

MyCoRe User Guide

Frank Lützenkirchen

Jens Kupferschmidt

Detlev Degenhardt

Johannes Bühler

Ulrike Krönert

Ute Starke

Andreas Trappe

Kathleen Krebs

Release 1.1

21. Juni 2005

Inhaltsverzeichnis

1. Voraussetzungen für eine MyCoRe Anwendung.....	8
1.1 Vorbemerkungen.....	8
1.2 Umgebungs- und Softwarevoraussetzungen.....	10
1.3 Hinweise zu den OpenSource-Komponenten.....	11
1.3.1 Installation auf Linuxsystemen.....	11
1.3.1.1 Java.....	11
1.3.1.2 MySQL.....	11
1.3.1.3 Lucene.....	12
1.3.1.4 Tomcat.....	13
1.3.1.5 Apache.....	13
1.3.2 Installation auf MS Windows-Systemen.....	14
1.3.2.1 Java.....	14
1.3.2.2 MySQL.....	14
1.3.2.3 Lucene.....	15
1.3.2.4 Apache.....	15
1.3.2.5 Tomcat.....	15
1.4 IBM Content Manager 8.2 als Backend.....	16
2. Download und Installation des MyCoRe Kerns.....	17
2.1 Download des MyCoRe Kerns.....	17
2.2 Konfiguration und Übersetzen des Kerns.....	18
3. Die MyCore Beispielanwendung MyCoReSample.....	20
3.1 Grundlegender Aufbau und Ziel der Beispielanwendung.....	20
3.1.1 Allgemeines.....	20
3.1.2 Weiterführende Erläuterungen.....	20
3.2 Download der Beispielanwendung.....	21
3.3 Konfiguration im Anwendungsteil DocPortal.....	21
3.3.1 Allgemeines.....	21
3.3.2 Installation.....	21
3.4 Konfiguration im Anwendungsteil MyCoReSample.....	22
3.4.1 Grundlegendes zur Konfiguration.....	22
3.4.1.1 CM8 Store konfigurieren.....	22
3.4.1.2 eXist Store Konfigurieren.....	23
3.4.1.3 JDBC-Treiber konfigurieren.....	23
3.4.1.4 Die SQL-Tabellennamen.....	23
3.4.1.5 User-Management.....	23
3.4.1.6 Remote-Access.....	23
3.4.1.7 Speicherung von Objekten und Dokumenten konfigurieren.....	24
3.4.1.7.1 File System Store.....	24
3.4.1.7.2 Lucene Store.....	24
3.4.1.8 Editor.....	24
3.4.1.9 Workflow.....	24
3.4.1.10 Appletsignatur.....	24
3.4.1.11 WCMS (Web Content Management System).....	25
3.4.1.12 Debug konfigurieren.....	25
3.4.2 Initialisieren des MyCoReSamples.....	25

3.4.3 Erzeugen der Skripte mycore.sh / mycore.cmd.....	26
3.4.4 Aufruf der MyCoRe-Kommandozeile.....	26
3.5 Erzeugen und Konfigurieren der Web-Anwendung.....	27
3.5.1 Erzeugen der Web-Anwendung.....	27
3.5.2 Konfiguration des Web Application Server.....	27
3.5.2.1 Tomcat.....	27
3.5.2.2 Websphere.....	28
3.6 Laden der Beispieldaten.....	28
4.Vom Sample zum eigenen Dokumenten-Server.....	29
4.1 Allgemeines.....	29
4.2 Die Installation der freien XML:DB eXist.....	30
4.2.1 Allgemeines.....	30
4.2.2 Vorbedingungen.....	30
4.2.3 Installation.....	30
4.2.4 Konfigurationsänderungen in den MyCoRe-Teilen.....	31
4.2.5 Füllen des Search-Backends für eXist.....	31
4.3 Die Zusammenarbeit mit anderen MyCoReSample-Installationen.....	31
4.3.1 Die eigene Installation.....	32
4.3.2 Weitere Server benachbarter Einrichtungen.....	32
4.3.3 Standard-Server des MyCoRe-Projektes.....	32
4.4 Einbindung virtueller Host-Namen mit Hilfe des Apache-Web-Server.....	32
4.4.1 Installation von httpd-2.0.54 mit Einbindung von mod_proxy.....	33
4.4.2 Die Verbindung von Tomcat5 und Apache2.....	33
4.5 Nutzung der OAI Schnittstelle.....	34
4.5.1 Grundlagen.....	34
4.5.2 Der OAI Data Provider.....	35
4.5.3 Installation.....	35
4.5.4 Der Deployment Descriptor.....	35
4.5.5 Die mycore.properties.oai.....	36
4.5.6 Instanzunabhängige Properties.....	36
4.5.7 Instanzabhängige Properties.....	36
4.5.8 Test.....	37
4.5.9 Zertifizierung und Registrierung.....	37
5.Weiterführende Informationen zum Aufbau von MyCoRe-Anwendungen.....	38
5.1 XML-Syntax des Datenmodells.....	38
5.1.1 Die MCRObjectID.....	38
5.1.2 Das Klassifikationen-Datenmodell.....	39
5.1.3 Das Metadatenmodell.....	40
5.1.3.1 XML-Syntax eines Metadatenobjektes.....	41
5.1.3.2 XML-Syntax des XML-Knotens structure.....	41
5.1.3.3 XML-Syntax des XML-Knotens metadata.....	42
5.1.3.4 MyCoRe Metadaten-Basistypen	43
5.1.3.4.1 XML-Syntax des Metadatentyps MCRMetaAddress.....	44
5.1.3.4.2 XML-Syntax des Metadatentyps MCRMetaBoolean.....	45
5.1.3.4.3 XML-Syntax des Metadaten-Basistyps MCRMetaClassification.....	46
5.1.3.4.4 XML-Syntax des Metadaten-Basistyps MCRMetaDate.....	46
5.1.3.4.5 XML-Syntax des Metadatentyps MCRMetaInstitutionName.....	47

