Datenschutz und Datensicherheit in einem landesweiten Data-Warehouse-System für das Hochschulwesen

Elmar J. Sinz, Markus Plaha, Achim Ulbrich-vom Ende[1]

1 Einleitung

Erweiterte Autonomie, verstärkter Wettbewerb, Einführung der Kosten- und Leistungsrechnung, leistungsorientierte Mittelvergabe, Erstellung von Lehrberichten, Profilbildung und Verpflichtung zur permanenten Weiterentwicklung sind Beispiele für aktuelle Entwicklungen, welche die Hochschulen vor bisher nicht bekannte Herausforderungen stellen. Eine zentrale Voraussetzung zur Bewältigung dieser Herausforderungen sind effektive Entscheidungsprozesse auf einer umfassenden, gesicherten und aktuellen Informationsgrundlage. Zur Schaffung dieser Informationsgrundlage können Data-Warehouse-Systeme einen wesentlichen Beitrag leisten.

Data-Warehouse-Systeme sind Teil des Führungsinformationssystems einer Organisation [Sinz02]. Allgemein ist ein Data-Warehouse-System ein Anwendungssystem, das entscheidungsrelevante Daten umfassend und flexibel auswertbar in multidimensionaler Form zur Verfügung stellt. Die Datenbasis des Data-Warehouse-Systems, das eigentliche Data-Warehouse, wird dabei unabhängig von den Datenbanken der operativen Anwendungssysteme verwaltet.

Data-Warehouse-Systeme enthalten im Allgemeinen sensible und zum Teil personenbezogene Daten, die eines besonderen Schutzes bedürfen. Zudem sind Data-Warehouse-Systeme äußerst komplexe Anwendungssysteme, an deren Erstellung, Einführung und Betrieb unterschiedliche Personengruppen mit möglicherweise zum Teil divergierenden Interessen beteiligt sind. Voraussetzung für eine erfolgreiche Einführung und einen effektiven Betrieb eines Data-Warehouse-Systems ist es daher, diese unterschiedlichen Interessen auszugleichen, Vertrauen zu schaffen und damit eine breite Akzeptanz aller Beteiligten zu erreichen.

Eine bekannte Erfahrung aus dem Bereich der Einführung von Anwendungssystemen besagt, dass diese – hinreichende technische und funktionale Einsatzfähigkeit vorausgesetzt - nur dort effektiv eingesetzt werden, wo eine umfassende Akzeptanz der Beteiligten erreicht wird. In der Akzeptanzforschung wird in diesem Zusammenhang zwischen der Einstellungsakzeptanz (innere Einstellung zur Nutzung des Systems) und der Verhaltensakzeptanz (tatsächliche Nutzung des Systems) unterschieden.

[1] Univ.-Prof. Dr. Elmar J. Sinz, 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 | markus.plaha | achim.ulbrich}@wiai.uni-bamberg.de

		Verhaltensakzeptanz	
		positiv	negativ
Einstellungs-akzeptanz	positiv	überzeugter Nutzer	verhinderter Nutzer
	negativ	gezwungener Nutzer	überzeugter Nicht-Nutzer

Bild 1: Akzeptanzarten und Nutzertypen (nach [Krüg90, 280])

Nur wenn eine positive Einstellungsakzeptanz auf eine ebenso positive Verhaltensakzeptanz trifft, sind die Voraussetzungen für einen effektiven Systemeinsatz erfüllt. Die jeweilige Ausprägung des Nutzerverhaltens ist einem Akzeptanzkontinuum entnommen, das von Begeisterung bis Missbrauch/Sabotage reicht [Krüg90, 279 f.].

An der Entwicklung, der Einführung und am Betrieb eines Data-Warehouse-Systems sind insbesondere folgende Personengruppen beteiligt:

- Auftraggeber des Data-Warehouse-Systems,
- Entwickler des Data-Warehouse-Systems,
- Betreiber des Data-Warehouse-Systems,
- Nutzer des Data-Warehouse-Systems (Entscheider),
- Betreiber der operativen Datenquellen (Datenlieferanten),
- Betroffene (über die personenbezogene Daten gespeichert werden; soweit es sich dabei um die Mitarbeiterinnen und Mitarbeiter einer Organisation handelt, werden diese durch die Personalvertretung repräsentiert) und
- der Datenschutzbeauftragte (der Organisation oder des Landes).

Für die breite Akzeptanz eines Data-Warehouse-Systems spielen neben der Erfüllung des eigentlichen Systemzwecks (hinreichende Deckung des objektiven und subjektiven Informationsbedarfs der Entscheider) und der Einsetzbarkeit des Systems (effektive Erlern- und Handhabbarkeit, wirtschaftlicher Betrieb usw.) Fragen von Datenschutz und Datensicherheit eine herausragende Rolle. Diesbezügliche Vertrauensdefizite führen schnell zu einer negativen Einstellungs- oder Verhaltensakzeptanz.

Fragen von Datenschutz und Datensicherheit in Data-Warehouse-Systemen stehen im Mittelpunkt des vorliegenden Beitrags. In einer ersten Abgrenzung wird

- unter Datenschutz der Schutz des Persönlichkeitsrechts durch den Schutz personenbezogener Daten,
- unter Datensicherheit der Schutz von Daten vor missbräuchlicher Verwendung

verstanden. Datenschutz und Datensicherheit stellen dabei Formalziele bei der Entwicklung, der Einführung und beim Betrieb von Data-Warehouse-Systemen dar. Zur Erreichung dieser Ziele werden in diesem Beitrag folgende Maßnahmen herangezogen:

- Zweckbegründete Auswahl der zu speichernden Daten,
- Gestaltung als verteiltes Data-Warehouse-System,
- Entwicklung eines differenzierten Berechtigungskonzepts,
- Schaffung von Vertrauensbereichen und vertrauensbasierten Schnittstellen,
- Anonymisierung von personenbezogenen Daten,
- geeignetes Entwicklungs- und Betriebskonzept und
- effektiver Schutz der Datenbestände und Übertragungswege.

Den Hintergrund des Beitrags bildet das Projekt CEUS^{HB} („Computerbasiertes Entscheidungs Unterstützungs System für die Hochschulen in Bayern“), welches Ende 1998 vom Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst initiiert wurde. Der Projektauftrag umfasst in einer ersten Phase die prototypische Entwicklung eines Data-Warehouse-Systems als Kern eines Führungsinformationssystems für das Hochschulwesen in Bayern sowie dessen Einführung beim Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst und bei zwei Pilotuniversitäten. Als Pilothochschulen wurden die Universität Bamberg und die Technische Hochschule München gewählt. In einer zweiten Phase ist die landesweite Einführung des Systems bei allen bayerischen Universitäten geplant. Mit der Projektdurchführung wurde das Bayerische Staatsinstitut für Hochschulforschung und Hochschulplanung (Prof. Dr. H.-U. Küpper; Projektgruppe München) und der Lehrstuhl für Wirtschaftsinformatik, insbesondere Systementwicklung und Datenbankanwendung der Universität Bamberg (Prof. Dr. E. J. Sinz; Projektgruppe Bamberg) beauftragt. Die Projektgruppe Bamberg ist insbesondere für die Systementwicklung und Betreuung des Data-Warehouse-Systems zuständig.

Der Beitrag ist wie folgt gegliedert: In Kapitel 2 werden Grundlagen des Datenschutzes und Maßnahmen für den Umgang mit personenbezogenen Daten vor dem Hintergrund von Data-Warehouse-Systemen behandelt. Die Konzeption des Data-Warehouse-Systems CEUS wird in Kapitel 3 vorgestellt. Anhand der Leistungsprozesse der Hochschulen werden die für CEUS relevanten Domänen abgeleitet. Ausgehend von der Organisations- und Führungsstruktur des Hochschulwesens wird die Systemarchitektur von CEUS erläutert. In Kapitel 4 wird die Berücksichtigung von Datenschutz und Datensicherheit in CEUS anhand der oben genannten Maßnahmen diskutiert. Eine kurze Zusammenfassung und ein Ausblick in Kapitel 5 schließen den Beitrag ab.

2 Datenschutz

2.1 Gesetzliche Grundlagen des Datenschutzes

Der Datenschutz ist gesetzlich auf Bundesebene im Bundesdatenschutzgesetz (BDSG) und zusätzlich auf Landesebene in den Datenschutzgesetzen der einzelnen Bundesländer geregelt (z.B. in Bayern durch das Bayerische Datenschutzgesetz (BayDSG)). Generell befasst sich das Datenschutzgesetz mit der Beeinträchtigung des Persönlichkeitsrechts von Einzelpersonen durch den missbräuchlichen Umgang mit deren personenbezogenen Daten. Personenbezogene Daten sind nach §3 Abs.1 BDSG „Einzelangaben über persönliche und sachliche Verhältnisse einer bestimmten oder bestimmbaren Person“. Eine Person ist durch die eindeutige Zuordenbarkeit personenbezogener Daten (z.B. *Name, Vorname und Anschrift eines Studierenden* oder *Matrikelnummer eines Studierenden*) **bestimmt**. Dem Schutz unterliegen aber auch **bestimmbare Personen**, d.h. Personen, denen sich personenbezogene Daten mit Hilfe von Zusatzinformationen zuordnen lassen ([Büll00, 12], [Möll98]).

