Betriebskonzept des Digitalen Monumentalbau-Archivs MonArch

Alfons Ruch, Alexander Stenzer, Burkhard Freitag

Lehrstuhl für Informationsmanagement, Universität Passau

Alfons.Ruch@uni-passau.de

Alexander.Stenzer@uni-passau.de

Burkhard.Freitag@uni-passau.de

Technischer Bericht, Nummer MIP-0911
Fakultät für Informatik und Mathematik
Universität Passau, Deutschland
August 2009

Inhaltsverzeichnis

1 Einleitung

Das Digitale Monumentalbau-Archiv (DMA) ist ein im Rahmen des DFG geförderten Projekts MonArch[1] entstandenes Archivsystem, für jegliche Art von Dokumenten, die im Bereich von Monumentalbauten archiviert werden sollen.

Kapitel 2 beschreibt die Architektur des Digitalen Monumentalbau-Archivs. In Kapitel 3 wird eine kurze Übersicht über das Datenmodell gegeben. Kapitel 4 enthält das verwendete Zugriffskonzept, um anschließend in Kapitel 5 kurz auf die Installation des DMA einzugehen.

2 Architektur

Das Digitale Monumentalbau-Archiv (DMA) setzt eine klassische Client-Server Architektur ein. Diese ermöglicht einen Mehrbenutzerbetrieb und etabliert eine klare Trennung zwischen interaktiver Nutzung und Visualisierung der Daten (Client) einerseits und der eigentlichen Datenhaltung (Server) andererseits.
Die Grobarchitektur des DMA ist in Abbildung 1 zu sehen.

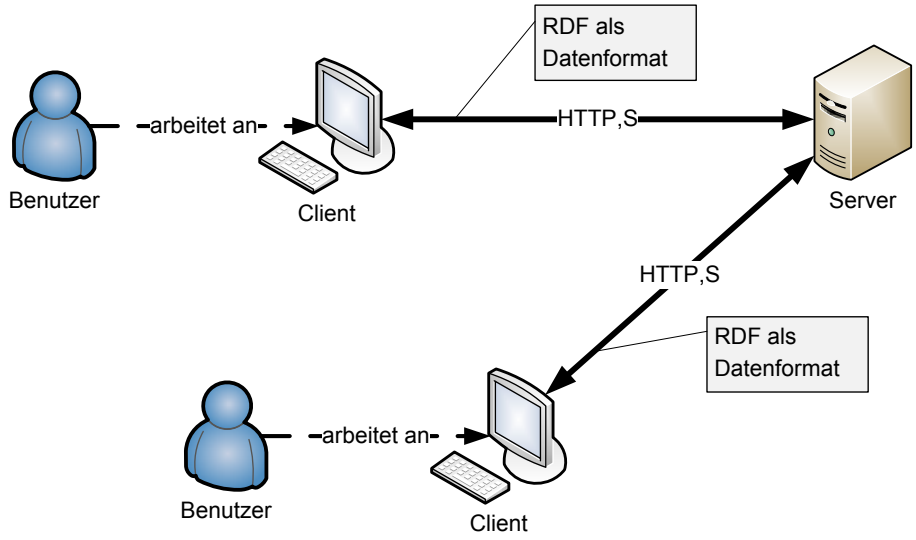

Abbildung 1: Client-Server Architektur des DMA

Als Kommunikationsprotokoll zwischen Client und Server wird das Hypertext Transfer Protocol (HTTP) [W3C99] eingesetzt.

Die Wahl von HTTP hat für den Betrieb des DMA folgende Vorteile:

- Der Zugriff vom Client auf den Server kann mit HTTP auch über das Internet erfolgen, ohne dass hierfür das Archiv zusätzlich angepasst werden müsste. Bei Verwendung von Firewalls ist der HTTP-Transfer meistens

[1]Dieses Projekt wird gefördert von der Deutschen Forschungsgemeinschaft (DFG) unter dem Aktenzeichen FR 1012/8-1.

freigeschaltet, so dass im Normalfall keine Anpassungen notwendig sein dürften.