5.1.3.4.6 XML-Syntax des Metadatentyps MCRMetaISBN.....	47
5.1.3.4.7 XML-Syntax des Metadatentyps MCRMetaLangText.....	48
5.1.3.4.8 XML-Syntax der Metadatentypen MCRMetaLink und MCRMetaLinkID.....	48
5.1.3.4.9 XML-Syntax des Metadatentyps MCRMetaNBN.....	49
5.1.3.4.10 XML-Syntax des Metadatentyps MCRMetaNumber.....	50
5.1.3.4.11 XML-Syntax des Metadatentyps MCRMetaPersonName.....	50
5.1.3.4.12 XML-Syntax des Metadatentyps MCRMetaXML.....	52
5.1.3.5 XML-Syntax des XML-Knotens service.....	53
5.1.4 Das Speichermodell für die Multimediadaten (IFS).....	54
5.2 Die Verwendung der Kommandozeilenschnittstelle	56
5.2.1 Basiskommandos des Kommandozeilensystems.....	56
5.2.2 Kommandos zur Arbeit mit der Benutzerverwaltung	56
5.2.2.1 Allgemeine Verwaltungskommandos.....	57
5.2.2.2 Kommandos zum Arbeiten mit XML-Dateien.....	57
5.2.2.3 Kommandos zum direkten Arbeiten mit Objekten der Benutzerverwaltung.....	58
5.2.2.4 Das Sichern und Restaurieren der Benutzerverwaltungsdaten.....	59
5.2.3 Kommandos zur Arbeit mit den Daten.....	59
5.2.3.1 Kommandos zum Verwalten von Klassifikationen.....	60
5.2.3.2 Kommandos zur Verwaltung der Objekte.....	60
5.2.3.3 Anfragen an das System per Kommandozeilen.....	61
5.2.3.4 Sonstige Kommandos.....	61
5.3 Das SimpleAccessControl-System.....	61
5.3.1 Benutzer und Privilegien.....	62
5.3.2 Zugriffsdaten eines Objektes.....	63
5.3.2.1 Das Lese-Zugriffs-Privileg.....	64
5.3.2.2 Das Status-Privileg.....	64
5.3.2.3 Das IP-Flag.....	64
5.3.2.4 Die Benutzer.....	64
5.3.3 Testmechanismen.....	64
5.3.4 Die konkrete Umsetzung im MyCoReSample.....	64
5.3.4.1 Administrative Privilegien.....	64
5.3.4.2 Privilegien zur Bearbeitung der Daten.....	65
5.3.4.3 Lese-Privilegien.....	66
5.3.4.4 Benutzer und Gruppen des Beispiels.....	67
5.3.4.5 Übersicht der Zusammenhänge von Benutzern, Gruppen und Privilegien im MyCoReSample.....	67
5.4 Das SimpleWorkflow-System zur interaktiver Autorensarbeit.....	68
5.4.1 Das MCRStartEditorServlet.....	69
5.4.2 Abläufe für neue Datenobjekte.....	72
5.4.3 Abläufe für Datenobjekte aus dem Server.....	72
5.4.4 Einbindung in das SimpleAccessControl-System.....	72
6.Hints & Tips / Troubleshooting.....	73
6.1 Eine zweite Instanz von eXist auf einem System einrichten.....	73

Abbildungsverzeichnis

Abbildung 1: MyCoRe-Schichtenmodell.....	20
Abbildung 2: Ausschnitt der httpd.conf.....	33
Abbildung 3: Änderungen in der server.xml.....	34
Abbildung 4: Einbindung des OAI-Data-Providers in web.xml.....	35
Abbildung 5: XML-Syntax eines Klassifikations-Objektes.....	40
Abbildung 6: XML-Syntax eines Metadaten-Objektes.....	41
Abbildung 7: XML-Syntax eines structure XML-Knotens.....	42
Abbildung 8: XML-Syntax eines metadata XML-Knotens.....	43
Abbildung 9: XML-Syntax des Metadatentyps MCRMetaAddress.....	44
Abbildung 10: Beispiel des Metadatentyps MCRMetaAddress.....	45
Abbildung 11: XML-Syntax des Metadatentyps MCRMetaBoolean.....	45
Abbildung 12: Beispiel des Metadatentyps MCRMetaBoolean.....	45
Abbildung 13: XML-Syntax des Metadaten-Basistyps MCRMetaClassification.....	46
Abbildung 14: Beispiel des Metadaten-Basistyps MCRMetaClassification.....	46
Abbildung 15: XML-Syntax des Metadaten-Basistyps MCRMetaDate.....	46
Abbildung 16: Beispiel des Metadaten-Basistyps MCRMetaDate.....	47
Abbildung 17: Syntax des Metadaten-Basistyps MCRMetaInstitutionName.....	47
Abbildung 18: Syntax des Metadaten-Basistyps MCRMetaISBN.....	48
Abbildung 19: XML-Syntax des Metadaten-Basistyps MCRMetaLangText.....	48
Abbildung 20: Beispiel des Metadaten-Basistyps MCRMetaLangText.....	48
Abbildung 21: XML-Syntax des Metadaten-Basistyps MCRMetaLink.....	49
Abbildung 22: Beispiel des Metadaten-Basistyps MCRMetaLink.....	49
Abbildung 23: Syntax des Metadaten-Basistyps MCRMetaNBN.....	50
Abbildung 24: XML-Syntax des Metadaten-Basistyps MCRMetaNumber.....	50
Abbildung 25: Beispiel des Metadaten-Basistyps MCRMetaNumber.....	50
Abbildung 26: XML-Syntax des Metadaten-Basistyps MCRMetaPersonName.....	51
Abbildung 27: Beispiel des Metadaten-Basistyps MCRMetaPersonName.....	52
Abbildung 28: XML-Syntax des Metadaten-Basistyps MCRMetaXML.....	52
Abbildung 29: Beispiel des Metadaten-Basistyps MCRMetaXML.....	53
Abbildung 30: XML-Syntax des service XML-Knotens.....	54
Abbildung 31: XML-Syntax des Derivate-Datenmodells.....	55
Abbildung 32: Das Prinzip des SimpleAccessControl-Systems.....	62
Abbildung 33: Zusammenhang von Gruppen, Benutzern und Privilegien.....	63
Abbildung 34: XML-Syntax des service XML-Knotens in Hinblick auf das SimpleAccessControl-System.....	63
Abbildung 35: Funktionsschema des SimpleWorkflow.....	69

Tabellenverzeichnis

Tabelle 1: MyCoRe Komponentenübersicht für freie Lösung auf XSLT-Basis.....	8
Tabelle 2: MyCoRe Komponentenübersicht für IBM CM 8.2.....	9
Tabelle 3: MyCoRe Komponentenübersicht für freie Lösung mit eXist.....	10
Tabelle 4: Aufbau des MyCoRe-Kerns.....	18
Tabelle 5: Feste Test-Instanzen für das MyCoRe-Beispiel.....	32
Tabelle 6: Aufbau der MCRObjectID.....	38
Tabelle 7: MyCoRe-Basisdatentypen.....	44
Tabelle 8: Liste der möglichen administrativen Privilegien.....	65
Tabelle 9: Liste der möglichen Bearbeitungs-Privilegien.....	66
Tabelle 10: Liste der möglichen Lese-Privilegien.....	67
Tabelle 11: Benutzer des MyCoReSample.....	67
Tabelle 12: Parameter des MCRStartEditorServlets.....	70
Tabelle 13: Mögliche Aktionen mit dem MCRStartEditorServlet auf dem Plattenbereich.....	71
Tabelle 14: Mögliche Aktionen mit dem MCRStartEditorServlet im Server.....	72

Vorwort

Dieses Dokument ist für einen Start in das MyCoRe-System konzipiert und soll dem Einsteiger den Weg zu einer ersten MyCoRe-Anwendung, einem Dokumenten-Server, aufzeigen. Begonnen wird mit der Beschreibung aller vorab notwendigen Installationen. Weiter geht es über die Installation des MyCoRe-Kerns und des mitgelieferten Beispiels. Ein weiteres Kapitel beschreibt die Gestaltung eines ersten Dokumenten-Servers. Abschließend sind noch einige Hinweise aus der MyCoRe-Gemeinde zu Fehlerfällen und der Umstieg auf ein anderes Search-Backend dokumentiert.