Die Beeinträchtigung von Individualrechtsgütern ist durch das Grundgesetz in Art.1 Abs.1 GG geregelt: „Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt.“ Durch Art.2 Abs.1 GG wird die freie Entfaltung von Personen definiert: „Jeder hat das Recht auf freie Entfaltung seiner Persönlichkeit, soweit er nicht die Rechte anderer verletzt.“. Das Bundesverfassungsgericht (BVerfG) hat 1984 im so genannten Volkszählungsurteil dieses Persönlichkeitsrecht in Bezug auf elektronische Datenverarbeitung durch das Recht auf **informationelle Selbstbestimmung** präzisiert. Durch die Speicherung und Verarbeitung personenbezogener Daten kann jedoch das Persönlichkeitsrecht des Betroffenen, insbesondere sein Recht auf informationelle Selbstbestimmung, verletzt werden. In einem demokratischen Gemeinwesen ist Selbstbestimmung eine elementare Bedingung für Handlungs- und Mitwirkungsfähigkeit der Bürger (BVerfGE 65, 1 (43)). Durch die Speicherung und Verarbeitung personenbezogener Daten wird dieses Recht eingegrenzt, was zu einer Verhaltensänderung führen kann, wenn sich Betroffene dadurch „beobachtet“ fühlen. Daraus resultiert eine Beeinträchtigung des Allgemeinwohls.

Ziel der Datenschutzgesetze ist nach §1 Abs.1 BDSG die Vermeidung von Beeinträchtigungen für das Allgemeinwohl und für Einzelpersonen. **Mittel** zur Erreichung dieses Ziels ist das generelle Verbot der Erhebung, Verarbeitung und Nutzung personenbezogener Daten, es sei denn, dass ausdrücklich eine im Gesetz festgeschriebene Erlaubnis vorliegt. Dieses Verbot mit Erlaubnisvorbehalt ist in §4 Abs.1 BDSG geregelt. Kritisch sind nach diesem Gesetz personenbezogene Daten, die Angaben über persönliche oder sachliche Verhältnisse einer Person darstellen (§4 Abs.1 BDSG).

2.2 Grundregeln des Datenschutzes

Zur Umsetzung des Rechts auf informationelle Selbstbestimmung können nach GOLA und JASPERS [GoJa01, 14 f.] die folgenden Grundregeln zur Erhebung, Verarbeitung und Nutzung personenbezogener Daten angegeben werden:

- **Zulässigkeit:** Die als „Verbot mit Erlaubnisvorbehalt“ bezeichnete Grundregel verbietet die Erhebung, Verarbeitung und Nutzung personenbezogener Daten. Zulässig ist dies nur, wenn entweder das Bundesdatenschutzgesetz die Verarbeitung erlaubt (§§28-31 BDSG), eine Rechtsvorschrift außerhalb des Bundesdatenschutzgesetzes diese Tätigkeiten gestattet bzw. anordnet oder der Betroffene selbst in die Erhebung, Verarbeitung und Nutzung seiner Daten einwilligt (§4a BDSG).

- **Datenvermeidung:** Durch Datenvermeidung soll der Umfang der gespeicherten personenbezogenen Daten möglichst gering gehalten werden. Zusätzlich sollen diese Daten möglichst anonymisiert oder pseudonymisiert werden. Laut §3 Abs. 7 BDSG wird unter **Anonymisierung** die Veränderung personenbezogener Daten derart verstanden, dass „die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“. Im Rahmen einer **Pseudonymisierung** werden die personenbezogenen Daten ebenfalls verändert. Im Gegensatz zur Anonymisierung ist jedoch der Rückschluss auf eine Person grundsätzlich möglich. Beispielsweise kann anhand der Matrikelnummer auf den betreffenden Studenten geschlossen werden, auch wenn dessen Name, Anschrift usw. nicht zur Verfügung stehen.

- **Transparenz:** Die Verarbeitung personenbezogener Daten darf nicht ohne die Kenntnis des Betroffenen erfolgen. Nach §4 Abs.3 BDSG ist dieser über den Zweck der Erhebung, Verarbeitung und Nutzung der Daten zu informieren. Allerdings definiert §33 BDSG einen Ausnahmenkatalog, der die Benachrichtigungspflicht einschränkt [GoJa01, 27].

- **Berichtigung, Löschung und Sperrung:** Unrichtig gespeicherte Daten sind zu korrigieren (§35 Abs. 1 BDSG). Weiterhin sind personenbezogene Daten, die nicht der Aufbewahrungspflicht unterliegen, zu löschen, wenn diese für den ursprünglichen Zweck der Erhebung nicht mehr erforderlich sind (§35 Abs. 2 BDSG). Daten mit Aufbewahrungspflicht müssen in diesem Fall für die weitere Verwendung gesperrt werden (§35 Abs. 3 BDSG).

- **Datensicherheit:** Um unbefugten Zugriff auf die Daten zu verhindern, sind geeignete technische (z.B. Zugriffsschutz durch Passwort, Verschlüsselung der Daten) und organisatorische (z.B. Berechtigungskonzept nach organisatorischer Zugehörigkeit) Sicherungsmaßnahmen zu treffen. Dabei soll der Aufwand stets in angemessenem Verhältnis zum Schutzzweck stehen (§9 BDSG).

Die Einhaltung des Datenschutzes wird durch ein Kontrollsystem überwacht. Dieses System ist mehrstufig und reicht von den innerbetrieblichen Datenschutzbeauftragten (§4f, §4g BDSG) über die zuständigen Aufsichtsbehörden nach Landesrecht (§38 BDSG) bis zu den Landesdatenschutzbeauftragten der einzelnen Bundesländer sowie den Bundesdatenschutzbeauftragten (§§21-26 BDSG). Verstöße gegen das Datenschutzrecht können mit Geld- und Freiheitsstrafen geahndet werden (§§43-44, §§7-8 BDSG).

2.3 Maßnahmen beim Umgang mit personenbezogenen Daten

Das Bundesdatenschutzgesetz gestattet die Erhebung, Verarbeitung und Nutzung personenbezogener Daten nach dem Prinzip „Verbot mit Erlaubnisvorbehalt“ (§4 Abs.1 BDSG), sofern die Einwilligung des Betroffenen vorliegt (§4a BDSG) oder Zulässigkeitstatbestände gegeben sind (§28 BDSG). Unter folgenden Voraussetzungen ist aus datenschutzrechtlicher Sicht ein Umgang mit personenbezogenen Daten zulässig ([Mönc98], [Burk00]):

- **Einwilligung des Betroffenen (§4a BDSG):** Eine rechtsgültige Einwilligung des Betroffenen zur Erhebung, Verarbeitung und Nutzung seiner Daten legalisiert den Umgang mit personenbezogenen Daten (siehe Abschnitt 2.2 „Zulässigkeit“). Dazu müssen die Betroffenen präzise über die jeweiligen Zwecke der Datenverarbeitung informiert werden (siehe Abschnitt 2.2 „Transparenz“).
- **Zulässigkeitstatbestände (§28 BDSG):** Ein legitimer Umgang mit personenbezogenen Daten kann auch ohne die Einwilligung der Betroffenen aufgrund gesetzlicher Erlaubnis erfolgen. Die Datenverarbeitung und –nutzung kann entweder durch vertragliche bzw. vertragsähnliche Zwecke (§28 Abs.1 Nr.1 BDSG) oder zur Wahrung berechtigter Interessen (§28 Abs.1 Nr.2 BDSG) gestattet sein [Bize99].
    - Steht die Verarbeitung und Nutzung der Daten in unmittelbarem Zusammenhang mit dem jeweiligen Vertragszweck, der sich aus übereinstimmenden Willenserklärungen ergibt, so ist diese zur Erfüllung der Pflichten aus dem Vertrag erlaubt (§28 Abs.1 Nr.1 BDSG). Mit Erfüllung des Vertragszwecks erlischt allerdings diese Erlaubnis, es sei denn eine längerfristige Aufbewahrung der Daten ist vorgeschrieben (siehe Abschnitt 2.2 „Berichtigung, Löschung und Sperrung“).
    - Eine Verarbeitung und Nutzung personenbezogener Daten ist auch gestattet, um berechtigte Interessen der Verwender zu wahren, wenn entgegenstehende schutzwürdige Interessen der betroffenen Personen nicht verletzt werden (§28 Abs.1 Nr.2 BDSG). Personenbezogene Daten, die für berechtigte eigene Interessen (z.B. Werbe- oder Marketingaktivitäten) gespeichert werden, dürfen nur für den angegebenen Zweck verwendet werden und müssen hierfür ausdrücklich erforderlich sein.