- Der Zugriff auf das DMA über einen HTTP-Proxyserver ist problemlos möglich.
- Mit der HTTP-Authentifizierung wird eine standardisierte Methode zur Benutzerauthentisierung bereitgestellt.
- Mit der HTTP-Erweiterung HTTPS (Hypertext Transfer Protocol Secure) steht eine standardisierte Methode für einen verschlüsselten und abgesicherten Datentransfer zur Verfügung.

Innerhalb des HTTP-Datenstroms werden alle Metadaten (z.B. Name des Dokuments, Autor des Dokuments, etc.) im RDF-Format [W3C04] übertragen.

RDF bietet Vorteile gegenüber proprietären Datenformaten für Metadaten:

- RDF ist ein W3C Standard.
- Bei einer Erweiterung des Metadaten-Schemas im Archiv muss das Datenformat nicht angepasst werden, weil RDF aufgrund seines Informationsmodells flexibel mit Schemaerweiterungen umgehen kann.
- Die Metadaten können von Webservices ebenfalls in einem RDF-Format (genauer: RDF/XML) bereitgestellt werden.
- Es existieren Programmbibliotheken für die Bearbeitung und Speicherung von RDF-Daten.

Genauere Informationen zur Speicherung der Metadaten und RDF sind in Abschnitt 3 zu finden.

2.1 DMA-Client

Die DMA-Client Komponente wurde in Java [SM09a] entwickelt. Durch die Verwendung von Java als Programmiersprache ist der DMA-Client unabhängig vom Betriebssystem ausführbar. Der DMA-Client kann lokal auf einem Arbeitsplatz-PC ausgeführt werden, ist aber auch in Terminal Server Umgebungen (z.B. Microsoft Terminal Server [Mic08] oder Citrix XenApp [Sys09]) einsetzbar. Der DMA-Client kann sogar im Browser über WebStart [SM01] ausgeführt werden. Für Java gibt es eine Vielzahl freier Bibliotheken, die wiederverwendet werden können. Dies reduziert einerseits den Entwicklungsaufwand, andererseits kann der Client von Weiterentwicklungen der Bibliotheken profitieren. Gerade bei sicherheitsrelevanten Funktionen kann der Einsatz freier Bibliotheken die Sicherheit erhöhen, da eine breite Benutzergemeinschaft diese Funktionen nutzt und so Sicherheitslücken früher entdeckt und beseitigt werden.
Durch ein Anmeldeverfahren beim Starten des Clients wird die erste Phase der Zugriffskontrolle realisiert. Verschiedene Benutzer können sich mit einer persönlichen Kennung am Server anmelden. Transaktionale Anfragebearbeitung und ein Sperrkonzept ermöglichen einen konkurrierenden Zugriff auf das Archivsystem.

2.2 DMA-Server

Der DMA-Server besteht konzeptionell aus den vier Komponenten Anwendungsserver, Datenbankserver, Benutzerverwaltung und Dateiserver (siehe Abbildung 2). Die Funktion des Dateiservers wird von dem jeweiligen Betriebssystem übernommen. Die ersten drei Komponenten werden in den folgenden Abschnitten näher beschrieben.