Bitte konsultieren Sie bei Detailfragen auch das mitgelieferte Programmer Guide, hier sind die Einzelheiten von MyCoRe genauer beschrieben.

ACHTUNG, diese Dokumentation basiert auf dem in Arbeit befindlichen Release 1.1 der MyCoRe-Software! Benutzen Sie bitte den aktuellen Code Stand!

1. Voraussetzungen für eine MyCoRe Anwendung

1.1 Vorbemerkungen

Das MyCoRe-Projekt ist so gestaltet, dass es dem einzelnen Anwender frei steht, welche Komponenten er für die Speicherung der Daten verwenden will. Dabei spielt natürlich das verwendete Betriebssystem eine wesentliche Rolle. Jedes System hat seine eigenen Vor- und Nachteile, die an dieser Stelle nur kurz diskutiert werden sollen. Wir wollen es dem Anwender überlassen, in welchem System er für seine Anwendung die größten Vorteile sieht. Nachfolgend finden Sie eine Tabelle der wesentlichen eingesetzten Komponenten entsprechend des gewählten Basissystems.

Allg. Basis	Lösung auf Basis von freien Komponenten (Standardinstallation)
Metadaten Store	Es wird die Metadaten-Suche via XSLT benutzt.
Volltextsuche in Datenbank	Es wird die freie Text-Suchmaschine Lucene benutzt. Es wird die freie Datenbank MySQL benutzt.
Dokument Stores	Filesystem, Lucene für Textdokumente
Web-Server	Apache / Tomcat
Servlet-Engine	Tomcat
OS	Linux oder Microsoft Windows alle weiteren UNIX-Systeme, sofern die benötigten Komponenten vorhanden sind
Vorteile	<ul style="list-style-type: none"> - Geringe Kosten durch die Nutzung freier Komponenten. - Keine direkte Bindung an einen Hersteller. - Die Komponenten sind plattformunabhängig, einfach zu installieren, gut getestet und dokumentiert. - Es kann schnell eine Anwendung fertig gestellt werden.
Nachteile	<ul style="list-style-type: none"> - Abhängigkeit von der Entwicklung in der OpenSource-Gemeinde für die freien Komponenten. - Keine Produkthaftung für Komponenten. - Derzeit werden pdf, ps, doc, OpenOffice-Formate und ascii-Formate für die Volltextsuche unterstützt. - Das System arbeitet nur mit kleinen Datenmengen derzeit vernünftig. Bei einer großen Anzahl von Datensätzen (>200) wird die Suche extrem verlangsamt. Diese Variante dient nur zu Demonstrationszwecken.

Tabelle 1: MyCoRe Komponentenübersicht für freie Lösung auf XSLT-Basis

Allg. Basis	kommerzieller IBM Content Manager
Metadaten Store	IBM Content Manager 8.2 mit zu integrierender Text-Suchmaschine IBM NSE 8.2
Volltextsuche	Diese ist für die Nutzung des IBM CM Ressource Managers voll integriert.
Datenbank	Es wird die mitgelieferte DB2 v8.2 Datenbank benutzt.
Dokument Stores	Filesystem, IBM CM Ressource Manager, IBM Video Charger
Web-Server	IBM WebSphere 5.1
Servlet-Engine	IBM WebSphere 5.1
OS	IBM AIX, Sun Solaris, Linux, Microsoft Windows
Vorteile	<ul style="list-style-type: none"> - Parametrische Metadaten-Werte können ggf. auch volltextindiziert und so gesucht werden werden. - Gute Implementierung der an XPath angelehnten Abfragesprache - Viele Dokumenttypen lassen sich automatisch volltextindizieren. - Das Basisprodukt kommt von einem Hersteller. - Viel unterstützte Dateiformate für die Volltextsuche - Die Verarbeitung großer Datenmengen ist möglich
Nachteile	<ul style="list-style-type: none"> - Sehr komplexe Installation des benötigten Content Manager /WebSphere Systems. - Abhängigkeit von den IBM Lizenz-Bedingungen.

Tabelle 2: MyCoRe Komponentenübersicht für IBM CM 8.2

Allg. Basis	Lösung auf Basis von freien Komponenten
Metadaten Store	Es wird die freie XML:DB eXist 1.02 benutzt.
Volltextsuche in Datenbank	Es wird die freie Text-Suchmaschine Lucene benutzt. Es wird die freie Datenbank MySQL benutzt.
Dokument Stores	Filesystem, Lucene für Textdokumente, Helix Server
Web-Server	Apache / Tomcat
Servlet-Engine	Tomcat
OS	Linux, Microsoft Windows alle weiteren UNIX-Systeme, sofern die benötigten Komponenten vorhanden sind
Vorteile	<ul style="list-style-type: none"> - Geringe Kosten durch die Nutzung freier Komponenten. - Keine direkte Bindung an einen Hersteller. - Die Komponenten sind plattformunabhängig, einfach zu installieren, gut getestet und dokumentiert. - Es kann schnell eine Anwendung fertig gestellt werden.

Nachteile	<ul style="list-style-type: none"> - Abhängigkeit von der Entwicklung in der OpenSource-Gemeinde für die freien Komponenten. - Keine Produkthaftung für Komponenten. - Derzeit werden nur pdf, doc, OpenOffice-Formate und ascii-Formate für die Volltextsuche unterstützt. - XPath-Abfragen unter eXist implementieren nicht alle Möglichkeiten der XQuery-Spezifikation. - eXist hat Performance-Probleme bei mehr als 4000 Dokumenten.
------------------	--

Tabelle 3: MyCoRe Komponentenübersicht für freie Lösung mit eXist

1.2 Umgebungs- und Softwarevoraussetzungen

In der nachfolgenden Beschreibung wird davon ausgegangen, dass ein Benutzer **mcradmin** für den MyCoRe-Kern und die Beispielanwendung angelegt wird. Für die Benutzung von MyCoRe muss der Benutzer mcradmin nachfolgend gelistete, frei verfügbare Software verwenden können:

- Java 2 Platform, Standard Edition, v1.4.2 (J2SE)¹ oder höher (erforderlich)
- BASH² (empfohlen unter Unix-Systemen) – Alle zusätzlichen Unix-Shell-Skripte basieren auf der bash-Shell.
- CVS-Client³ (empfohlen) – Dieser dient dem Checkout des aktuellen MyCoRe-Codes und ist für den Einbau von Bug-Fixes unerlässlich. CVS ist in den meisten Linux-Distributionen enthalten. Für andere Systeme muss es von einem separaten Server nachgeladen werden.
- ANT 1.6.x⁴ oder höher (erforderlich) – Unter Linux ist dieses Tool in den Distributionen enthalten, für die anderen Systeme muss es vom Server des Apache-Projektes geholt werden.
- OpenOffice 1.1.3⁵ oder höher (erforderlich für OpenSource Volltextindex) – Unter Linux ist dieses Paket mit in der Distribution enthalten. Wichtig ist, dass es als **mcradmin** einmal gestartet wurde, damit lokale Konfigurationen angelegt sind. Das Paket wird zur Volltextindizierung verwendet. Alternativ kann man das Produkt auch direkt beziehen.
- XPDF⁶ (erforderlich für OpenSource Volltextindex) – Das Paket ist in den gängigen Linux-Distributionen in Version 3.00x enthalten.
- GhostScript⁷ (erforderlich für OpenSource Volltextindex) - Das Paket ist in den gängigen Linux-Distributionen in Version 7.07x enthalten.
- ImageMagick⁸ (optional) – Ein mächtiges Bildkonverter-Programm, welches bei der Bearbeitung großer Bildmengen sehr hilfreich ist. Auch dieses Produkt ist in Linux-Distributionen mit enthalten.

Für das Betriebssystem AIX stellt die Firma IBM verschiedene Software-Downloads unter <http://ftp.software.ibm.com/> bereit.

1 <http://java.sun.com/j2se/>

2 <http://www.gnu.org/software/bash/bash.html>

3 <https://www.cvshome.org/dev/interface.html>

4 <http://ant.apache.org/>

5 <http://www.openoffice.org/>

6 <http://www.foolabs.com/xpdf/download.html>

7 <http://www.cs.wisc.edu/~ghost/>

8 <http://www.imagemagick.org/script/index.php>

1.3 Hinweise zu den OpenSource-Komponenten

1.3.1 Installation auf Linuxsystemen

Die folgende Beschreibung erläutert die vorbereitenden Arbeiten unter SuSE Linux. Getestet wurde unter der Version 9.2. Aus Sicherheitsgründen ist besonders zu empfehlen, auch die vom Distributor angebotenen Updates für die nachfolgenden Komponenten nachzuinstallieren.

1.3.1.1 Java

Als erster Schritt ist die *Java 2 Platform, Standard Edition, v1.4.1 (J2SE)* oder höher zu installieren. JRE und SDK befinden sich unter in der SuSE in der Distribution. Zum Test unter SuSE 9.2 wurden Java 1.4.2 verwendet. Probleme mit höheren Versionen sind bisher nicht bekannt.

Ob Java bereits auf dem System installiert ist und welche Version kann durch den Aufruf des nachfolgenden Befehls auf der Kommandozeile geprüft werden:

```
> java -version
java version "1.4.2"
Java(TM) 2 Runtime Environment, Standard Edition (build 1.4.2-b28)
Java HotSpot(TM) Client VM (build 1.4.2-b28, mixed mode)
```

Ist keine Java installiert oder nur die Runtime-Edition (jre), ist eine Installation des J2SE Software Development Kit (SDK) notwendig. Dazu sind folgenden Pakete zu installieren:

- **java2-jre-1.4.2-x**
- **java2-1.4.2-x**

1.3.1.2 MySQL

MySQL ist eine derzeit frei verfügbare relationale Datenbank, welche zur Speicherung von Daten innerhalb des MyCoRe-Projektes benötigt wird. Sie besitzt eine JDBC Schnittstelle und ist SQL konform. MySQL ist Bestandteil der meisten Linux-Distributionen. Sie könnten MySQL auch durch eine andere verfügbare Datenbank mit gleicher Funktionalität ersetzen, z. B. IBM DB2 oder Oracle. Der Test erfolgte mit einem MySQL 4.x System, höhere Versionen sollten keine Probleme bereiten.

1. Installieren Sie aus Ihrer Distribution die folgenden Pakete und danach ggf. noch vom Hersteller der Distribution per Netz angebotene Updates. Die angegebenen Versionsnummern sind nur exemplarisch.
 - **mysql-4.0.18-...**
 - **mysql-shared-4.0.18-...**
 - **mysql-client-4.0.18-...**
 - **mysqlcc-0.9.4-84**
2. Die Dokumentation steht nun unter `/usr/share/doc/packages/mysql`.
3. Starten Sie als **root** den Datenbankserver mit dem Kommando `rcmysql start` und/oder tragen Sie den Start des MySQL-Servers für den Systemstart ein.
4. Setzen Sie als **root** das mysql-root-Passwort wie folgt:
`/usr/bin/mysqladmin -u root password rootpassword`

```
/usr/bin/mysqladmin -u root -h <full_host_name> password
rootpassword
```

5. Die folgende Sequenz sorgt dafür, dass der mysql-**mcradmin**-Benutzer alle Rechte auf der Datenbank hat. Dabei werden bei der Ausführung von Kommandos von **localhost** aus keine Passwörter abgefragt. Von anderen Hosts aus muss *newpassword* eingegeben werden. (Hinweis: Dies ist die schnellste, aber keine sichere Methode. Dazu bitte die MySQL-Dokumentation kontaktieren)

```
mysql -uroot -prootpassword mysql
GRANT ALL PRIVILEGES ON *.* TO mcradmin@localhost WITH GRANT OPTION;
quit
```

6. Möchten Sie, dass auch externe Hosts auf Ihr System zugreifen, so nutzen Sie das folgende Kommando. Dabei muss von anderen Hosts aus *newpassword* eingegeben werden.

```
mysql -uroot -prootpassword mysql
GRANT ALL PRIVILEGES ON *.* TO mcradmin@'%' IDENTIFIED BY
'newpassword' WITH GRANT OPTION;
quit
```

7. Ist das Passwort einmal gesetzt, müssen Sie zusätzlich die Option -p verwenden.

8. Zum Verifizieren, ob der Server läuft, nutzen Sie folgende Kommandos

```
mysqladmin -u mcradmin version
mysqladmin -u mcradmin variables
```

9. Jetzt können Sie die Datenbasis für MyCoRe mit nachstehendem Kommando anlegen.

```
mysqladmin -u mcradmin create mycore
```

10. Falls weitere Benutzer noch das Recht auf Selects von allen Hosts aus haben sollen, verwenden Sie die Kommandos

```
mysql -u mcradmin mycore
GRANT SELECT ON mycore.* TO mcradmin@'%';
quit
```

Falls sie keine Connection auf ihren Rechnernamen (nicht localhost) aufbauen können, kann es auch mit ihrer Firewall oder TCPWrapper Einstellung zu tun haben. Bei einer Firewall sollte der Port 3306 freigegeben werden und bei einem TCPWrapper der entsprechende Dienst (**mysqld**) in die Datei */etc/hosts.allow* geschrieben werden.