Legal im Sinne des Datenschutzrechts ist der Umgang mit personenbezogenen Daten stets dann, wenn diese in anonymisierter Form vorliegen:

- **Anonymisierung:** Anonymisierte Daten sind nach §3 Abs.7 BDSG veränderte personenbezogene Daten, die eine Zuordnung von Einzelangaben über persönliche oder sachliche Verhältnisse zu einer bestimmten oder bestimmbaren Person nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft ermöglichen. Der Umgang mit anonymisierten Daten unterliegt nicht dem Datenschutzrecht und ist somit „uneingeschränkt zulässig" [Büll00, 13]. Das Kriterium für die Gültigkeit einer Anonymisierung ist der Aufwand, der aufgebracht werden muss, um einen Personenbezug wieder herstellen zu können (§3 Abs.6 BDSG). Hierzu ist allein das Entfernen der **Attribute zur externen Identifikation**, wie z.B. *Name, Vorname, Anschrift* nicht ausreichend, da über sogenannte künstliche Schlüssel (**Attribute zur internen Identifikation**), z.B. die *Matrikelnummer bei Studierenden*, die Möglichkeit der Reidentifikation einer realen Person erhalten bleibt. Um den Personenbezug aufzulösen stehen drei Alternativen zur Verfügung ([Mönc98, 565], [Mönc99]):
    - **Aggregation:** Im Rahmen der Aggregation werden Daten über Einzelpersonen zu Daten über Gruppen verdichtet. Dadurch sind nur noch Aussagen über die Gruppe der Personen mit gleichen Merkmalen möglich (z.B. *Anzahl der Studierenden, die dasselbe Studienfach studieren* oder *Anzahl der Studierenden, die dasselbe Studienfach studieren und aus demselben Bundesland stammen*). Ein Rückschluss auf Einzelpersonen ist dadurch nicht mehr möglich. Mit der Anzahl der gespeicherten Merkmale nimmt allerdings der Grad der Anonymisierung ab, da im ungünstigsten Fall nur eine einzige Person in eine Gruppe fällt. Sind umgekehrt nur wenige Merkmale gespeichert, so können vielfach keine sinnvollen Auswertungen mehr vorgenommen werden.
    - **Individuelle anonyme Profile:** Bei dieser Art der Anonymisierung bleiben die individuellen Datensätze im Datenbestand bestehen. Allerdings wird der Zusammenhang zwischen den unterschiedlichen Merkmalsausprägungen eines Individuums entfernt. Beispielsweise werden demographische und auf den Studiengang bezogene Attributwerte eines Studenten getrennt verwaltet. Dadurch sind gleichzeitige Aussagen über Wohnort und Studienrichtung eines Studierenden nicht mehr möglich, wodurch der Rückschluss auf den Studierenden im Allgemeinen nicht möglich ist.
    - **Faktische Anonymisierung:** Im Rahmen der faktischen Anonymisierung bleibt der Bezug zur einzelnen Person prinzipiell erhalten, ist jedoch nicht offensichtlich. Die **Deanonymisierung** sollte im Verhältnis zur erlangten Information entweder zu aufwendig sein, oder es muss weniger aufwendige Alternativmethoden geben, um an die gleichen Informationen zu gelangen. Beispielsweise bleibt der Personenbezug durch die Verschlüsselung der *Matrikelnummer eines Studierenden* erhalten, jedoch ist ein Rückschluss auf die ursprüngliche Matrikelnummer und damit auf den eigentlichen Studierenden nur über die entsprechende Dekodierung möglich. Der Aufwand für die Deanonymisierung ist maßgeblich für den Grad der faktischen Anonymisierung.

2.4 Datenschutz in Data-Warehouse-Systemen

Bisher wurde lediglich allgemein auf Datenschutzaspekte eingegangen. In diesem Abschnitt werden spezifische Anforderungen an Data-Warehouse-Systeme herausgearbeitet. Dazu werden zunächst die aus der Sicht des Datenschutzes relevanten Hauptmerkmale von Data-Warehouse-Systemen aufgezeigt [Mönc98, 562]:

- **Zweckbestimmung:** Im Gegensatz zu operativen Systemen, deren Zweck die Abwicklung bestimmter Geschäftsvorfälle ist (z.B. *Personalverwaltungssystem*), werden Data-Warehouse-Systeme zur Gewinnung von Informationen im Rahmen **strategischer Entscheidungen** genutzt.

- **Haltung historisierter Daten:** Aus der Zweckbestimmung resultiert die Art der Datenverarbeitung und damit die Art der Datenhaltung. Operative Systeme aktualisieren den Datenbestand während der Abwicklung der laufenden Geschäftsvorfälle. Typischerweise werden Daten eingefügt, modifiziert und nach Abschluss des Geschäftsvorfalls gelöscht. Es werden jeweils nur aktuelle Daten gehalten, die nach Ablauf einer Transaktion überschrieben oder gelöscht werden. Data-Warehouse-Systeme dienen der Unterstützung strategisch relevanter Auswertungen über lange Zeiträume, wozu nur ein lesender Zugriff auf den Datenbestand erfolgt. Historisiert gespeicherte Daten im Data-Warehouse ermöglichen Zeitreihenanalysen. Hierzu werden periodisch entscheidungsrelevante Daten aus unterschiedlichen operativen Systemen in das Data-Warehouse geladen, wodurch dieses sukzessive erweitert wird. Während der Datenübernahme aus den operativen Systemen muss ggf. eine Anpassung der Daten erfolgen, um eine konsolidierte Datenbasis zu erhalten.

- **Operationen des multidimensionalen Datenmodells:** Der Zugriff auf die Daten im Data-Warehouse erfolgt typischerweise über **Online Analytical Processing** (**OLAP**). Dabei werden die Daten den Nutzern in Form einer **multidimensionalen Datenstruktur** (**Hyperwürfel**) zur Verfügung gestellt [BöUl00]. Diese besteht aus mehreren **Dimensionen** mit zugehörigen **Attributen** (qualitative Daten) und **Kennzahlen** (quantitative Daten). Die Kennzahlen können bezüglich der Attribute ausgewertet werden. Attribute innerhalb einer Dimension können in hierarchischen Beziehungen stehen. Navigationsoperationen ermöglichen ein nahezu beliebiges „navigieren“ im Datenbestand. Beispielsweise lässt sich mittels der **Drill-Down**-Operation ein existierendes Anfrageergebnis bezüglich eines Attributs innerhalb der gleichen Dimension verfeinern. So kann ausgehend von den *Studierenden einer Hochschule* zu den *Studierenden der einzelnen Fakultäten* navigiert werden. Die **Roll-Up**-Operation bildet die inverse Funktion zur Drill-Down-Operation. Mittels zusätzlicher Operationen (z.B. **Drill-Anywhere**-Operation) ist eine beliebige Navigation innerhalb des Datenbestands des Data-Warehouse möglich *(z.B. Aufteilen der Studierenden einer Hochschule in männliche und weibliche Studierende)*.

- **Nutzerinteraktion:** Bei operativen Systemen bearbeitet ein bestimmter Nutzer im Allgemeinen Daten aus einem klar umrissenen Ausschnitt der betrieblichen Domäne. Da die

Nutzerinteraktion typischerweise über vordefinierte Bildschirmmasken und Menüs erfolgt, lässt sich der Zugriff auf personenbezogene Daten klar definieren und dokumentieren. Im Gegensatz dazu unterstützen Data-Warehouse-Systeme eine nahezu uneingeschränkte Datenrecherche über die gesamte betriebliche Domäne.

Es ist ersichtlich, dass die beschriebenen Merkmale von Data-Warehouse-Systemen im Zusammenhang mit personenbezogenen Daten Gefahren für das Recht auf informationelle Selbstbestimmung (Art.2 Abs.1 GG i.V.m. Art.1 Abs.1 GG) darstellen. Durch redundante Datenhaltung, durch Zusammenführung der Daten einer Person aus verschiedenen operativen Systemen und durch die Historisierung entsteht ein umfangreicher Datenbestand über die betroffene Person.