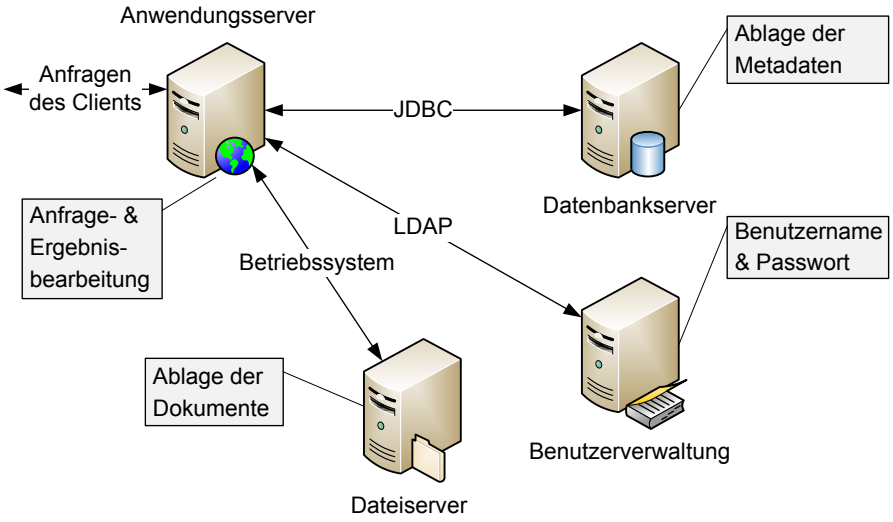

Abbildung 2: Übersicht der Server Komponenten des DMA

2.2.1 Anwendungsserver

Als Anwendungsserver kommt der Apache Tomcat Server in der Version 6.0 [ASF09] zum Einsatz. Er übernimmt dabei folgende Aufgaben:

- **Authentifizierung der Benutzer:** Meldet sich ein Benutzer am Client an, wird diese Anmeldung über eine HTTP-Authentifizierung an den Anwendungsserver delegiert. Dieser wiederum delegiert die Authentifizierung an die Benutzerverwaltung weiter. Ist die Überprüfung der Benutzerdaten bei der Benutzerverwaltung erfolgreich, wird der Zugang zum Archiv gewährt.
- **Überprüfung der Rechte:** Beim Zugriff auf ein Datenobjekt (z.B. ein Dokument) überprüft der Anwendungsserver die Zugriffsrechte. Die Überprüfung schützt den Datenbestand des Archivs vor unbefugtem Zugriff. Diese Zugriffsrechte können frei definiert und geändert werden (siehe dazu Abschnitt 4).
- **Abruf und Speicherung der Metadaten:** Sollen Metadaten aus dem Archiv ausgelesen werden, generiert der Anwendungsserver die entsprechenden Anfragen an den Datenbankserver und wandelt die Antworten in RDF um. Anschließend überträgt er die Daten an den Client per HTTP.

Werden Metadaten am Client verändert, werden diese ebenfalls vom Anwendungsserver aufbereitet und an den Datenbankserver geschickt, um sie dort abzuspeichern.

- **Abruf und Speicherung der Dokumente:** Wird ein Dokument abgerufen oder ins Archiv eingefügt, delegiert der Anwendungsserver den Aufruf an das Dateisystem des Betriebssystems.
- **Sperrkontrolle:** Der Anwendungsserver verwaltet die Sperren im System. Damit wird verhindert, dass zwei Personen dasselbe Objekt gleichzeitig bearbeiten.
- **Transaktionskontrolle:** Durch eine transaktionale Abarbeitung von Anwenderaktionen stellt der Anwendungsserver die Konsistenz der im Archiv gespeicherten Daten sicher.

2.2.2 Datenbankserver

Der Datenbankserver ist für die Speicherung der Metadaten verantwortlich. Er kommuniziert über eine JDBC-Schnittstelle [SM09b] mit dem Anwendungsserver. Als Datenbankserver kommt der Microsoft SQL Server in der Version 2005 [Mic05] zum Einsatz. Die Speicherung der RDF-Daten erfolgt analog zur Methode des Jena RDF Frameworks [McB01]. Alle Statements werden dabei in einer Statements-Tabelle abgelegt. Zur Effizienzsteigerung werden zusätzlich Indextabellen verwaltet, die die Beantwortung bestimmter, häufig auftretender Anfragen ohne Zugriff auf die Statements-Tabelle ermöglichen.
Der Zugriff auf den Datenbankserver wird durch spezielle Datenbankbenutzer abgesichert, die nicht in der Benutzerverwaltung, sondern direkt im Datenbankserver angelegt sind. Damit wird eine Trennung der Benutzer des Archivs (Endnutzer) und der Benutzer der Datenbank (administrative Nutzer) erreicht. Dieses Prinzip wird als *separation of duty* [BG09] bezeichnet. Bei den Datenbankbenutzern wird zwischen verschiedenen Benutzertypen und den dazugehörigen Rechten unterschieden.