1.3.1.3 Lucene

Um die frei verfügbare Text-Suchmaschine für die Volltextsuche Lucene des Apache Jakarta Projektes⁹ zu benutzen, bedarf es keiner zusätzlichen Installationen. Derzeit ist die Version Lucene 1.4.2 aktuell. Die erforderliche Datei **lucene-1.4.2.jar** wird unter dem MyCoRe-Kern im Verzeichnis *~/mycore/lib* mitgeliefert. Zur Nutzung ist lediglich die weiter hinten beschriebene Konfiguration in der MyCoRe-Anwendung nötig.

Im derzeitigen Entwicklungsstand können mit der MyCoRe Variante auf Basis freier Software nur nachfolgende Dateitypen textindiziert werden. Hierzu müssen die Tools OpenOffice und xpdf installiert worden sein. Diese sollten in den gängigen Linux-Distributionen enthalten sein.

Achtung, es können selbstverständlich nur PDF-Dateien indiziert werden, welche auch reinen Text und nicht den Text als Bild enthalten! Die Praxis hat gezeigt, dass einige PostScript Ausprägungen

⁹<http://jakarta.apache.org/lucene/docs/index.html>

nicht gelesen werden können (Fehler beim Öffnen)!

1.3.1.4 Tomcat

Wie Apache, dessen Installation im vorigen Abschnitt beschrieben wurde, ist auch Tomcat als Servlet-Engine im Linux-Umfeld ein Quasi-Standard. Für die Arbeit der Web-Anwendung des MyCoRe-Projektes wird dieses Tool zwingend benötigt. Getestet wurde unter SuSE 9.2 mit folgenden Komponenten:

- **jakarta-tomcat-5.0.19-...**
- **apache2-jakarta-tomcat-connectors-5.0.19-...**

Im File */etc/profile.local* ist folgender Eintrag vorzunehmen, damit Tomcat auch von **mcradmin** genutzt werden kann. Weiterhin werden der Memory Size für den Server erweitert.

```
export CATALINA_HOME=/opt/share/tomcat5
export CATALINA_OPTS="$CATALINA_OPTS -server -Xms256m -Xmx512m
-Xincgc"
```

MyCoRe arbeitet beim Encoding der Zeichen standardmäßig mit UTF-8. Um dies auch dem Tomcat zu vermitteln und damit alle Abfragen und Verarbeitungen, welche Umlaute verwenden, korrekt laufen, müssen Sie die Datei */usr/share/tomcat/conf/server.xml* anpassen. Ergänzen Sie das Tag für den **Connector** um das Attribut **URIEncoding="UTF-8"**.

Mit `rcatmcat [start|stop|restart]` als **root** kann der Server gestartet gestoppt bzw. neu gestartet werden und über `rcatmcat status` wird der aktuelle Staus abgefragt. Die Dokumentation von Tomcat steht unter */usr/share/doc/packages/jakarta-tomcat*.

Die Benutzung von Tomcat im Zusammenhang mit dem MyCoReSample wird auf Seite 27 Abschnitt 3.5.2.1 beschrieben.

1.3.1.5 Apache

Der Apache-Webserver ist in allen gängigen Linux-Distributionen enthalten. Die allerneuesten Apache 2 Pakete sind zu finden unter <ftp://ftp.suse.com/pub/projects/apache/apache2/9.0-i386>. Wenn sie Ihre Anwendung soweit erweitern möchten, dass über eine virtuelle Host-Adresse zugegriffen werden kann, so benötigen Sie z. B. eine Apache-Installation, wie sie hier beschrieben wird. Der Test erfolgte unter SuSE 9.2 mit folgenden Versionen:

- **apache2-2.0.49-...**
- **apache2-doc-2.0.49-...** (dieses Paket ist optional)

Mit `rcapache2 start` wird der Server per Standardeinstellungen gestartet und sollte jetzt über `http://localhost` erreichbar sein. Etwaige Fehlermeldungen werden in die Fehler-Datei */var/log/apache2/error_log* geschrieben. Die zentrale Konfigurationsdatei des Apache2 ist die unter */etc/apache2* liegende Datei *httpd.conf*. Bevor irgendwelche Änderungen vorgenommen werden bitte immer ein Sicherheitskopie anlegen. Der Befehl (als **root**) `/usr/sbin/apache2ctl configtest` überprüft die Syntax ihrer geänderten Einstellungen. Vorerst sollten jedoch die Standardeinstellungen genügen. Mit `rcapache2 [start|stop|restart]` als **root** kann der Server gestartet gestoppt bzw. neu gestartet werden und über `rcapache2 status` wird der aktuelle Staus abgefragt.

Die Einbindung von virtuellen Web-Servern und Tomcat wird in Abschnitt 4.4 auf Seite 32

besprochen.

Anmerkung: Für den Betrieb von MyCoRe ist es nicht zwingend notwendig Apache 2 zu installieren. Der Apache ist nur notwendig, wenn ein Proxy oder eine Weiterleitung auf die ServletEngine eingerichtet werden soll. Dazu genügt eine ältere Version des Apache.

1.3.2 Installation auf MS Windows-Systemen

1.3.2.1 Java

Als erster Schritt ist die *Java 2 Platform, Standard Edition, v1.4.1 (J2SE)* oder höher zu installieren.. Die aktuelle Version kann unter <http://java.sun.com/j2se/> heruntergeladen werden. Ob Java bereits auf dem System installiert ist und welche Version kann durch den Aufruf des nachfolgenden Befehls auf der Kommandozeile geprüft werden:

```
> java -version  
java version "1.4.2"  
Java(TM) 2 Runtime Environment, Standard Edition (build 1.4.2-b28)  
Java HotSpot(TM) Client VM (build 1.4.2-b28, mixed mode)
```

Fällt das Echo nicht so oder ähnlich aus, muss das *J2SE Software Development Kit (SDK)* installiert werden. Ist nur die *J2SE Java Runtime Environment (JRE)* installiert, muss ebenfalls das Java SDK nachinstalliert werden.

1.3.2.2 MySQL

MySQL kann von <http://www.mysql.de> bezogen werden.