Korrespondierend zu Abschnitt 2.3 werden im Folgenden Maßnahmen zur Umsetzung des Datenschutzes in Data-Warehouse-Systemen vorgestellt:

- **Einwilligung des Betroffenen (§4a BDSG):** Voraussetzung für die Einwilligung des Betroffenen ist dessen Aufklärung über den Zwecke der Datenverarbeitung (siehe Abschnitt 2.2 „Transparenz“). Dieser lässt sich für Data-Warehouse-Systeme nicht immer exakt angegeben. Häufig ergibt sich der genaue Zweck erst als Ergebnis einer strategischen Auswertung [Büll00, 15]. Ein weiterer Punkt, der eine Einwilligung der Betroffenen bei einer Datenverarbeitung in Data-Warehouse-Systemen erschwert, ist der große Umfang der gespeicherten Daten, der aus der Historisierung und der Zusammenführung von Daten aus verschiedenen operativen Systemen resultiert. Eine Einwilligung aller Personen, deren Daten in einem Data-Warehouse gespeichert werden, erscheint daher nahezu unmöglich.
- **Zulässigkeitstatbestände (§28 BDSG):** Die Datenverarbeitung und –nutzung kann entweder durch vertragliche bzw. vertragsähnliche Zwecke (§28 Abs.1 Nr.1 BDSG) oder zur Wahrung berechtigter Interessen (§28 Abs.1 Nr.2 BDSG) gestattet sein.
    - **Vertragliche bzw. vertragsähnliche Zwecke (§28 Abs.1 Nr.1 BDSG):** Eine (längerfristige) Verarbeitung der Daten in einem Data-Warehouse ist durch den Vertragszweck nicht gegeben [Büll00, 13]. Eine Datenverarbeitung, die zur Erfüllung des Vertragszwecks dient, findet in den operativen Systemen statt. Auch für eine dauerhafte Speicherung der Daten aufgrund von Aufbewahrungspflichten ist ein Data-Warehouse-System nicht geeignet, da dieses Auswertungsfunktionen zur Verfügung stellt, die zur reinen Archivierung nicht benötigt werden [Mönc98, 566].
    - **Wahrung berechtigter Interessen (§28 Abs.1 Nr.2 BDSG):** Personenbezogene Daten, die für berechtigte eigene Interessen (z.B. Werbe- oder Marketingaktivitäten) im Data-Warehouse gespeichert werden, dürfen nur für den angegebenen Zweck verwendet werden. Für Data-Warehouse-Systeme kann jedoch nicht immer ein eindeutiger Verwendungszweck angegeben werden [Büll00, 14].

Bei der Speicherung personenbezogener Daten in Data-Warehouse-Systemen lässt sich die Einwilligung der Betroffenen nach §4a BDSG nur schwer erreichen; die Zulässigkeitstatbe-

stände nach §28 BDSG sind kaum gegeben. Daher bleibt für eine legale Datenverarbeitung in Data-Warehouse-Systemen in der Regel nur die Anonymisierung der personenbezogenen Daten, da anonymisierte Daten nicht dem Datenschutzrecht unterliegen. Neben dem Entfernen der personenbezogenen Attribute, die eine externe Identifikation der Person ermöglichen, muss zusätzlich eine Reidentifikation der Person über künstliche Attribute ausgeschlossen werden. Hierzu können die in Abschnitt 2.3 genannten Alternativen „Aggregation“, „Individuelle anonyme Profile“ und „Faktische Anonymisierung“ angewandt werden. In Abschnitt 4 wird auf die entsprechende Umsetzung dieser Maßnahmen zur Gewährleistung des Datenschutzes im Projekt CEUS eingegangen. Hierzu wird zunächst die Architektur des Data-Warehouse-Systems CEUS vorgestellt.

3 Konzeption des Data-Warehouse-Systems CEUS

3.1 Domänen von CEUS

Kernaufgabe des Hochschulmanagements ist die Gestaltung und Lenkung der Leistungsprozesse der Hochschulen. Diese Leistungsprozesse bilden damit den Ausgangspunkt für die Konzeption eines landesweiten Data-Warehouse-Systems für das Hochschulwesen.

Leistungsprozesse (Geschäftsprozesse) einer Hochschule können nach Haupt- und Serviceprozessen unterschieden werden [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.

- **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 Rahmen des Hochschulmanagements sind vor allem die Serviceprozesse **Personalwirtschaft** und **Mittelbewirtschaftung** von Interesse. Die Gestaltung der Serviceprozesse und ihre Abgrenzung gegenüber den Hauptprozessen unterliegt Art. 43 Abs. 2 Satz 1 Bayerisches Hochschulgesetz (BayHSchG), wonach die Verwaltung so einzurichten ist, „dass die Fachbereiche, wissenschaftlichen Einrichtungen [...] möglichst von Verwaltungsaufgaben entlastet werden“.

Ausgehend von den Leistungsprozessen wurden folgende Domänen bestimmt, die durch das Data-Warehouse-System CEUS unterstützt werden sollen:

- **Studium und Lehre:** Diese Domäne umfasst Daten über Studenten und Studienfachbelegungen sowie die erbrachten Prüfungsleistungen.

- **Stellen und Personal:** In dieser Domäne werden Daten über Stellen und Personal für die Zwecke der Personalwirtschaft subsummiert.
- **Sachmittel:** Die Domäne Sachmittel beinhaltet haushaltsrelevante Daten zur Mittelbewirtschaftung.

3.2 Organisations- und Führungsstruktur des Hochschulwesens

Die Leistungserstellung im Hochschulwesen erfolgt im Wesentlichen gemäß der in Bild 2 dargestellten Organisations- und Führungsstruktur.

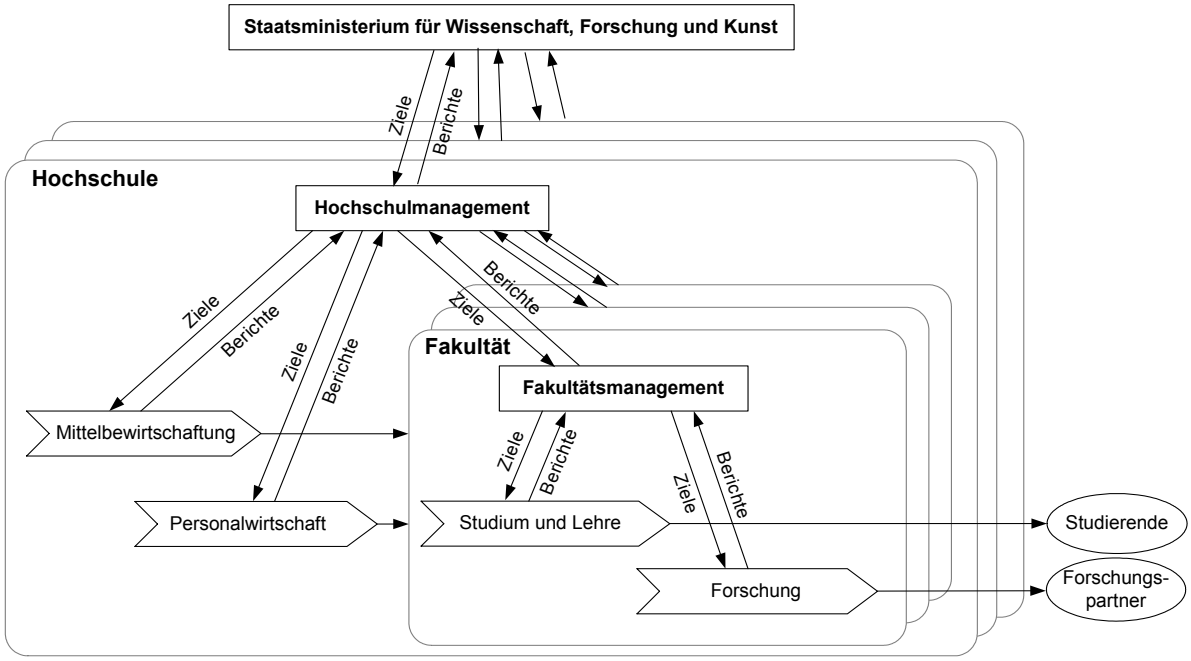

Bild 2: Organisations- und Führungsstruktur des Hochschulwesens ([SiBU99], [SBPU01])

Hochschulen sind in Fakultäten (bzw. Fachbereiche) gegliedert. Das Fakultätsmanagement bezieht sich auf die Hauptprozesse *Studium und Lehre* sowie *Forschung* der jeweiligen Fakultät. Das Hochschulmanagement bezieht sich auf alle Fakultäten einer Hochschule sowie auf die Serviceprozesse des Verwaltungsbereichs (*Mittelbewirtschaftung*, *Personalwirtschaft* und andere). Die Managementaufgaben des Bayerischen Staatsministeriums für Wissenschaft, Forschung und Kunst beziehen sich auf alle Hochschulen des Landes.

Das Hochschulwesen eines Landes weist 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**.

Die primären Nutzerkreise von CEUS lassen sich unmittelbar aus den Organisationseinheiten und aus der Führungsstruktur des Hochschulwesens ableiten:

- **Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst:** Die Entscheidungsträger des Staatsministeriums benötigen zur Koordination des gesamten Hochschulwesens im Wesentlichen die nach dem Bayerischen Hochschulstatistikgesetz und den zugehörigen Verordnungen definierten Daten.
- **Hochschulmanagement:** Entscheidungsträger an den Hochschulen sind die Hochschulleitung, zentrale Gremien und beauftragte Referate. Der Umfang der benötigten Daten ergibt sich aus der Erfüllung der jeweiligen Aufgaben nach dem Bayerischen Hochschulgesetz. Die Entscheidungsträger erhalten ausschließlich Zugriff auf den sie betreffenden Bereich des Data-Warehouse.
- **Fakultätsmanagement:** Entscheidungsträger an den Fakultäten sind die Fakultätsleitung (Dekan, Studiendekan) und die diversen Gremien (Prüfungsausschuss, Promotionsausschuss usw.). Auch hier ergibt sich der Umfang der benötigten Daten aus der jeweiligen Aufgabenerfüllung.

