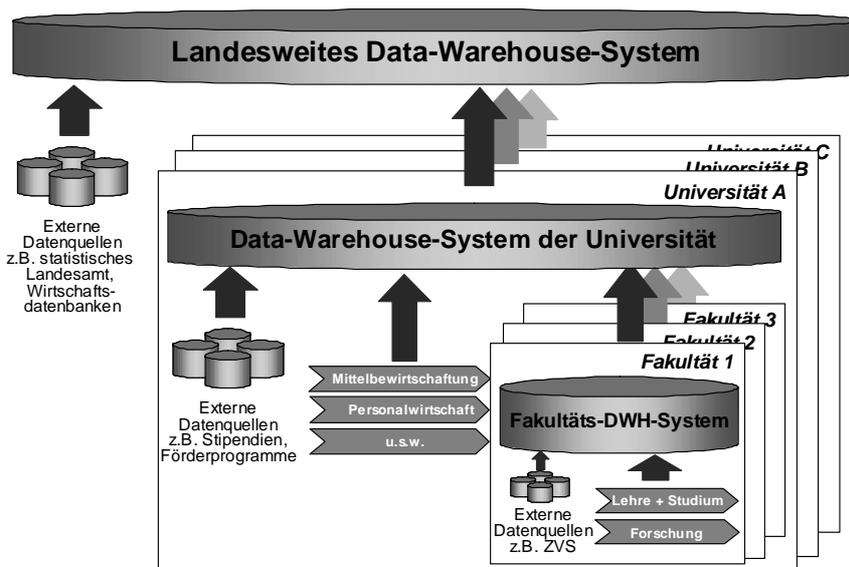


Abbildung 3: Gesamtarchitektur (Schalenmodell) des verteilten DWH-Systems aus funktionaler Sicht [SiBU99]

Aufgrund der in Abschnitt 2 genannten Anforderungen ist die zentrale Architektur für ein DWH-System des Hochschulwesens ungeeignet. Bezüglich der hierarchischen Koordination des Hochschulwesens gilt, dass der Informationsbedarf eines Managementobjekts (Ministerium, Hochschulmanagement, Fakultätsmanagement) mit dessen Entscheidungsbefugnis und –reichweite abzustimmen ist (Abstimmung von Zielen und Berichten; siehe Abbildung 1). Daneben resultiert aus der Teilau-

tonomie von Organisationseinheiten (Hochschule, Fakultät) das Erfordernis einer geschützten Privatsphäre von Daten. Diese Merkmale bleiben beim zentralen Ansatz unberücksichtigt. Daher wurde ein mehrstufiges, schalenförmiges, verteiltes DWH-System mit abgestimmter Reichweite und Aggregation der verwalteten Daten konzipiert (Abbildung 3). Die Architektur des DWH-Systems folgt der Führungsstruktur des Hochschulwesens [SiBU99]:

- Auf der untersten Ebene befinden sich die DWH-Teilsysteme der einzelnen Fakultäten einer Hochschule. Diese basieren auf den detaillierten Daten der von dieser Fakultät betriebenen Prozesse *Lehre und Studium* sowie *Forschung*.
- Die nächst höhere Ebene bildet das DWH-Teilsystem der jeweiligen Hochschule mit den detaillierten Daten der Serviceprozesse und den durch die DWH-Teilsysteme der Fakultäten – ggf. eingeschränkt oder aggregiert – bereitgestellten Daten.
- Auf der obersten Ebene werden die Daten aller Hochschulen des Landes – wiederum ggf. eingeschränkt oder aggregiert – in einheitlicher und konsolidierter Form zur Verfügung gestellt. Die Datenstrukturen dieser Ebene sind standardisiert und orientieren sich an der Struktur der amtlichen Hochschulstatistik des Statistischen Landesamts.

Gemäß dem Prinzip der Objektorientierung kapselt jedes DWH-(Teil-)System ein lokales DWH und stellt Schnittstellen zu den DWH-Systemen der über- und untergeordneten Ebenen zur Verfügung. Die DWHs einer Ebene bilden somit Quellen für das DWH der nächst höheren Ebene. Darüber hinaus können auf jeder Ebene zusätzliche externe Datenquellen einbezogen werden.