1. mysql-essential-4.1.9-win32 oder höher laut mitgelieferter Dokumentation installieren
2. MySQL-Client starten
3. Die folgende Sequenz sorgt dafür, dass der MyCoRe-User **mcradmin** alle Rechte auf der Datenbank hat. Dabei werden bei der Ausführung von Kommandos von **localhost** aus keine Passwörter abgefragt. Von anderen Hosts aus muss *newpassword* eingegeben werden.

```
mysql -uroot -prootpassword mysql  
GRANT ALL PRIVILEGES ON *.* TO mcradmin@localhost WITH GRANT  
OPTION;  
quit
```

4. Sollen auch externe Hosts zugreifen dürfen, sind noch die folgenden Kommandos erforderlich.

```
mysql -uroot -prootpassword mysql  
GRANT ALL PRIVILEGES ON *.* TO mcradmin@'%' IDENTIFIED BY  
'newpassword' WITH GRANT OPTION;  
quit
```

5. Ist das Passwort einmal gesetzt, müssen Sie zusätzlich die Option **-p** verwenden. Zum Verifizieren, ob der Server läuft, nutzen Sie folgende Kommandos

```
mysqladmin -u mcradmin version  
mysqladmin -u mcradmin variables
```

6. Jetzt können Sie die Datenbasis für MyCoRe mit nachstehendem Kommando anlegen.

```
mysqladmin -u mcradmin create mycore
```

7. Falls weitere Benutzer noch das Recht auf Selects von allen Hosts aus haben sollen, verwenden Sie die Kommandos

```
mysql -u mcradmin mycore
GRANT SELECT ON mycore.* TO mcradmin@'%';
quit
```

1.3.2.3 Lucene

Um die frei verfügbare Text-Suchmaschine für die Volltextsuche Lucene des Apache Jakarta Projektes¹⁰ zu benutzen, bedarf es keiner zusätzlichen Installationen. Derzeit ist die Version Lucene 1.4.2 aktuell. Die erforderliche *lucene-1.4.2.jar* Datei wird unter dem MyCoRe-Kern im Verzeichnis *~/mycore/lib* mitgeliefert. Zur Nutzung ist lediglich die in Abschnitt 3.4.1.7.2 auf Seite 24 beschriebene Konfiguration in der MyCoRe-Anwendung nötig. Im derzeitigen Entwicklungsstand können mit der MyCoRe Variante auf Basis freier Software nur nachfolgende Dateitypen textindiziert werden. Hierzu müssen die Tools **OpenOffice** und **xpdf** installiert worden sein. Diese sollten in den gängigen Linux-Distributionen enthalten sein. Achtung, es können selbstverständlich nur PDF-Dateien indiziert werden, welche auch reinen Text und nicht den Text als Bild enthalten.

1.3.2.4 Apache

1. Apache2 für Windows oder höher über <http://httpd.apache.org/> holen
2. laut mitgelieferter Dokumentation installieren und testen

Die Einbindung von virtuellen Web-Servern und Tomcat wird in Abschnitt 4.4 auf Seite 32 besprochen.

Anmerkung: Für den Betrieb von MyCoRe ist es nicht zwingend notwendig Apache 2 zu installieren. Der Apache ist nur notwendig, wenn ein Proxy oder eine Weiterleitung auf die ServletEngine eingerichtet werden soll. Dazu genügt eine ältere Version des Apache.

1.3.2.5 Tomcat

Wie Apache ist auch Tomcat als Servlet-Engine ein Quasi-Standard. Für die Arbeit der Web-Anwendung des MyCoRe-Projektes ist eine Servlet-Engine Voraussetzung.

- **jakarta-tomcat-5** oder höheren
- **apache2-jakarta-tomcat-connectors-5** oder höher

Tomcat kann über <http://jakarta.apache.org/tomcat/index.html> geholt werden. Die Installation ist laut mitgelieferter Dokumentation vorzunehmen.

Die Umgebungsvariablen sind wie folgt beschrieben zu setzen, damit Tomcat auch von **mcradmin** genutzt werden kann. Weiterhin werden der Memory Size für den Server erweitert.

```
CATALINA_HOME = Pfad zum Home-Verzeichnis
CATALINA_OPTS = "$CATALINA_OPTS -server -Xms256m -Xmx512m -Xincgc"
```

MyCoRe arbeitet beim Encoding der Zeichen standardmäßig mit UTF-8. Um dies auch dem Tomcat zu vermitteln und damit alle Abfragen und Verarbeitungen, welche Umlaute verwenden, korrekt laufen, müssen Sie die Datei *\$CATALINA_HOME/conf/server.xml* anpassen. Ergänzen Sie

¹⁰ Link zu jakarta einfügen

das Tag für den **Connector** um das Attribut **URIEncoding='UTF-8'** .

1.4 IBM Content Manager 8.2 als Backend

Die Installation des Content Managers v8.2 inkl. DB2 v8.2, NSE v8.2 und WebSphere v5.1 mit Java 1.4 ist in der jeweils mitgelieferten Dokumentation hinreichend beschrieben. Daher wird an dieser Stelle nicht weiter auf die Installation dieser Komponenten eingegangen. Es ist darauf zu achten, dass

- die genannten Versionen der einzelnen Komponenten inkl. aktueller Fixes verwendet wird
- bei DB2 und Content Manager Unicodeunterstützung aktiviert wird
- der Content Manager „textsearch enabled“ ist (bzw. die entsprechenden Datenbanken icmnlbdb und rmdb)

Eine speziell auf das Backend Content Manager abgestimmte MyCoRe-Installationsanleitung mit weiteren Hinweisen befindet sich, aufgrund des Umfangs, ausgelagert im Dokumentationsverzeichnis `mycore/documentation/ContentManager` .

2.Download und Installation des MyCoRe Kerns

2.1 Download des MyCoRe Kerns

Derzeit steht der MyCoRe-Kern unter server.mycore.de. Sie sollten also unter einem Unix-System zuerst ein kleines Shell-Script *cvsmymcore.sh* erstellen, welches die erforderlichen Variablen setzt.

```
# Script zum Zugriff auf das CVS in Essen
export CVS_CLIENT_LOG=~/.cvs_log
export TERM=vt100
export CVSROOT=:pserver:anoncvs@server.mycore.de:/cvs
export CVS_RSH=ssh
cvs $*
```

Für MS Windows sollte alternativ folgendes Script als *cvsmymcore.cmd* werden:

```
rem Script zum Zugriff auf das CVS in Essen
CVS_CLIENT_LOG=cvs_log
CVSROOT=:pserver:anoncvs@server.mycore.de:/cvs
cvs $*
```

Führen Sie zuerst folgendes Kommando aus (**das Passwort ist anoncvs**)

```
touch ~/.cvspass ; cvsmymcore.sh login
oder
cvsmymcore login
```

Der MyCoRe Kern wird für alle unterstützten Systeme über das CVS Repository ausgeliefert. Das Holen der aktuellen Version erfolgt mit dem Kommando

```
cvsmymcore.sh checkout mycore
oder
cvsmymcore checkout mycore
```

Nach dem erfolgreichen CheckOut erhalten Sie unter dem Verzeichnis *~/mycore* folgende Dateistruktur.