3.3 Systemarchitektur von CEUS

CEUS ist als hierarchisch verteiltes Data-Warehouse-System konzipiert (Bild 3). Dabei sind den in Bild 2 dargestellten Managementinstanzen (Staatsministerium für Wissenschaft, Forschung und Kunst, Hochschulmanagement, Fakultätsmanagement) Teil-Data-Warehouses zugeordnet, die in ihrem Inhalt, Umfang und Aggregationsgrad auf den Informationsbedarf des Managements auf der jeweiligen Ebene abgestimmt sind. Gründe für die Wahl dieser Architektur sind:

- Auf den einzelnen Managementebenen existieren unterschiedliche Informationsbedarfe für die jeweils durchzuführenden Entscheidungsaufgaben.
- Die Komplexität des Gesamtsystems ist leichter handhabbar.
- Die Architektur unterstützt eine sukzessive Entwicklung und Einführung des Data-Warehouse-Systems.
- Die Architektur ermöglicht vertrauensbasierte Schnittstellen zwischen den einzelnen Teil-Data-Warehouse-Systemen. Jede für den Betrieb eines Teil-Data-Warehouse verantwortliche Person hat dabei die vollständige Information und die Kontrolle darüber, welche Daten in welcher Verdichtung an ein anderes Teil-Data-Warehouse übermittelt werden.

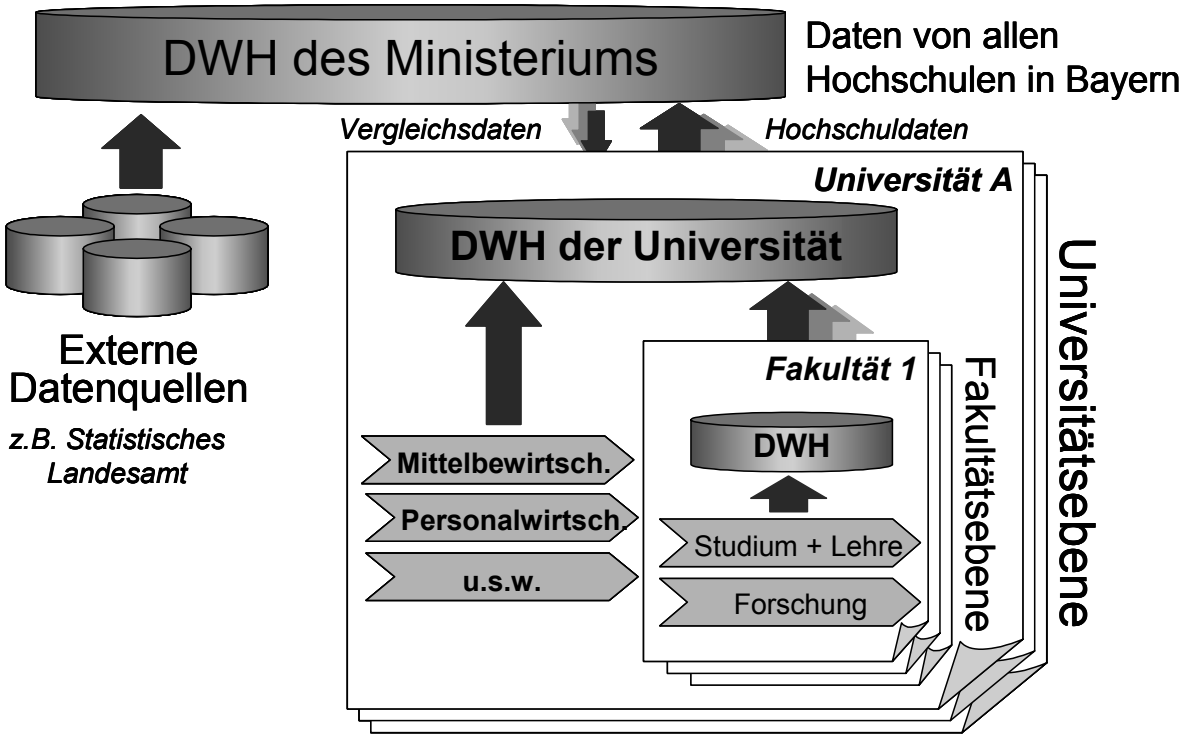

Bild 3: Architektur des Data-Warehouse-Systems CEUS ([SiBU99], [SBPU01])

Im vollen Ausbau umfasst CEUS damit ein Data-Warehouse auf Landesebene (Data-Warehouse des Ministeriums), je ein Data-Warehouse der Hochschulen sowie ggf. Data-Warehouses der einzelnen Fakultäten. In jedes Data-Warehouse können neben den aus den Haupt- und Serviceprozessen gewonnenen Daten auch zusätzliche Daten aus externen Datenquellen eingestellt werden.

Als operative Systeme zur Unterstützung der Haupt- und Serviceprozesse werden insbesondere berücksichtigt:

- *Studium und Lehre*: HIS-SOS, HIS-POS, FlexNow!.
- *Mittelbewirtschaftung*: HIS-MBS, SAP R/3 (HER).
- *Personalwirtschaft*: DIAPERS, SAP R/3 (HER).

Die Architektur eines herkömmlichen zentralen Data-Warehouse-Systems lässt sich in drei Ebenen untergliedern [SiBU99]: eine *Datenerfassungsebene* mit der Schnittstelle zu den operativen Systemen, eine *Datenhaltungsebene* mit dem eigentlichen Data-Warehouse und eine *Datenbereitstellungsebene* mit den Schnittstellen zu den Präsentationswerkzeugen und anderen Anwendungssystemen. Allen Ebenen sind Administrationsfunktionen zugeordnet, die durch eine *Metadatenbank* unterstützt werden.

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 Bild 2). Daneben resultiert aus der Teilautonomie von Organisationseinheiten (Hochschule, Fakultät) die Erfordernis einer geschützten Privatsphäre von Da-

ten. Diese Merkmale bleiben beim zentralen Ansatz unberücksichtigt. Daher wurde ein mehrstufiges, schalenförmiges, verteiltes Data-Warehouse-System mit abgestimmter Reichweite und Aggregation der verwalteten Daten konzipiert (Bild 3).

Die Architektur des Data-Warehouse-Systems folgt der Führungsstruktur des Hochschulwesens [SiBU99]:

- Auf der untersten Ebene befinden sich die Teil-Data-Warehouse-Systeme der einzelnen Fakultäten einer Hochschule. Diese basieren auf den Daten der von der jeweiligen Fakultät betriebenen Prozesse *Studium und Lehre* sowie *Forschung (die Unterstützung der Domäne Forschung ist im Projektauftrag von CEUS nicht enthalten*).
- Die nächst höhere Ebene bildet das Teil-Data-Warehouse-System der jeweiligen Hochschule mit den detaillierten Daten der Serviceprozesse und den durch die Teil-Data-Warehouse-Systeme 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 Landesamts für Statistik und Datenverarbeitung.

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

4 Datenschutz und Datensicherheit in CEUS

Belange des Datenschutzes im eingangs beschriebenen Sinn werden in CEUS durch eine Reihe von konzeptuellen Merkmalen berücksichtigt [SBPU02]. Im Folgenden werden diese Merkmale vorgestellt.

4.1 Auswahl der Datenbestände

Zentrale Aufgabe des Data-Warehouse-Systems CEUS ist die Versorgung der Entscheidungsträger im Hochschulwesen mit entscheidungsrelevanten Informationen. Grundlage hierfür ist eine Abstimmung zwischen **Informationsbedarf**, **Informationsnachfrage** und **Informationsangebot** [PiRW98, 106 f.]. Beim Informationsbedarf wird zwischen **subjektivem** und **objektivem Informationsbedarf** unterschieden, die in der Regel nicht deckungsgleich sind. Der subjektive Informationsbedarf beinhaltet diejenigen Informationen, die einem Entscheidungsträger zur Erfüllung einer Aufgabe relevant erscheinen. Dagegen spiegelt der objektive

Informationsbedarf alle Informationen wider, die bei einer systematischen Analyse zur optimalen Erfüllung einer Aufgabe erforderlich sind [BiMR94, 7 f.]. Der Teil des subjektiven Informationsbedarfs, den ein Entscheidungsträger bei der Erfüllung seiner Aufgabe tatsächlich anfordert, ist die Informationsnachfrage. Das Informationsangebot ist dagegen die gesamte Menge der zur Verfügung stehenden Informationen.

Um den in das Data-Warehouse-System CEUS aufzunehmenden Datenbestand zu bestimmen, wurde vom Institut für Hochschulforschung und Hochschulplanung eine Informationsbedarfsanalyse durchgeführt. Zur Bestimmung des subjektiven Informationsbedarfs dienten induktive Analysemethoden, wie Fragebögen, Interviews und Workshops. Dabei wurde u.a. festgestellt, dass die Entscheidungsträger keinerlei personenbezogene Daten benötigen. Vielmehr wurden aggregierte Daten nachgefragt, die Aussagen über Personengruppen (siehe Abschnitt 2.3 „Aggregation“) mit bestimmten Merkmalen liefern (z.B. *Anzahl der Studierenden je C4-Lehrstuhl*). Ergänzend dazu dienten deduktive Verfahren zur Ermittlung des objektiven Informationsbedarfs. Diese Analyse lieferte insbesondere Hinweise für die Festlegung der Zugriffsrechte auf die einzelnen Domänen im CEUS-System.