- **DMA_READ:** Dieser Benutzer darf nur lesend auf das Archiv zugreifen und hat keinerlei Schreibrechte. Auf diese Kennung wird bei jeder lesenden Aktion zurückgegriffen. Bevor ein Zugriff auf Metadaten für einen Endbenutzer des DMA erlaubt wird, werden die Zugriffsrechte mit Hilfe des DMA_RIGHTS Datenbankbenutzers überprüft.
- **DMA_WRITE:** Dieser Benutzer darf zusätzlich die Metadaten der Dokumente verändern. Auch hier erfolgt zuerst eine Rechteüberprüfung.
- **DMA_MANAGE:** Im Gegensatz zu DMA_WRITE darf dieser Benutzer zusätzlich die Gebäudestruktur und den Themenkatalog verändern. Auch hier erfolgt zuerst eine Rechteüberprüfung.
- **DMA_RIGHTS:** Dieser Benutzer ist ausschließlich für die Überprüfung von Zugriffsrechten vorgesehen. Der Benutzer kann nicht auf die Metadaten, die in der Datenbank gespeichert sind, zugreifen, sondern nur auf die Rechte, welche für die Objekte im Archiv vergeben sind.

2.2.3 Benutzerverwaltung

Für die Komponente der Benutzerverwaltung kann zwischen zwei Ansätzen gewählt werden.

- **Einsatz der Benutzerverwaltung des Apache Tomcat Servers:** Diese Variante sollte gewählt werden, wenn eine separate Benutzerverwaltung erwünscht ist oder keine externe Benutzerverwaltung existiert, die integriert werden könnte.
- **Anbindung einer externen Benutzerverwaltung über das LDAP Protokoll:** Dieser Ansatz ermöglicht die Einbeziehung des Digitalen Monumentalbau-Archivs in eine bestehende Benutzerverwaltung. Die externe Benutzerverwaltung muss das LDAP Protokoll unterstützen, wie es z.B. von Microsoft Active Directory oder dem OpenLDAP Directory bereitgestellt wird. Durch diese Art der Anbindung wird eine doppelte Benutzerverwaltung vermieden, wodurch der Administrationsaufwand verringert werden kann.

2.3 Verteilungskonzept

Das DMA kann sowohl auf einem zentralen Server als auch als verteiltes System betrieben werden. Im ersten Fall werden alle beschriebenen Komponenten auf einem einzelnen Rechner installiert. Im zweiten Fall können die Komponenten beliebig auf mehrere Rechner verteilt werden. Durch diese Flexibilität ist die Integrierbarkeit des DMA in beliebige bestehende Systemumgebungen gewährleistet.

3 Datenmodell

Das in diesem Abschnitt dargestellte, etwas vereinfachte DMA-Datenmodell dient als Grundlage für die folgenden Abschnitte. Abbildung 3 zeigt das zugehörige ER-Diagramm.