Bezogen auf die Datenunabhängigkeit zwischen den Teilsystemen benachbarter Ebenen besitzt diese Architektur folgende Eigenschaften:

- **Inhaltliche Datenunabhängigkeit:** Die Weitergabe von Daten an das (Teil-)DWH der nächst höheren Ebene erfolgt über vereinbarte Schnittstellen. Diese sind den beteiligten Managementebenen bekannt und bilden die Grundlage für deren Vertrauensbeziehung (Abschnitt 4.2). Zur Erfüllung spezifischer Informationsbedarfe auf den jeweiligen Managementebenen können auf jeder Warehouse-Ebene zusätzliche externe Datenquellen eingebunden werden.
- **Strukturelle Datenunabhängigkeit:** Auf den einzelnen Ebenen müssen die Daten ggf. nach unterschiedlichen Strukturmerkmalen gegliedert werden. Beispielsweise ist für das landesweite DWH eine einheitliche Datenstruktur über alle Universitäten notwendig. Aufgrund der strukturellen Unabhängigkeit des universitären Warehouses können trotzdem dort universitätsspezifische Details berücksichtigt werden, die dem landesweiten Warehouse nicht zur Verfügung stehen.

- **Zeitliche Datenunabhängigkeit:** Die Aktualisierung der einzelnen (Teil-) DWHs kann zu unterschiedlichen Zeitpunkten erfolgen. So können die DWHs der Fakultäten in kürzeren Zeitabständen aktualisiert werden als das der Hochschule.

Nach diesem Überblick über die Gesamtarchitektur des DWH-Systems für das Hochschulwesen wird im folgenden Abschnitt die Architektur der DWH-Teilsysteme aus systemtechnischer Sicht behandelt.

4 Die Architektur eines DWH-Teilsystems

Die Architektur eines einzelnen DWH-Teilsystems ist in Abbildung 4 dargestellt. Nach einer kurzen Einführung in die detaillierte Architektur werden konkrete Problemfelder und zugehörige Lösungsansätze behandelt. Gegenüber der in Abbildung 3 dargestellten Makroarchitektur verdeutlicht Abbildung 4 die zugehörige Mikroarchitektur. Dabei werden auf der Datenhaltungsebene folgende Schichten unterschieden:

- **Interface-Layer:** Diese Schicht realisiert die Schnittstelle zwischen einem DWH-Teilsystem und den operativen Systemen bzw. einem untergeordneten DWH-Teilsystem. Der Interface-Layer stellt einen temporären Datenspeicher zum Laden des DWH dar. Nach der Extraktion der Daten aus den operativen Datenquellen werden diese dort bereinigt, ggf. transformiert und anschließend an den Base-Layer weitergereicht. In der Literatur finden sich für den Begriff „Interface-Layer“ häufig die Bezeichnungen Staging-Area [Kurz99] oder Update-Tables [TeU197].
- **Base-Layer:** Der Base-Layer bildet die bereinigte und historisierte Datenbasis des gesamten DWH-Teilsystems. Diese konsistente Datenhaltungsschicht abstrahiert von der eigentlichen Datenhaltung für Analysezwecke (Analytical-Layer). Das zugrundeliegende Datenschema ist weder für die spezifischen Nutzeranforderungen, noch für das verwendete Analysewerkzeug angepasst. Häufig wird für diese Schicht synonym auch der Begriff „Operational-Data-Store“ ([Inmo99], [Dres00], [JLVV00]) verwendet.
- **Analytical-Layer:** Diese Schicht stellt die Daten in der für das verwendete Analysewerkzeug benötigten Form zur Verfügung. Beim Einsatz eines Auswertungswerkzeugs, das auf multidimensionalem Online-Analytical-Processing (MOLAP) [Cham98] basiert, handelt es sich bei der Analytical-Layer um dessen spezifische Datenhaltung in einer vom eigentlichen DWH unabhängigen, multidimensionalen Datenbank. Im Gegensatz dazu bildet der Analytical-Layer bei relationalem OLAP (ROLAP) [Cham98] eine zusätzliche Schicht im eigentlichen DWH. In diesem Fall liegt der Datenhaltung häu-

fig das Star- oder Snowflake-Schema [Rade96] zugrunde. Im Weiteren wird von einer ROLAP-Lösung ausgegangen.