Darüber hinaus wurde das Informationsangebot der jeweiligen Teil-Data-Warehouses betrachtet. Die Datenbestände der zugrunde liegenden operativen Systeme an den Hochschulen definieren das Informationsangebot für die Data-Warehouses der Hochschulen. Das Informationsangebot des Data-Warehouses auf Landesebene wird durch das Hochschulstatistikgesetz (HStatG) geregelt. Der endgültige Datenbestand der Data-Warehouses resultiert aus der Schnittmenge der Informationsbedarfe und des Informationsangebots.

4.2 Anonymisierung personenbezogener Daten

In Abschnitt 2.4 wurde bereits festgestellt, dass die Maßnahmen **Einwilligung des Betroffenen** (§4a BDSG) sowie die **Zulässigkeitstatbestände** (§28 BDSG) - sowohl **vertragliche bzw. vertragsähnliche Zwecke** (§28 Abs.1 Nr.1 BDSG) als auch die **Wahrung berechtigter Interessen** (§28 Abs.1 Nr.2 BDSG) - zur Umsetzung des Datenschutzes in Data-Warehouse-Systemen nicht geeignet sind. Daher muss der Personenbezug im Data-Warehouse-System CEUS mittels Anonymisierung eliminiert werden.

Die Anonymisierung wird wie folgt realisiert: **Personenbezogene Daten zur externen Identifikation** (z.B. *Name, Anschrift, Telefonnummer*) gelangen überhaupt nicht in das Data-Warehouse von CEUS. Sie werden bei der Datenübernahme aus den operativen Systemen nicht erhoben.

Darüber hinaus muss die Reidentifikation einer betroffenen Person über **Attribute zur internen Identifikation** (z.B. *Matrikelnummer*) unterbunden werden (siehe Abschnitt 2.3). Dabei wird zwischen der Übernahme der Daten aus den operativen Systemen und der Verwendung der Daten an der Nutzerschnittstelle unterschieden:

- **Datenübernahme:** Hinsichtlich der Übernahme der Daten aus den entsprechenden Vorsystemen sind das landesweite und das hochschulweite Data-Warehouse getrennt zu betrachten:
    - **Hochschulebene:** Die Schlüsselattribute zur internen Identifikation werden beim Laden des hochschulweiten Data-Warehouse zur Qualitätssicherung und zur Verknüpfung mit anderen Tabellen benötigt (z.B. die Verknüpfung der Studentendaten mit den jeweiligen Studienverlaufsdaten erfolgt über die Matrikelnummer). Aus Gründen des Datenschutzes werden während der Datenübernahme aus den operativen Systemen alle Daten zur internen Identifikation durch Verschlüsselung **faktisch anonymisiert** (siehe Abschnitt 2.3 „Faktische Anonymisierung"). Dadurch bleibt der Zusammenhang zwischen den Tabellen erhalten und die Datenqualität ist gewährleistet. Ein Rückschluss auf die ursprüngliche Matrikelnummer und damit auf einen bestimmten Studierenden ist nur über eine Dekodierung möglich.
    - **Landesebene:** Auf der Landesebene von CEUS werden diese Attribute nicht benötigt, da die Qualitätssicherung und die Zusammenführung der Daten bereits in den Vorsystemen bzw. durch das Landesamt für Statistik und Datenverarbeitung erfolgt. Daher werden diese Daten nicht in das landesweite CEUS-System übernommen.
- **Nutzerschnittstelle:** Aus der Informationsbedarfsanalyse (siehe Abschnitt 4.1) ergab sich, dass die Attribute zur internen Identifikation für die Entscheidungsfindung nicht benötigt werden. Aus diesem Grund werden diese dem Nutzer nicht zur Verfügung gestellt. Darüber hinaus kann der Nutzer nur auf **aggregierte Datenbestände** (siehe Abschnitt 2.3 „Aggregation") zugreifen. Statt Aussagen über Einzelpersonen sind somit nur noch Auswertungen über Personengruppen mit gleichen Merkmalen möglich.

4.3 Nutzerkreise von CEUS

Durch die Konzeption von CEUS als hierarchisch verteiltes Data-Warehouse-System (siehe Abschnitt 3.3) kann jede Managementebene gezielt mit den dort zur Entscheidungsfindung benötigten Informationen in der geeigneten Aggregationsstufe versorgt werden (siehe Abschnitt 4.1). Neben der Architektur lassen sich auch die Nutzerkreise von CEUS direkt aus der Managementstruktur der Hochschulen (siehe Abschnitt 3.2) ableiten. Es werden folgende Nutzerkreise unterschieden:

- **Bayerisches Staatsministerium für Wissenschaft, Forschung und Kunst:** Die Entscheidungsträger des Staatsministeriums benötigen im wesentlichen die nach dem Bayerischen Hochschulstatistikgesetz und den zugehörigen Verordnungen definierten Daten. Im Vergleich zu den Daten an den einzelnen Hochschulen weisen diese eine höhere Aggregationsstufe auf; in der Regel werden weniger Merkmale gespeichert.

- **Hochschulen:** Entscheidungsträger an den Hochschulen sind die Hochschulleitung, zentrale Gremien und beauftragte Referate; an den Fakultäten die Fakultätsleitung (Dekan, Studiendekan) und diverse Gremien (Prüfungsausschuss, Promotionsausschuss usw.). Die für diesen Nutzerkreis benötigten Informationen ergeben sich aus der Erfüllung der jeweiligen Aufgaben nach dem Bayerischen Hochschulgesetz (BayHSchG). Aus diesem Grund ist der Datenbestand in den Teil-Data-Warehouses für die Hochschulen gegenüber dem Datenbestand des Staatsministeriums weniger stark aggregiert. Außerdem erhalten die Entscheidungsträger ausschließlich Zugriff auf den für sie relevanten Bereich des Data-Warehouse.
- **Interessierte Öffentlichkeit (z.B. Studenten):** Dieser Personenkreis erhält keinen direkten Zugriff auf das Data-Warehouse. Es werden ausschließlich statische Standardberichte zur Verfügung gestellt, deren Inhalt sich am Statistischen Jahrbuch orientiert.

Die Architektur von CEUS unterstützt explizit die Datenunabhängigkeit zwischen den Teilsystemen benachbarter Ebenen und damit zwischen den Informationsbedarfen der jeweiligen Nutzerkreise. Auf den einzelnen Ebenen müssen die Daten ggf. nach unterschiedlichen Strukturmerkmalen gegliedert werden. Beispielsweise ist für das landesweite Data-Warehouse eine einheitliche Datenstruktur über alle Universitäten notwendig. Aufgrund der strukturellen Unabhängigkeit des universitären Warehouse können dort trotzdem universitätsspezifische Details und Datenquellen berücksichtigt werden, die dem landesweiten Warehouse nicht zur Verfügung stehen.

4.4 Datenübernahme aus externen Datenquellen

4.4.1 Vertrauensbereiche

Um den Datenschutzproblemen auch auf der Akzeptanzebene zu begegnen, wird in CEUS das Konzept „**Vertrauensbereich“** eingeführt. Wesentliche Merkmale eines Vertrauensbereichs sind:

- Ein Vertrauensbereich schützt den Betrieb und die Nutzung eines Anwendungssystems (automatisiertes Verfahren im Sinne des Datenschutzgesetzes) und kapselt diese gegenüber Dritten.
- Die Schnittstellen eines Vertrauensbereichs sind öffentlich bekannt (welche Daten werden wann an wen übergeben).
- Die Verantwortlichkeit für die Realisierung eines Vertrauensbereichs ist klar geregelt und wird im Allgemeinen von den Betreibern und Nutzern eines Anwendungssystems gemeinsam wahrgenommen.

Bei den meisten Data-Warehouse-Systemen ist die Aufgabe der Datenübernahme (ETL = Extraktion, Transformation, Laden) den Betreibern des jeweiligen Teil-Data-Warehouse-Systems zugeordnet. Bei CEUS ergäben sich in diesem Fall datenschutzrelevante Probleme,

weil mit Hilfe der Extraktionswerkzeuge der Durchgriff auf die personenbezogenen Daten der operativen Systeme prinzipiell möglich wäre (Bild 4). Neben der formalen Ebene des Datenschutzes würde insbesondere auch die Akzeptanzebene in Form von Vertrauensproblemen berührt.