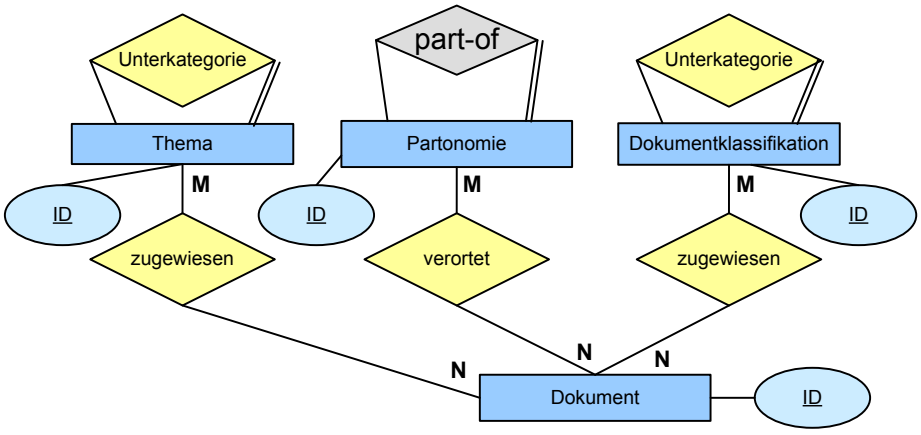

Abbildung 3: Vereinfachtes DMA-Datenmodell

Die folgenden Objekte bilden das DMA-Datenmodell:

- **Dokument**: Das zentrale Objekt im Archiv ist das Dokument. Ein Dokument besteht dabei aus einer beliebigen Datei sowie Metadaten (u.a. Autor, Kommentartext, Erstellungsdatum), die das Dokument beschreiben. Bei Monumentalbauten handelt es sich bei Dokumenten hauptsächlich um Pläne, Bilder und Textdokumente.
- **Partonomie**: Dokumente werden einer Gebäude-Partonomie zugeordnet, d.h. einer Repräsentation der Zerfallsstruktur des betrachteten Gebäudes. Die Partonomie wird als Monohierarchie entwickelt.
- **Thema**: Einem Dokument können zusätzlich ein oder mehrere Themen zugeordnet werden. Damit eröffnen Themen eine Möglichkeit zur semantischen Auszeichnung und bilden zugleich ergänzende Selektionskriterien für das Wiederauffinden (Retrieval) der Dokumente.
- **Dokumentklassifikation**: Neben den Themen gibt es weitere Dokumentklassifikationen und -beschreibungen, die technische Eigenschaften oder Entstehungs- und Erfassungsinformationen des Dokuments beinhalten, z.B. Dokumenttyp oder Einstellungsdatum.

4 Zugriffskontrolle

Im Folgenden wird das im DMA eingesetzte Konzept der Zugriffskontrolle definiert, das sich stark an [BG09] orientiert. Die einzelnen Bestandteile sind in Abbildung 4 (vgl. [BG09]) dargestellt.

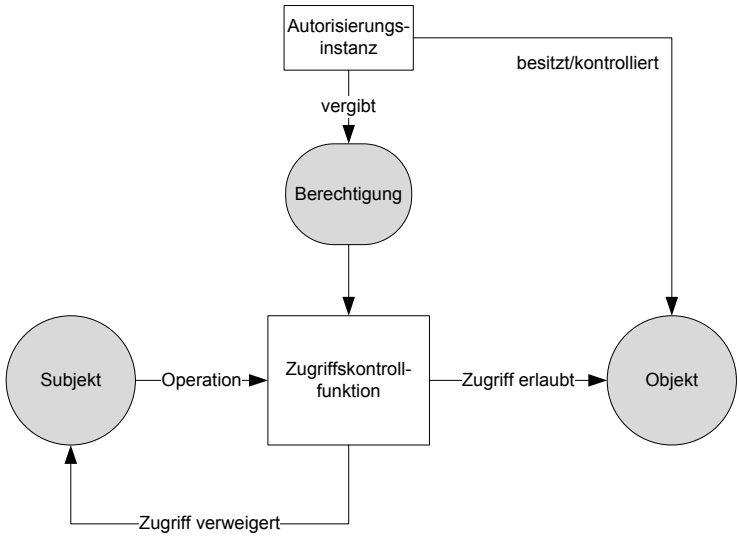

Abbildung 4: Grundprinzip der Autorisierung und Zugriffskontrolle

4.1 Berechtigung

Im DMA werden Berechtigungen über Rollen in Form von Zugriffssteuerungslisten (ACL; Access Control List) vergeben. Der Zugriff auf ein Objekt im DMA wird über drei Berechtigungsstufen definiert.