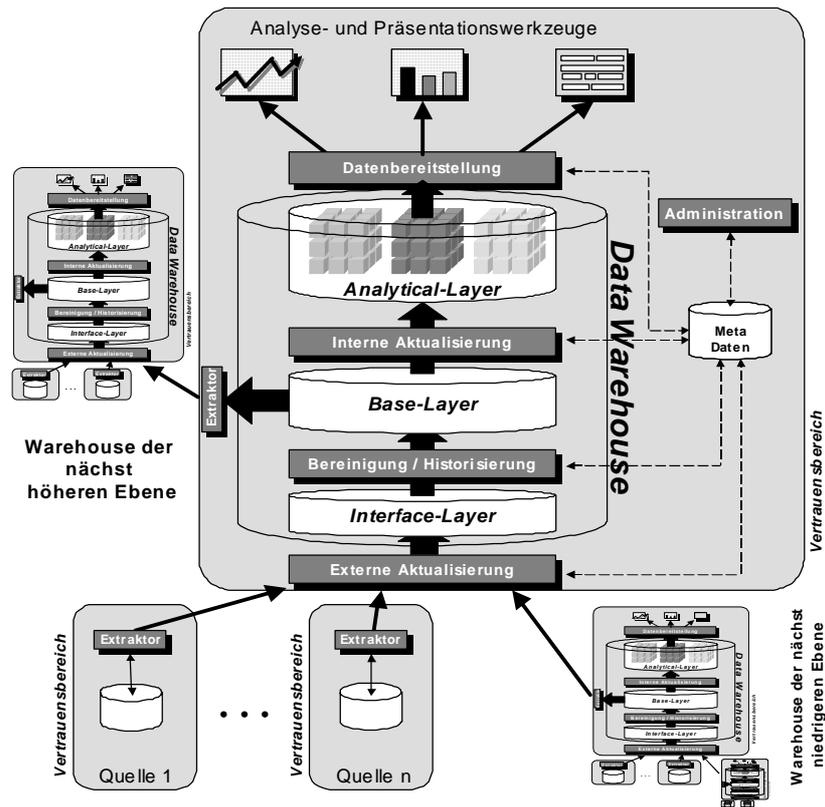


Abbildung 4: Architektur eines DWH-Teilsystems

Durch diese Einteilung können folgende Funktionen beim Laden der Daten explizit differenziert werden:

- **Externe Aktualisierung:** In diesem Kontext wird unter der externen Aktualisierung die Übernahme der Daten aus operativen Systemen in den Interface-Layer verstanden. Hierzu werden mit Hilfe einer Extraktor-Komponente die relevanten Daten aus den Quellen extrahiert und in das DWH übernommen.
- **Bereinigung / Historisierung:** Nachdem die Daten aus den Quellsystemen extrahiert und in den Interface-Layer geladen wurden, müssen sie in einem nächsten Schritt bereinigt und transformiert werden. Dies ist notwendig, um die Korrektheit der Daten im DWH sicherzustellen und somit die Korrektheit der auf diesen Daten aufbauenden Analysen zu gewährleisten [ChDa97]. Zu-

sätzlich werden beim Übergang zwischen Interface- und Base-Layer die Daten historisiert, indem die nichttemporalen Daten des Interface-Layers beim Übergang in den Base-Layer mit Zeitstempeln versehen werden [YaWi98].

- **Interne Aktualisierung:** Entsprechend den nutzerspezifischen Anforderungen bezüglich der Auswahl der Attribute, deren Verdichtung und Zeitbezug werden die Daten während der internen Aktualisierung in den Analytical-Layer übertragen. Die explizite Trennung zwischen interner und externer Aktualisierung ermöglicht einen weitgehend unterbrechungsfreien Betrieb für den Nutzer. Während des Ladens des Base-Layers können die Nutzer weiterhin auf die konsistenten Daten im Analytical-Layer zugreifen [TeUI98]. Darüber hinaus ist eine leichtere Anpassung des Systems an wechselnde Nutzeranforderungen möglich. Diesem Vorteil steht ein höherer Speicherbedarf gegenüber, der in diesem konkreten Fall jedoch vernachlässigt werden kann.

In den folgenden Abschnitten wird auf die Umsetzung der in Kapitel 2 vorgestellten Anforderungen genauer eingegangen.