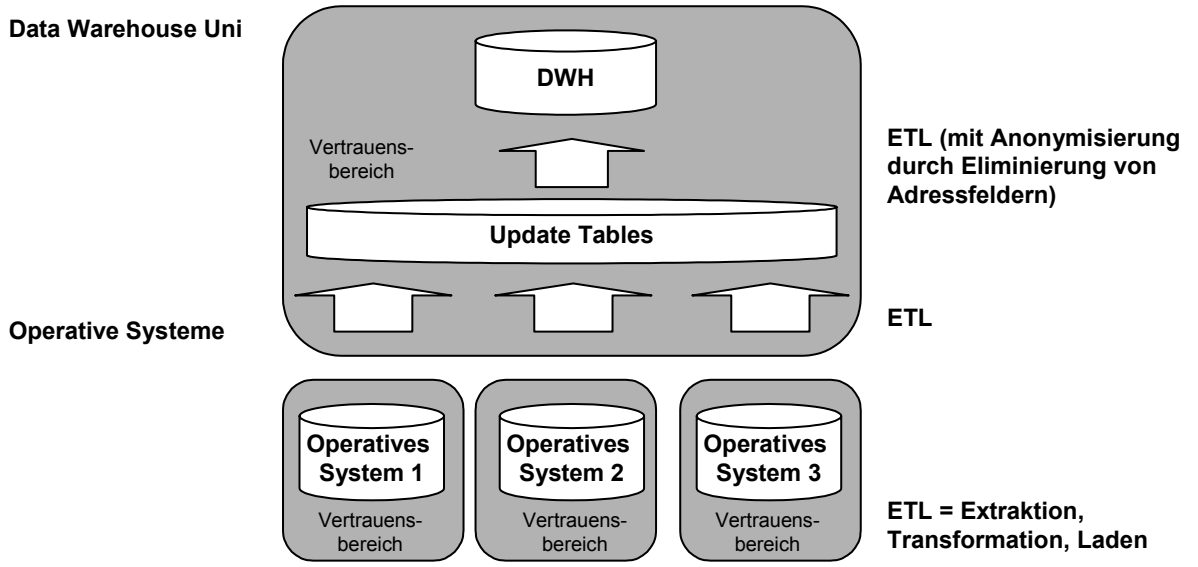

Bild 4: ETL im Vertrauensbereich des Data-Warehouse-Systems

Um diesen Problemen konzeptuell zu begegnen, werden in CEUS wesentliche Funktionen der Datenerfassung in die Vertrauensbereiche der operativen Systeme verlagert (Bild 5). Dazu werden den Betreibern der operativen Systeme Schnittstellen zur Verfügung gestellt, welche es gestatten, die Extraktion und die Anonymisierung der in das Data-Warehouse aufzunehmenden Daten innerhalb des jeweiligen Vertrauensbereichs durchzuführen.

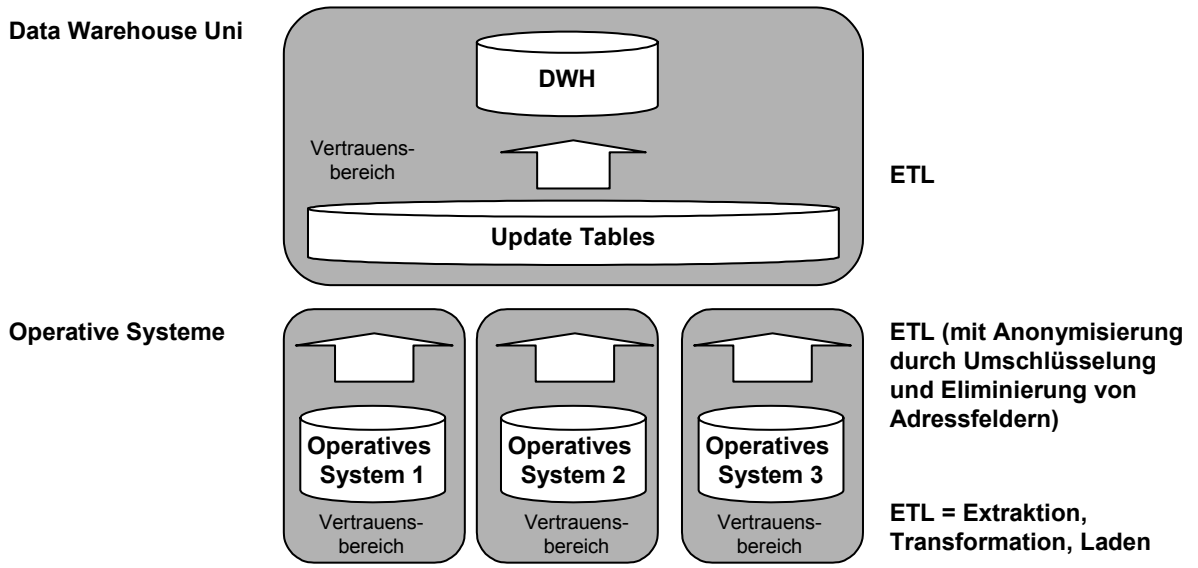

Bild 5: ETL im Vertrauensbereich des operativen Systems

Die Übergabe der Daten an CEUS erfolgt über klar definierte Schnittstellen, die zwischen den beteiligten Betreibern vereinbart werden. Beispielsweise basiert die Schnittstelle zur Übernahme der Daten aus den einzelnen Hochschulen in das landesweite Data-Warehouse auf den

gesetzlichen Bestimmung des Hochschulrahmengesetzes (HRG) und des Hochschulstatistikgesetzes (HStatG).

Ein befürchteter Durchgriff auf Data-Warehouses untergeordneter Ebenen ist durch die verteilte Architektur des Data-Warehouse-Systems ausgeschlossen. Dadurch bleibt die Autonomie der Teilsysteme gewahrt. Dies ist eine der wichtigsten Voraussetzungen für die Akzeptanz eines Data-Warehouse-Systems für das Hochschulwesen.

4.4.2 Übernahme von Daten aus den operativen Systemen in die Teil-Data-Warehouses der Hochschulen

Die Daten für die Teil-Data-Warehouses der Hochschulen werden aus den operativen Systemen der jeweiligen Hochschule extrahiert. Der Umfang der Daten ist über vertrauensbasierte Schnittstellen festgelegt, so dass die Betreiber der operativen Systeme die vollständige Information und die Kontrolle darüber haben, welche Daten in welcher Verdichtung übermittelt werden.

Für die Übernahme der Daten aus den operativen Systemen einer Hochschule in ein Teil-Data-Warehouse dieser Hochschule werden im Hinblick auf die faktische Anonymisierung (siehe Abschnitt 4.2) der Daten folgende Regeln beachtet:

- Es werden keine personenbezogenen Attribute zur externen Identifizierung (z.B. *Name, Adresse*) in CEUS übernommen.

- Attribute zur internen Identifikation, anhand derer eine Reidentifzierung von Personen möglich wäre (z.B. *Matrikelnummer* oder *Personalnummer*), werden verschlüsselt.

4.4.3 Übernahme von Daten aus der Amtlichen Statistik in das landesweite CEUS Data-Warehouse

Das landesweite Data-Warehouse enthält ausgewählte und zum Teil aggregierte Daten aus den Data-Warehouses der einzelnen Hochschulen. Die Vereinbarung über die Schnittstelle zwischen den Data-Warehouses der Hochschulen und dem landesweiten Data-Warehouse beruht auf der Spezifikation des Bayerischen Landesamts für Statistik und Datenverarbeitung. Diese Spezifikation gilt auf der Basis des Landesstatistikgesetzes als Zulässigkeitstatbestand (siehe Abschnitt 2.3) im Rahmen des Datenschutzes und dient zur Erstellung der Amtlichen Hochschulstatistik.

Statt die Daten für das landesweite Teil-Data-Warehouse direkt aus den Teil-Data-Warehouses der Hochschulen zu übernehmen, werden die Daten des Bayerischen Landesamts für Statistik und Datenverarbeitung verwendet. Die von den Hochschulen gemeldeten Daten werden vom Landesamt vor Aufnahme in die Amtliche Statistik mit Hilfe von Prüfroutinen auf Plausibilität geprüft. Hierzu werden unverschlüsselte Attribute zur internen Identifikation (z.B. *Matrikelnummer*) verwendet. Durch die vorgelagerte Qualitätssicherung durch das Lan-

desamt ist die Übernahme dieser Attribute in das landesweite Data-Warehouse nicht mehr erforderlich.

Längerfristig könnte die Datenübernahme direkt über die Data-Warehouses der Hochschulen erfolgen. In diesem Fall müssten jedoch ebenfalls verschlüsselte Attribute zur internen Identifikation (z.B. *Matrikelnummer*) eingesetzt werden. Die Daten aus den Data-Warehouses der Hochschulen werden dann in einer vorläufigen Form an das landesweite Data-Warehouse übermittelt. Nach Fertigstellung der Amtlichen Statistik durch das Bayerische Landesamt für Statistik und Datenverarbeitung werden diese durch die korrigierten Daten der Amtlichen Statistik überschrieben. Mögliche Abweichungen bis zur Aktualisierung durch das Landesamt müssen notwendigerweise toleriert werden. Der Vorteil dieser Lösung besteht darin, dass die Daten zeitlich vor der Veröffentlichung der Amtlichen Statistik zu Analysezwecken zur Verfügung stehen.