Relativer Pfad	Inhalt
bin	Das Verzeichnis für die Shell-Skripts des Kerns.
build.xml	Die Konfigurationsdatei für den Build-Prozess.
changelog.txt	Enthält Hinweise zu Produktänderungen.
config/build.properties.template	Konfigurationsdatei für den Kern, welche die Pfade zu den verwendeten Datenspeichern enthält.
documentation	Enthält alle MyCoRe-Dokumentationen.
documentation/StartingGuide	Eine kurze Einleitung in das Projekt.
documentation/UserGuide	Das Handbuch für die Standard-Installation von MyCoRe.
documentation/QuickGuide	Einige Kurzbeschreibungen zu ganz bestimmten Spezialthemen rund um MyCoRe.
documentation/ProgGuide	Dieses Handbuch enthält tiefere Hintergrundinformationen zu MyCoRe.

Relativer Pfad	Inhalt
lib	Java-Archive von Komponenten, welche in MyCoRe genutzt werden.
license.txt	Die verbindliche Lizenz von MyCoRe.
modules	Zusätzliche modulare Programmkomponenten.
schema	XML Schema Dateien des MyCoRe-Kerns.
sources/	Verzeichnis des Java-Quellcodes. Eine Beschreibung der Pakete steht im ProgrammerGuide.
stylesheets	XSLT Stylesheets des MyCoRe-Kerns.

Tabelle 4: Aufbau des MyCoRe-Kerns

Es ist erforderlich, diese Pfadangabe in der Umgebungsvariablen MYCORE_HOME im User-Profile abzulegen.

```
export MYCORE_HOME=~/.mycore
oder
MYCORE_HOME=mycore
```

Es wird immer wieder Korrekturen am Code geben, welche bestehende Fehler beseitigen. Es ist daher immer mal sinnvoll, ein Update zu fahren. Ggf. sollten Sie vorher eine Anfrage per Mail an mycore-user@lists.sourceforge.net stellen, ob derzeit ein stabiler Code Stand vorhanden ist. Das folgend Kommando kann in abgewandelter Form auch für die anderen MyCoRe-Komponenten benutzt werden.

```
cvsmycore.sh update -dP mycore
oder
cvsmycore update -dP mycore
```

2.2 Konfiguration und Übersetzen des Kerns

1. MyCoRe verwendet das Apache Ant Build-Tool, um den Quellcode zu übersetzen und eine vollständige Beispiel-Applikation zu erzeugen. Entsprechend der Installationsanleitung des Ant-Paketes sollten Sie zunächst die Umgebungsvariable JAVA_HOME und ANT_HOME gesetzt haben. Der Aufruf `ant usage` im `mycore`-Verzeichnis zeigt bei richtiger Konfiguration, welche Aufrufe möglich sind.
2. Es ist nicht nötig, weitere Umgebungsvariablen wie etwa den Java CLASSPATH zu setzen. Das MyCoRe Ant Build-Skript ignoriert den lokal gesetzten CLASSPATH völlig und generiert stattdessen einen eigenen CLASSPATH entsprechend Ihrer Konfiguration. Somit können wir sicherstellen, dass nur die erforderlichen Pakete und Klassen in der richtigen Version verwendet werden. Die Konfiguration der Systemumgebung der verwendeten Datenbanken für die XML-Speicherung (JDOM, IBM CM8, eXist) und die Speicherung von Tabellen über JDBC in einer relationalen Datenbank (IBM DB2, MySQL, optional auch andere) wird in der Datei `config/build.properties` festgelegt.
Nach dem erstmaligen Checkout des Kerns befindet sich im `config`-Verzeichnis nur das Template `build.properties.template`. Diese Datei muss nach `build.properties` kopiert werden.
3. In der Regel werden Sie nur die beiden entsprechenden Blöcke für die verwendete XML-Datenbank (`MCR.XMLStore.*`) - **Standard ist JDOM** - und die verwendete relationale

Datenbank (MCR.JDBCStore.*) - **Standard ist MySQL** - durch kommentieren bzw. auskommentieren der vorgegebenen Zeilen und Anpassen der beiden Variablen MCR.XMLStore.BaseDir und MCR.JDBCStore.BaseDir an die lokalen Installationsverzeichnisse Ihrer Datenbanksysteme anpassen müssen. Zu Testzwecken können die vorgegebenen Standards verwendet werden. Die weiteren Variablen steuern die für den Betrieb notwendigen JAR-Dateien (MCR.*Store.Jars), eventuell zusätzlich in den CLASSPATH einzubindende class-Dateien oder Ressourcen (MCR.*Store.ClassesDirs) und zur Laufzeit erforderliche native Libraries bzw. DLLs (MCR.*Store.LibPath). Passen Sie die Werte entsprechend der Dokumentation Ihres Datenbankproduktes und der Kommentare in der Datei selbst an.

4. Sie sollten zunächst prüfen, ob ihre Systemumgebung korrekt eingerichtet ist, indem Sie

```
ant info
```

ausführen. Das Ant Build Tool zeigt Ihnen daraufhin die verwendeten JDK- und Ant-Software-Versionen und den generierten CLASSPATH und LIBPATH (für Unix Systeme) an.

5. Testen Sie mit dem Aufruf: `java org.apache.xalan.xslt.EnvironmentCheck` welche Xalan-Version ihr installiertes Java mitbringt. Sollte dieser Befehl nicht möglich sein

```
Exception in thread "main" java.lang.NoClassDefFoundError:
    org/apache/xalan/xslt/EnvironmentCheck
```

oder die installierte Version ist kleiner als 2.6.0, führen Sie bitte folgende Kommandos aus und prüfen Sie danach Ihr System erneut.

- `cd $JAVA_HOME/jre/lib`
- `mkdir endorsed`
- `cd endorsed`
- `cp $MYCORE_HOME/lib/xerces* .`
- `cp $MYCORE_HOME/lib/xalan* .`
- `cd $JAVA_HOME/lib`
- `ln -s ../jre/lib/endorsed endorsed`

6. Eine Übersicht über alle wesentlichen Build-Ziele erhalten Sie mit

```
ant usage
```

7. Übersetzen Sie alle MyCoRe Quellcode-Dateien mit dem Befehl

```
ant jar
```

Dabei entsteht, abhängig von dem von Ihnen gewählten Datenbank-System zur Speicherung der XML-Daten eine Jar-Datei `lib/mycore-for-[jdom|cm8|xmlldb].jar`.

8. Optional können Sie auch JavaDoc Quellcode-Dokumentation im HTML-Format generieren lassen, indem Sie

```
ant javadocs
```

aufrufen. Dabei entstehen HTML-Dateien im Verzeichnis `documentation/html`.