4.1 Heterogenität der Datenquellen

Eines der Hauptmerkmale von DWH-Systemen liegt in der konsolidierten Zusammenführung von Daten aus unterschiedlichen operativen Anwendungssystemen (entspricht dem Merkmal der Integration nach Inmon [Inmo96]). Sowohl syntaktische als auch semantische Inkonsistenzen zwischen den Quellen müssen im DWH beseitigt werden. Dies erfordert ein einheitliches Datenschema, in das Daten aus operativen Systemen transformiert werden.

Bei einer zentralisierten DWH-Architektur müssen die unterschiedlichen Informationsbedürfnisse der einzelnen Managementebenen (Land, Hochschulen, Fakultäten) in einem Gesamtschema integriert werden. Aufgrund der Teil-Autonomie von Hochschulen und Fakultäten und der daraus resultierenden Notwendigkeit einer „Privatsphäre“ ist eine zentralisierte Architektur für ein DWH-System für das Hochschulwesen ungeeignet. Im Gegensatz dazu steht bei einem verteilten DWH-System, das sich an der Führungsstruktur orientiert, für jedes Managementobjekt ein separates Warehouse zur Verfügung. Dadurch können die individuellen Informationsbedürfnisse leichter berücksichtigt werden.

Die Heterogenität der im Hochschulwesen eingesetzten Anwendungssysteme und deren Nutzungsformen erfordern eine individuelle Entwicklung der DWH-Teilsysteme. Um den damit verbundenen Aufwand zu reduzieren, wird zunächst ein standardisierter DWH-Kern als Basis für alle Hochschulen realisiert, der sich an die individuellen Bedürfnisse der einzelnen Hochschulen anpassen lässt. Dabei dient der Interface-Layer als Puffer zwischen der Datenhaltung in den operativen Systemen und dem DWH und definiert somit eine einheitliche Schnittstelle für die Datenübernahme. Das Schema dieser Schicht orientiert sich an den Informations-

bedarfen zur Entscheidungsunterstützung in der jeweiligen Anwendungsdomäne (*Forschung, Studium und Lehre* sowie *Mittelbewirtschaftung* und *Personalwirtschaft*) und nicht am Inhalt der verwendeten operativen Quellsysteme.

4.2 Autonomie der Teilsysteme

Klassische DWH-Systeme verwenden zur Datenlieferung das Pull-Prinzip. Dies bedeutet, dass ein DWH-System seine Daten selbst aus den operativen Datenquellen extrahiert.

Im Gegensatz dazu unterliegt die Datenlieferung in der vorgestellten verteilten Architektur dem Push-Prinzip. Um den Vertrauensbereich (Abbildung 4) der operativen Systeme und der DWH-Teilsysteme zu wahren, wird die Datenextraktion vom Betreiber des betroffenen Systems durchgeführt. Die Übergabe der Daten an die nächst höhere Ebene erfolgt über klar definierte Schnittstellen, die zwischen den beteiligten Managementobjekten vereinbart werden. Beispielsweise basiert die Schnittstelle zur Übernahme der Daten aus den einzelnen Hochschulen in das landesweite DWH auf den gesetzlichen Bestimmungen des Hochschulrahmengesetzes (HRG) und des Hochschulstatistikgesetzes (HStatG).

Die Betreiber der Datenquellen besitzen somit stets die Kontrolle über die Daten, die an das DWH-System der nächst höheren Ebene übergeben werden. Dadurch werden der jeweiligen Managementebene nur bestimmte Informationen in einer klar definierten Aggregationsstufe zur Verfügung gestellt. Ein Zugriff auf detaillierte Informationen (Disaggregation) andererseits ist ohne explizite Einwilligung des geführten Objekts nicht möglich. Ein befürchteter Durchgriff auf DWHs untergeordneter Ebenen ist durch die verteilte Architektur des DWH-Systems ausgeschlossen. Dadurch bleibt die Autonomie der Teilsysteme gewahrt. Dies ist eine der wichtigsten Voraussetzungen für die Akzeptanz eines DWH-Systems für das Hochschulwesen.

4.3 Datenschutz

In einem DWH-System ist der Datenschutzproblematik besondere Aufmerksamkeit zu widmen, da die dort angebotenen vielfältigen Navigationsmöglichkeiten in einem meist äußerst umfangreichen Datenbestand zur nahezu beliebigen Kombination von Einzeldaten herangezogen werden können.