- **Lesen:** Das *Lese*-Recht auf ein Dokument im Archiv ermöglicht das Öffnen und Anzeigen der Metadaten des Dokuments. Ein Dokument wird bei einer Suche nur dann in die Ergebnismenge übernommen, wenn der Benutzer das *Lese*-Recht auf dieses Dokument besitzt. Für die Partonomie, Themen und Dokumentklassifikationen ermöglicht das *Lese*-Recht das Anzeigen dieser Elemente im DMA.
- **Schreiben:** Das *Schreib*-Recht erweitert das *Lese*-Recht um folgende Berechtigungen: Ein Dokument kann mit dem *Schreib*-Recht gelöscht und dessen Metadaten verändert werden. Für die Partonomie, Themen und Dokumentklassifikationen ermöglicht das *Schreib*-Recht zusätzlich das Erweitern, das Löschen und die Veränderung (z.B. Umbenennen oder Verschieben).
- **Besitz:** Das *Besitz*-Recht erlaubt zusätzlich zum Lese- und Schreibzugriff auch das Ändern der Rechte für ein Objekt (Partonomie, Themen, Dokumente, etc.).

4.2 Autorisierungsinstanz - DMA-Rollen

Das DMA besitzt vier vordefinierte Rollen:

- **DMA-Administratoren:** Die Rolle der DMA-Administratoren ist die Rolle mit dem umfassendsten Zugriff auf die Daten im DMA. Diese Rolle hat für jedes Objekt im Archiv das *Besitz*-Recht. Beim Anlegen eines Objekts wird dieses Recht automatisch zu der ACL hinzugefügt und kann auch nicht mehr entfernt werden. Zusätzlich können Mitglieder dieser Rolle neue Rollen anlegen und Benutzer oder Gruppen einer Rolle zuweisen.
- **DMA-Verwalter:** Die Rolle DMA-Verwalter kann keine Rechte vergeben, hat aber Schreibzugriff auf jedes Objekt im Archiv. Beim Anlegen eines Objekts wird dieses Recht automatisch zu der ACL hinzugefügt. Ein Löschen dieses ACL-Eintrags ist jederzeit möglich.
- **DMA-Benutzer:** Die Rolle DMA-Benutzer hat einen Lesezugriff auf jedes Objekt im Archiv. Beim Anlegen eines Objekts wird dieses Recht automatisch zu der ACL hinzugefügt. Ein Löschen dieses ACL-Eintrags ist jederzeit möglich.
- **DMA-Gäste:** Die Rolle DMA-Gäste hat keinerlei Rechte im DMA. Diese Rolle muss manuell in der ACL der betroffenen Objekte eingetragen werden. Daher eignet sich diese Rolle, um für bestimmte Bereiche des Archivs einen Gastzugang einzurichten.

Rechte im DMA werden nur auf Basis von Rollen vergeben. Neue Rollen können jederzeit angelegt werden. Um eine Verbindung zwischen Rechteverwaltung und Benutzerverwaltung herzustellen, müssen Benutzer oder Gruppen aus der Benutzerverwaltung den DMA-Rollen zugewiesen werden. Das Konzept der rollenbasierten Zugriffskontrolle wurde in [SCFY96] vorgestellt.

4.3 Autorisierungssubjekte

Autorisierungssubjekte sind im DMA die Benutzer. Benutzer werden in der Benutzerverwaltung (siehe Abschnitt 2.2.3) angelegt. Jeder Benutzer muss sich beim Anmelden durch seine Kennung (login) und sein Passwort authentifizieren. Einem Benutzer kann ein Zugriffsrecht nur indirekt erteilt werden, indem ihm eine DMA-Rolle zugewiesen wird. Ein Benutzer kann mehrere Rollen besitzen. Sollte es hierbei zu Rechtekonflikten zwischen verschiedenen Rollen kommen, wird das höchste Recht gewählt.
Wenn ein Benutzer durch seine Rollen für ein Autorisierungobjekt (siehe Abschnitt 4.4) keine Berechtigung besitzt, wird ihm bzw. ihr der Zugriff verweigert.