4.5 Entwicklung und Betrieb von CEUS

In der Entwicklungsphase wird CEUS als Forschungsprojekt betrieben. Das entwickelte Data-Warehouse-System wird als Systemprototyp im Pilotbetrieb am Bayerischen Staatsministerium für Wissenschaft, Forschung und Kunst und an den beiden Pilothochschulen, der Universität Bamberg und der Technischen Universität München, eingesetzt.

Zur Analyse des Informationsangebots ist der Einblick in die Datenhaltung der jeweiligen operativen Systeme und damit auch in personenbezogene Daten notwendig. Die alleinige Kenntnis der Datenstrukturen reicht erfahrungsgemäß nicht aus, da sich die Semantik der Attribute zum Teil erst aus den konkreten Datenwerten erschließt. Die Einsicht in die Datenhaltung der operativen Systeme erfolgt aber innerhalb des Vertrauensbereichs des jeweiligen operativen Systems. Eine externe Bereitstellung nicht anonymisierter personenbezogener Daten ist nicht erforderlich.

Auf der Grundlage der Datenquellenanalyse werden für die einzelnen operativen Systeme Extraktions- und Anonymisierungsschnittstellen entwickelt. Entsprechende Werkzeuge extrahieren die benötigten Daten über diese Schnittstellen aus den operativen Systemen und führen die faktische Anonymisierung der Daten durch Entfernung von Adressmerkmalen und Verschlüsselung von identifizierenden Merkmalen durch (siehe Abschnitt 4.2).

Die Entwicklung von CEUS erfolgt bereits unter Nutzung faktisch anonymisierter Daten. Die Entwicklungsarbeiten werden von der Projektgruppe Bamberg durchgeführt. Im Rahmen der Entwicklungsarbeiten werden angemessene Vorkehrungen zum Datenschutz getroffen. Hierzu gehören u.a. Aufstellung der Server in einem separaten, gesicherten Raum, Sicherung der Arbeitsplatzrechner gegen unbefugten Zugriff, Zugriff auf personenbezogene Daten während der Analyse des Informationsangebots nur für die Entwickler. Die Mitglieder des Projektteams sind aufgrund ihres Beschäftigungsverhältnisses im öffentlichen Dienst zur Geheimhaltung verpflichtet.

In der Betriebsphase wird CEUS an den Pilothochschulen und weiteren Hochschulen des Landes im produktiven Betrieb eingesetzt. Die Aufnahme der Betriebsphase setzt eine datenschutzrechtliche Prüfung und Freigabe des Systems voraus. Im Einzelnen gelten in der Betriebsphase folgende Verantwortlichkeiten:

- Der Betrieb der Extraktions- und Anonymisierungsschnittstellen erfolgt in der Verantwortung der Betreiber der operativen Systeme (siehe Abschnitt 4.4. „Vertrauensbereiche“). Die Datenübertragung zwischen den Vertrauensbereichen (d.h. vom Betreiber eines operativen Systems zum Data-Warehouse der Hochschule und von diesem zum Data-Warehouse des Landes) erfolgt auf geschützten Übertragungswegen.
- Der Betrieb der Data-Warehouse-Systeme der Hochschulen und des Landes erfolgt in deren Verantwortung. Der Nutzerzugriff auf das System erfolgt über das Internet (Darstellung in HTML über HTTP). Da typischerweise auf jedem Arbeitsplatzrechner ein WWW-Browser zur Verfügung steht, entfällt eine zusätzliche Softwareinstallation für den Nutzer. Dies fördert die Akzeptanz des Systems (siehe Abschnitt 1). Die Datenübertragung zwischen Client und Server erfolgt ebenfalls verschlüsselt.

5 Zusammenfassung und Ausblick

Im Zusammenhang mit Data-Warehouse-Systemen ist der Datenschutzproblematik grundsätzlich hohe Aufmerksamkeit zu widmen, da die dort angebotenen vielfältigen Navigationsmöglichkeiten in einem meist äußerst umfangreichen Datenbestand zur nahezu beliebigen Kombination von Einzeldaten herangezogen werden können. Dies gilt insbesondere für ein landesweites Data-Warehouse-System für das Hochschulwesen wie das System CEUS, bei dem u.a. die Interessen (teil-)autonomer Hochschulen mit den Interessen auf Landesebene abzustimmen sind.

Fragen des Datenschutzes und der Datensicherheit reichen bei CEUS weit über rechtliche Erfordernisse hinaus. Es ist davon auszugehen, dass ein derart umfassendes System nur dann effektiv betrieben und genutzt werden kann, wenn eine hohe Akzeptanz aller Beteiligten erreicht wird.

Voraussetzung für Akzeptanz ist Vertrauen. Wesentliches Ziel dieses Beitrags ist es, die konzeptuellen und konstruktiven Maßnahmen darzulegen, die in CEUS zur Erreichung von Datenschutz und Datensicherheit gewählt wurden.

CEUS ist ein „lebendes System“; es ist zu erwarten, dass sich aus der zunehmenden Nutzung des Systems immer wieder Anforderungen an dessen Weiterentwicklung ergeben werden. Fragen von Datenschutz und Datensicherheit werden daher das System CEUS auf seinem Lebensweg begleiten.

6 Literatur

[BiMR94] Biethahn, J.; Muksch, H.; Ruf, W.: Ganzheitliches Informationsmanagement. Band 1: Grundlagen, München et al., 3. Auflage, 1994.

[Bize99] Bizer, J.: Datenschutz im Data Warehouse – Die Verwendung von Kunden- und Nutzerdaten zu Zwecken der Marktforschung. In: Horster, P.; Fox, D. (Hrsg.): Datenschutz und Datensicherheit – Konzepte, Realisierungen, Rechtliche Aspekte, Anwendungen, Vieweg, Braunschweig, 1999, S. 60-81.

[BöUl00] Böhnlein, M.; Ulbrich-vom Ende, A.: Grundlagen des Data Warehousing: Modellierung und Architektur, Bamberger Beiträge zur Wirtschaftsinformatik Nr. 55, Bamberg, Februar 2000.

[Büll00] Büllesbach, A.: Datenschutz bei Data Warehouses und Data Mining. In: Computer und Recht, Heft 1, 16. Jg., 2000, S. 11-17.

[Burk00] Burkert, H.: Datenschutz. In: Jung, R.; Winter, R. (Hrsg.): Data Warehousing Strategie - Erfahrung, Methoden, Visionen. Springer, Heidelberg, 2000, S. 117-125.

[GoJa01] Gola P., Jaspers A.: Das neue BDSG im Überblick – Erläuterungen und Schaubilder für die Datenschutzpraxis. Gesellschaft für Datenschutz und Datensicherheit e.V., Datakontext-Fachverlag, Frechen 2001.

[Krüg90] Krüger, W.: Organisatorische Einführung von Anwendungssystemen. In: Kurbel, K.; Strunz H. (Hrsg.): Handbuch der Wirtschaftsinformatik. Stuttgart, Poeschel 1990, S. 275-288.

[Möll98] Möller, F.: Data Warehouse als Warnsignal an die Datenschutzbeauftragten. In: Datenschutz und Datensicherheit, 22. Jg., 10/1998, S. 555-560.

[Mönc98] Möncke, U.: Data Warehouses – eine Herausforderung für den Datenschutz? In: Datenschutz und Datensicherheit, 22. Jg., 10/1998, S. 561-569.

[Mönc99] Möncke, U.: Sicherheit im Data Warehouse – Profilbildung und Anonymität. In: Horster, P.; Fox, D. (Hrsg.): Datenschutz und Datensicherheit – Konzepte, Realisierungen, Rechtliche Aspekte, Anwendungen, Vieweg, Braunschweig, 1999, S. 30-59.

[PiRW98] Picot, A.; Reichwald, R.; Wigand, R.T.: Die grenzenlose Unternehmung - Information, Organisation, Management, 3. Auflage, Gabler, Wiesbaden, 1998.

[SBPU01] Sinz, E.J.; Böhnlein, M.; Plaha, M.; Ulbrich-vom Ende, A.: Architekturkonzept eines verteilten Data Warehouse-Systems für das Hochschulwesen, in: Proceedings of Wirtschaftsinformatik 2001, (WI 2001, Augsburg, 19.-21. September), 2001.

[SBPU02] Sinz, E.J.; Böhnlein, M.; Plaha, M.; Ulbrich-vom Ende, A.: Leitlinien für den Datenschutz im Projekt CEUS. Version 1.9. Unveröffentlicht, Universität Bamberg 2002.

[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.

[Sinz02] Sinz, E.J.: Data Warehouse. In: Küpper H.-U., Wagenhofer U. (Hrsg.): Handwörterbuch Unternehmensrechnung und Controlling. 4., neu gestaltete Auflage, Schäffer-Poeschel, Stuttgart 2002, S. 313 - 318