Der Datenschutz stellt den Persönlichkeitsschutz von Individuen in den Mittelpunkt. Neben dem Bundesdatenschutz (BDSG) regeln länderspezifische Gesetze, wie z.B. das Bayerische Datenschutzgesetz (BayDSG), den Umgang mit personenbezogenen Daten im Rahmen der elektronischen Datenverarbeitung: „Zweck dieses Gesetzes ist es, die einzelnen davor zu schützen, dass sie bei der Erhebung, Verarbeitung oder Nutzung ihrer personenbezogenen Daten durch öffentliche

Stellen in unzulässiger Weise in ihrem Persönlichkeitsrecht beeinträchtigt werden.“ (BayDSG §1)

Um eine datenschutzkonforme Anwendung zu erreichen, muss wenigstens eine der folgenden Voraussetzungen erfüllt sein [Burk00]. Es muss entweder ein legitimer Zweck und die Einwilligung der Betroffenen für die Übernahme von personenbezogenen Daten vorliegen oder es ist der Personenbezug zu eliminieren. Da die DWH-Daten zur Entscheidungsunterstützung im Management und nicht im Rahmen der operativen Durchführung der Leistungsprozesse verwendet werden, ist eine Bezugnahme auf einzelne Personen im Allgemeinen nicht erforderlich. Um den Anforderungen des Datenschutzes zu genügen, muss der Personenbezug im DWH eliminiert werden. Dies ist durch eine faktische Anonymisierung gewährleistet. Hierbei werden alle personenbezogenen Attribute (Name, Vorname, Adresse usw.) bereits bei der Übernahme der Daten noch innerhalb des Vertrauensbereichs des jeweiligen operativen Systems eliminiert. Für die Durchführung von Konsistenzprüfungen werden beim Laden des Warehouse jedoch zusätzlich eindeutige Merkmale benötigt (z.B. Matrikelnummer). Um hierbei der faktischen Anonymisierung genüge zu tun, werden diese Merkmale mit Hilfe eines Public-Key Verfahrens auf der Basis des RSA-Algorithmus nach Rivest, Shamir und Adleman [Selk00] verschlüsselt. Dadurch ist ein Rückschluss auf die Person im Verhältnis zum Wert der erlangten Information zu aufwendig [Moen98]. Zudem werden diese Merkmale nach Durchführung der Konsistenzprüfungen beim Laden des Warehouse eliminiert, wodurch eine missbräuchliche Verwendung verhindert wird.

4.4 Strukturdynamik

Ein wichtiges Merkmal von DWH-Systemen ist ihr Zeitraumbezug [Inmo96], d.h. DWHs speichern historisierte Daten, um Trends im Zeitverlauf erkennen und angemessen darauf reagieren zu können. Im Rahmen der Historisierung stellt vor allem die Strukturdynamik das zentrale Problem dar. Hierbei müssen strukturelle Änderungen im Hochschulwesen durch adäquate Anpassungen in der verteilten DWH-Systemarchitektur berücksichtigt werden.

Vor allem sich im zeitlichen Verlauf ändernde Klassifikationsstrukturen sind von diesem Problem betroffen. Bereits KIMBAL [Kimb96] weist im Rahmen von „slowly changing dimensions“ auf Dimensionen hin, die logisch von der Zeitdimension abhängig sind.

Beispielsweise kann sich die Zuordnung von Studienfächern zu Fakultäten im Zeitablauf ändern. Auch ein Wegfall, eine Umwidmung oder das Neuentstehen von Studienfächern sind im Hochschulwesen nicht selten. Abhängig von den nutzerspezifischen Anforderungen kann dieses Problem auf zwei unterschiedliche Weisen behandelt werden:

- Die bestehende Klassifikationsstruktur wird mit den neuen Werten überschrieben. Eine Auswertung der vergangenen Daten unter den Gesichtspunkten der ehemaligen Klassifikationsstruktur ist dadurch nicht mehr möglich. (entspricht Typ I nach Kimball [Kimb96]).
- Durch Versionierung können gleichzeitig mehrere Klassifikationsstrukturen für unterschiedliche Zeitbereiche existieren. Statt die bisherige Struktur zu überschreiben, wird beim Laden der Daten eine zusätzliche Version mit den neuen Strukturdaten generiert. (entspricht Typ II nach Kimball [Kimb96]).