4.4 Autorisierungsobjekte

Im Folgenden werden Autorisierungsobjekte in Anlehnung an [BG09] definiert (siehe auch Abbildung 5). Im DMA werden drei Kategorien von Autorisierungsobjekten verwendet: Dokumente, Dimensionen und Schema.

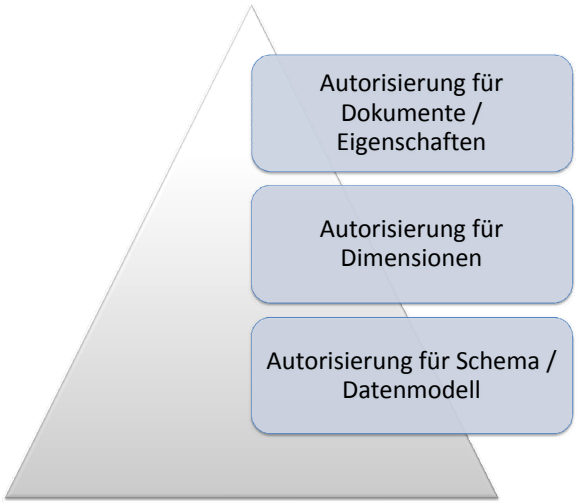

Abbildung 5: Autorisierungshierarchie

- **Dokumente:** Zugriffsrechte für Dokumente umfassen das *Lese*-Recht als schwächstes Recht und das *Schreib*-Recht, welches das *Lese*-Recht einschließt. Ein Autorisierungssubjekt, das für ein bestimmtes Dokument weder *Schreib*- noch *Lese*-Recht besitzt, kann auf dieses Dokument nicht zugreifen. Insbesondere ist in diesem Fall das Dokument für dieses Autorisierungssubjekt nicht sichtbar.
- **Dimensionen:** Bei der Suche (Retriveal) nach Dokumenten können sowohl die Partonomie als auch die Themen und die Dokumentklassen, denen ein Dokument zugeordnet ist, dazu dienen, die Ergebnismenge genauer zu spezifizieren bzw. einzuschränken. Als "Dimension"werden im

Folgenden daher zusammenfassend die Partonomiezuordnung, die Themenzuordnung und die Dokumentklassifikation bezeichnet. Zugriffsrechte für Dimensionen umfassen das *Lese*-Recht als schwächstes Recht und das *Schreib*-Recht, welches das *Lese*-Recht einschließt. Ein Autorisierungssubjekt, das für eine bestimmte Dimension weder *Schreib*- noch *Lese*-Recht besitzt, kann auf diese Dimension nicht zugreifen. Insbesondere ist in diesem Fall die Dimension für dieses Autorisierungssubjekt nicht sichtbar und kann nicht für die Suche verwendet werden.

- **Schema:** Das Schema des DMA kann nicht verändert werden. Ein Autorisierungssubjekt hat keinen direkten Zugriff auf das Schema der Datenbank, da alle Zugriffe auf die Datenbank der Anwendungsserver durchführt. Selbst der Anwendungsserver kann das Schema nicht verändern, da den Datenbankbenutzern (siehe Abschnitt 2.2.2) dieses Recht nicht gewährt wird..

4.5 Zugriffskontrollfunktion

Die Zugriffskontrollfunktion wird im DMA auf drei Schichten realisiert (siehe Abbildung 6).

	Verantwortlich	Aktion
Anmeldung	 Benutzerverwaltung	**Überprüfung der Benutzerdaten wird an Benutzerverwaltung delegiert.**
Zugriff auf Objekte	 Anwendungsserver	**Der Anwendungsserver überprüft bei Zugriff auf ein Objekt, ob in der ACL des Objektes eine Rolle das entsprechende Recht hat.**
Zugriff auf Daten in der DB	 Datenbankserver	**Der Datenbankserver überprüft bei Zugriff auf die Datenbank die Zugriffsrechte anhand der Datenbankbenutzer.**

Abbildung 6: Schichten der Zugriffskontrolle des DMA

1. **Autorisierung der Benutzer:** Meldet sich ein Benutzer am DMA an, wird vom Anwendungsserver Benutzername und Passwort überprüft, indem er diese Aufgabe an die Benutzerverwaltung delegiert (siehe Abschnitt 2.2.1).

2. **Zugriff auf Objekte:** Versucht ein erfolgreich angemeldeter Benutzer einen Zugriff auf ein Objekt (z.B. ein Dokument) durchzuführen, überprüft der Anwendungsserver, ob die Rollen, denen dieser Benutzer angehört, die nötigen Rechte besitzten. Dazu werden die ACL des Objekts herangezogen.

3. **Zugriff auf die Daten in der DB:** Nachdem der Anwendungsserver die Zugriffsrechte überprüft hat, erfolgt der Zugriff auf die Datenbank. Je nach Zugriffsrecht (Lese- oder Schreibzugriff) wird der entsprechende Datenbankbenutzer für den Zugriff auf die Daten ausgewählt (siehe Abschnitt 2.2.2).

5 Installation

Für den Betrieb des DMA sind die folgenden Voraussetzungen erforderlich:
Hardware Minimalanforderung:

- PC ab 1 GHz Taktfrequenz, 256 MB RAM, 500 MB freier Festplattenspeicher
- Display ab 1024 x 768 Bildpunkte

Hardware Empfehlung:

- PC ab 2 GHz Taktfrequenz, 512MB RAM, 1 GB freier Festplattenspeicher
- Display ab 1280 x 1024 Bildpunkte

Software:

- Betriebssystem Windows 2000, Windows XP, Windows Vista, Windows Server 2003 oder Windows Server 2008
- Sun Java Runtime Environment (JRE) 1.6.11
- Apache Tomcat 6.0.18
- Microsoft SQL Server 2005 Express Edition
- Microsoft SQL Server JDBC Driver 1.2
- Microsoft Visual C Runtime 7.1 DLL

Eine ausführliche Installationsanleitung ist auf der Installations-CD zu finden, weshalb hier nicht näher auf die Installation des DMA eingegangen wird.

Literatur

[ASF09] The Apache Software Foundation. Apache tomcat 6.0. http://tomcat.apache.org, July 2009.

[BG09] Bauer and Günzel. *Data Warehouse Systeme*. dpunkt.verlag, 2009.

[McB01] Brian McBride. Jena: Implementing the rdf model and syntax specification. In *Semantic Web Workshop, WWW*, 2001.

[Mic05] Microsoft. Sql server 2005. http://www.microsoft.com/germany/sql, 2005.

[Mic08] Microsoft. Terminal services. http://www.microsoft.com/windowsserver2008, 2008.

[SCFY96] Ravi S. Sandhu, Edward J. Coyne, Hal L. Feinstein, and Charles E. Youman. Role-based access control models. *IEEE Computer*, 29(2):38–47, 1996.

[SM01] Java Software: A Division of Sun Microsystems. *Java(TM) Network Launching Protocol & API Specification (JSR-56): Version 1.5*. Sun Microsystems, 21. May 2001.

[SM09a] Inc. Sun Microsystems. Java. http://java.sun.com/, July 2009.

[SM09b] Inc. Sun Microsystems. The java database connectivity (jdbc). http://java.sun.com/javase/technologies/database/, July 2009.

[Sys09] Citrix Systems. Citrix xenapp. http://www.citrix.de/produkte/schnellsuche/xenapp/, July 2009.

[W3C99] W3C. Http - hypertext transfer protocol, http://www.w3.org/protocols/, 1999.

[W3C04] W3C. Resource description framework (rdf), http://www.w3.org/rdf/, 2004.