Abhängig von den Anforderungen ist eine dieser beiden Alternativen zu wählen. Beispielsweise wird beim Geschlecht eines Studenten von einer statischen Struktur ohne Versionierung ausgegangen, da unabhängig vom zeitlichen Verlauf lediglich die drei Ausprägungen *weiblich*, *männlich* und *UNBEKANNT* existieren können.

Im Folgenden wird nur auf die zweite Alternative eingegangen, da diese zur Behandlung der Strukturndynamik beitragen kann. Hierzu findet sich in Abbildung 5 ein veranschaulichendes Beispiel.

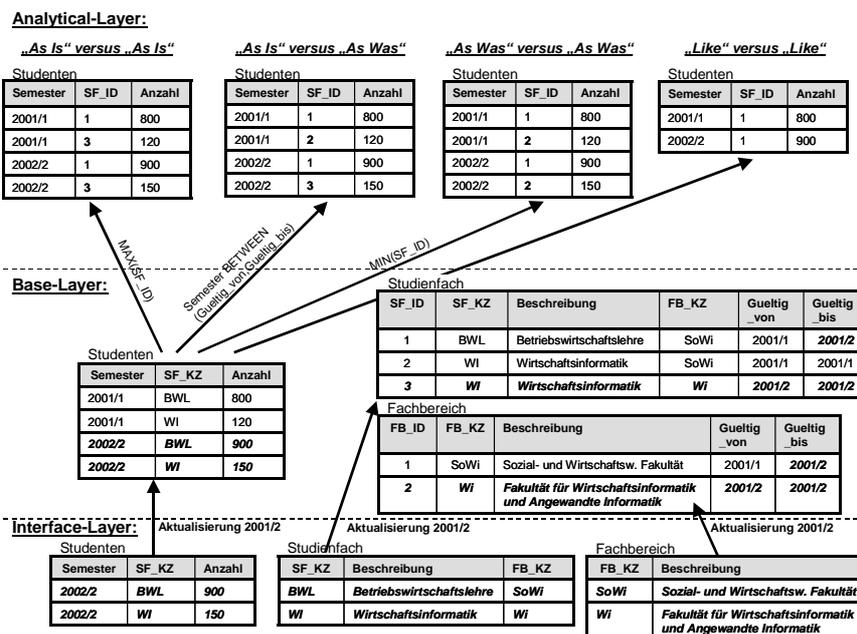


Abbildung 5: Versionierte Aktualisierung

In einer Hochschule erfolgt die Neugründung einer *Fakultät für Wirtschaftsinformatik und Angewandte Informatik* (Semester 2001/2). Ein bereits existierendes

Studienfach *Wirtschaftsinformatik* wird von der *Sozial- und Wirtschaftswissenschaftlichen Fakultät* (Semester 2001/1) der neuen Fakultät zugeordnet (Semester 2001/2). Dies führt zu einer neuen Version des Studienfachs *Wirtschaftsinformatik*.

In einer relationalen Datenbank werden zur Verwaltung mehrerer Versionen zwei zusätzliche Attribute benötigt. Zur Identifikation des Gültigkeitszeitraums eines Wertes werden die beiden Zeitstempel *Gueltig_von* und *Gueltig_bis* verwendet (s. auch [Snod99]). Darüber hinaus wird hier zusätzlich zum ursprünglich identifizierenden Attribut (*Fakultaet_KZ* und *Studienfach_KZ*) ein künstlicher Schlüssel (Surrogatschlüssel) eingeführt (*Fakultaet_ID* und *Studienfach_ID*). Dieser ermöglicht eine leichtere Anpassung des Analytical-Layers an das eingesetzte OLAP-Werkzeug.

Entsprechend den temporalen Anforderungen an die Auswertungsmöglichkeiten ergeben sich auf der Basis eines versionierten Base-Layers vier Alternativen für die Konstruktion der Analytical-Layer (Abbildung 5) [BaGu01, S. 188]:

- Die Daten aller Semester werden entsprechend der aktuellen Klassifizierung strukturiert. Das bedeutet, dass das Studienfach *Wirtschaftsinformatik* auch für die vergangenen Semester der neuen *Fakultät für Wirtschaftsinformatik und Angewandte Informatik* zugeordnet wird, obwohl diese zu diesem Zeitpunkt noch nicht existierte („*As Is versus As Is*“).
- Die Daten werden entsprechend ihrer zeitlich gültigen Klassifizierung strukturiert. Bis zum Semester 2001/1 sind die Studenten mit dem Studienfach *Wirtschaftsinformatik* der *Sozial- und Wirtschaftswissenschaftlichen Fakultät* zugeordnet, danach der *Fakultät für Wirtschaftsinformatik und Angewandte Informatik* („*As Is versus As Was*“).
- Manchmal sind Auswertungen gewünscht, in denen die vergangene Struktur auch auf die neuen Daten angewendet wird. Selbst in diesem Fall können die Daten für eine beliebige Klassifikation aus dem Base-Layer erzeugt werden („*As Was versus As Was*“).
- Es werden nur diejenigen Daten ausgewertet, deren Strukturen im Zeitablauf unverändert blieben („*Like versus Like*“).

Unterstützt das eingesetzte OLAP-Werkzeug zusammengesetzte Schlüssel (Compound Keys), so kann statt des Surrogat-Schlüssels die Kombination aus ehemaligem Schlüssel (z.B. *SF_KZ*) und Zeitattributen verwendet werden. In diesem Fall ist auch eine dynamische Generierung der jeweiligen Analysestruktur mit Hilfe von Sichten (Views) möglich.

Darüber hinaus kann durch die explizite Trennung zwischen Base-Layer und Analytical-Layer besser auf zusätzliche nutzerspezifische Anforderungen, wie beispielsweise die Erweiterung des Schemas, reagiert werden.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag untersucht anhand der Entwicklung eines DWH-Systems für das Hochschulwesen die Beziehung zwischen (a) den Leistungsprozessen, der Organisations- und der Führungsstruktur einer verteilten Organisation und (b) der Architektur eines zugehörigen DWH-Systems. Dabei wird deutlich, dass zentrale DWH-Systeme zur Unterstützung verteilter Organisationen nur bedingt geeignet sind. Zentrale DWH-Systeme besitzen eine relativ starre, inflexible Struktur, welche die Systemevolution behindert. Individuelle Informationsbedarfe der Teilorganisationen werden nicht unterstützt.

Als Lösung wird ein verteiltes, schalenförmiges DWH-System mit speziellen Architekturmerkmalen vorgestellt. Das DWH-System wird im Projekt CEUS seit 1998 entwickelt. Zum 2. Quartal 2001 beginnt die Testphase für die Domäne des Prozesses *Studium und Lehre*, der sich der Pilotbetrieb anschließen wird.

Bezüglich der Anforderungen an die Architektur eines DWH-Systems sehen wir eine Vielzahl von Parallelen zwischen dem Hochschulwesen und verteilten Organisationen in der Industrie, im Handel und im Dienstleistungsbereich. Neben den generellen Anforderungen an flexible, evolutionsfähige Strukturen und an die Unterstützung individueller Informationsbedarfe ist dabei insbesondere die Autonomie der Teilorganisationen in der Konzeption des DWH-Systems zu berücksichtigen. DWH-Teilsysteme, deren Privatsphäre durch Vertrauensbereiche geschützt sind, dürften sich als wesentlich für die Akzeptanz von DWH-Systemen in verteilten Organisationen und damit als kritischer Erfolgsfaktor für den Aufbau und den Betrieb von DWH-Systemen erweisen.

LITERATUR

- [BaGu01] Bauer, A.; Günzel, H. (Hrsg.): Data-Warehouse-Systeme – Architektur, Entwicklung, Anwendung. dpunkt, Heidelberg 2001.
- [Burk00] Burkert, H.: Datenschutz. In: Jung, R.; Winter, R. (Hrsg.): Data Warehousing Strategie – Erfahrungen, Methoden, Visionen. Springer, Berlin 2000, S. 117-125.
- [Cham98] Chamoni, P.: Entwicklungslinien und Architekturkonzepte des On-Line Analytical Processing. In: Chamoni, P., Gluchowski, P. (Hrsg.): Analytische Informationssysteme. Springer, Berlin 1998, S. 231-250.
- [ChDa97] Chaudhuri, S.; Dayal, U.: An Overview of Data Warehousing and OLAP Technology. In: Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data (SIGMOD'97, Tucson, USA, 13.-15. Mai), 1997, S. 65-74.
- [Dres00] Dressler, C.: Schnittstelle zwischen dem SAP Business Information Warehouse und SAP R/3. In: 3. Workshop des Arbeitskreises „Modellierung und Nutzung von Data Warehouse-Systemen“, Siegen, 2000.
- [FeSi01] Ferstl, O. K.; Sinz, E. J.: Grundlagen der Wirtschaftsinformatik. Band 1, 4. Auflage, Oldenbourg, München 2001.
- [Inmo96] Inmon, W. H.: Building the Data Warehouse. 2. Auflage, Wiley, New York 1996.
- [Inmo99] Inmon, W. H.: Building the Operational Data Store. 2. Auflage, Wiley, New York 1999.
- [JLVV00] Jarke, M.; Lenzerini, M.; Vassiliou, Y.; Vassiliadis, P.: Fundamentals of Data Warehouses. Springer, Berlin 2000.
- [Kimb96] Kimbal, R.: The Data Warehouse Toolkit. Wiley, New York 1996.
- [Kurz99] Kurz, A.: Data Warehousing – Enabling Technology. MITP, Bonn 1999.
- [LoDi87] Lockemann, P. C.; Dittrich, K. R.: Architektur von Datenbanksystemen. In: Lockemann P. C., Schmidt J. W. (Hrsg.): Datenbankhandbuch. Springer, Berlin 1987, S. 85-161.
- [Moen98] Möncke, U.: Data Warehouses – eine Herausforderung für den Datenschutz? In: Datenschutz und Datensicherheit 22 (1998) 10, S. 561-569.
- [Rade96] Raden, N.: Modeling the Data Warehouse. 1996 (http://members.aol.com/nraden/iw0196_1.htm).
- [RiPo00] Rieger, B.; Postert, S.: Technisch-Organisatorische Infrastruktur für effektives Hochschulmanagement. In: Steinberger C., Mayr H.C., Marquardt U., Beyer R., Appellath H.-J. (Hrsg.): Unternehmen Hochschule 2000. Berlin, 19. September 2000. Workshop im Rahmen der Jahrestagung der Gesellschaft für Informatik "Informatik 2000", Berlin 19.-22.09.2000, S. 19 – 32.

- [Selk00] Selke, G. W.: Kryptographie – Verfahren, Ziele, Einsatzmöglichkeiten. O'Reilly, Köln 2000.
- [SiBU99] Sinz, E. J.; Böhnlein, M.; Ulbrich-vom Ende, A.: Konzeption eines Data Warehouse-Systems für Hochschulen. In: Workshop "Unternehmen Hochschule" (Informatik'99, 29. Jahrestagung der Gesellschaft für Informatik, Paderborn, 5.-9. Oktober), 1999, S. 111-124.
- [Sinz98] Sinz, E. J.: Universitätsprozesse. In: Küpper, H.-U.; Sinz, E.J. (Hrsg.): Gestaltungskonzepte für Hochschulen. Effizienz, Effektivität, Evolution. Schäffer-Poeschel, Stuttgart 1998.
- [Snod99] Snodgrass, R. T.: Developing Time-Oriented Database Applications in Sql. Morgan Kaufmann, New York 1999.
- [TeU197] Teschke, M.; Ulbrich-vom Ende, A.: Using Materialized Views to Speed Up Data Warehousing. Technical Report, IMMD 6 Universität Erlangen-Nürnberg, 1997.
- [TeU198] Teschke, M.; Ulbrich-vom Ende, A.: Concurrent Warehouse Maintenance without Compromising Session Consistency. In: Proceedings of the 9th International Conference of Database and Expert Systems Applications (DEXA'98, Vienna, Austria, 24.-28. August), 1998, S. 776-785.
- [YaWi98] Yang, J.; Widom, J.: Maintaining Temporal Views over Non-Temporal Information Sources for Data Warehousing. In: Proceedings of the 6th International Conference on Extending Database Technology (EDBT'98, Valencia, Spanien, 23.-27. März), 1998, S. 389-403.