3. Die MyCore Beispielanwendung MyCoReSample

3.1 Grundlegender Aufbau und Ziel der Beispielanwendung

3.1.1 Allgemeines

Um die Funktionsweise des MyCoRe-Kernes zu testen wurde eine Beispiel-Anwendung basierend auf diesem Kern entwickelt. Sie soll dem Anwender eine voll funktionsfähige Applikation in die Hand geben, welche eine Vielzahl von Komponenten des MyCoRe-Projektes nutzt und deren Arbeitsweise klar werden lässt. Um die Anwendung, im weiteren MyCoReSample genannt, gleichzeitig sinnvoll einsetzen zu können, wurde als Beispielszenario ein Dokumenten-Server gewählt, wie er bei vielen potentiellen Nutzern zur Anwendung kommt. Auch soll das MyCoReSample die Nachfolge des erfolgreichen MILESS-Dokumenten-Servers sein und den Migrationspfad zu MyCoRe hin aufzeigen.

Die gesamte Applikation ist in mehrere Ebenen aufgeteilt. Dies gestattet eine flexible Konfiguration auf einem Zielsystem bei gleichzeitiger Nutzung mehrerer gleichartiger Anwendungen auf diesem System. Beispielsweise wollen Sie einen Dissertations-Server neben einem Server für eine Bildsammlung und einem allgemeinen Dokumenten-Server auf dem selben System laufen lassen. Alle drei basieren auf einem gemeinsamen Anwendungskern DocPortal und nutzen große Teile wie Datenmodell oder Editormasken davon gemeinsam. Diese Komponenten werden in der DocPortal-Basis untergebracht, während die Konfiguration der Tabellennamen jeweils unterschiedlich ist. Die möglichen Szenarien verdeutlicht die folgende Grafik. Für jede Webanwendung gibt es im CVS noch ein zugehöriges Tomcat-Modul, welches die vollständige Installation der OpenSource-Variante von MyCoRe beschleunigen soll.

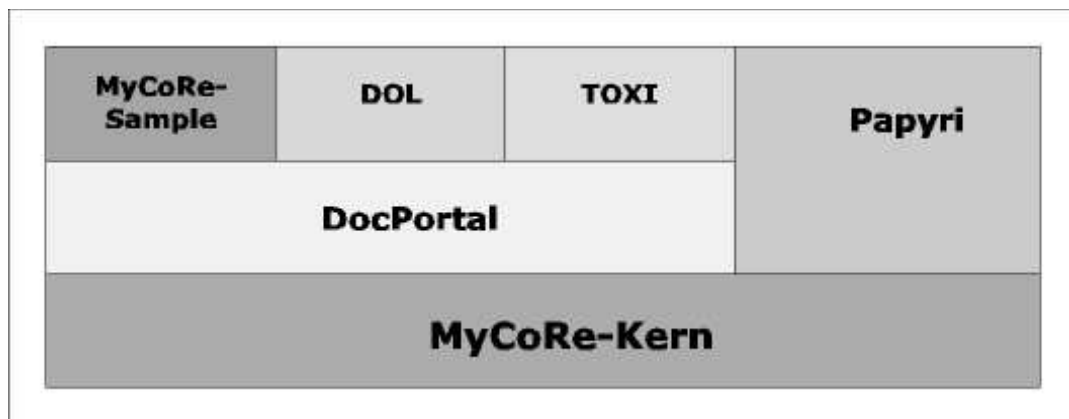


Abbildung 1: MyCoRe-Schichtenmodell

3.1.2 Weiterführende Erläuterungen

Die weiterführenden Erläuterungen zum Datenmodell und der Struktur des MyCoReSamples finden Sie in der Dokumentation zu DocPortal¹¹. In dieser Schrift ist das Konzept des Dokumenten-Servers und seiner individuellen Gestaltung als MyCoReSample beschrieben.

Auf der offiziellen Web-Seite des MyCoRe-Projektes¹² finden sie die in den nachfolgenden

¹¹ <http://www.mycore.de/content/main/documentation.xml> bzw. docportal/documentation/DokumentServer.pdf

¹² <http://mycoresamplelinux.dl.uni-leipzig.de/>

Abschnitten zu installierende Anwendung bereits lauffähig vor. Nach Abschluss der Arbeiten sollte auch Ihr System so funktionieren.

3.2 Download der Beispielanwendung

Nachdem Sie den MyCoRe-Kern erfolgreich installiert haben, ist nun die Installation der mitgelieferten Beispiel-Anwendung sinnvoll. Hier können Sie ein erstes Gefühl dafür gewinnen, wie eine eigene Anwendung gestaltet sein könnte. Das MyCoReSample wird für alle unterstützten Systeme über das CVS Repository ausgeliefert.

Das Holen der aktuellen Version von **DocPortal** und **MyCoReSample** erfolgt mit den Kommandos

```
cvsmycore.sh checkout docportal
cvsmycore.sh checkout mycoresample

oder

cvsmycore checkout docportal
cvsmycore checkout mycoresample
```

Nach erfolgreichem CheckOut finden Sie nun zwei Verzeichnisse – docportal für die DocPortal-Basis-Applikation und mycoresample als spezielle Applikation. Der folgende Abschnitt beschreibt deren Konfiguration. Für Updates dieser Komponenten verfahren Sie wie in Abschnitt 2.1 beschrieben.

3.3 Konfiguration im Anwendungsteil DocPortal

3.3.1 Allgemeines

Die Anwendungskomponente DocPortal enthält alle Dateien, welche für alle Applikationen der Klasse Dokumenten-Server gemeinsam genutzt werden sollen. Dies sind u. a. die Datenmodelldefinition, allgemeine Web-Seiten, Stylesheets zur Darstellung von Suche und Ergebnisanzeige, allgemeingültige Klassifikationen.

3.3.2 Installation

Die Installation der Zwischenschicht DocPortal gestaltet sich denkbar einfach. Führen Sie die nachfolgenden Arbeitsschritte einfach nacheinander durch.

1. Setzen Sie die Umgebungsvariable DOCPORTAL_HOME in Ihrem System, günstiger Weise in einer Profile-Datei.

```
export DOCPORTAL_HOME=~/.docportal

oder

DOCPORTAL_HOME=docportal
```

2. Prüfen Sie die Systemumgebung in DocPortal mit

```
ant info
```

3. Compilieren Sie die zusätzlichen Java-Klassen mit dem Kommando

```
ant jar
```

4. Generieren Sie die XML Schema Dateien für das Datenmodell mit dem Kommando

ant schema

3.4 Konfiguration im Anwendungsteil MyCoReSample

3.4.1 Grundlegendes zur Konfiguration

Bevor Sie beginnen, sollte als erstes die Umgebungsvariable für das MyCoReSample-Home-Verzeichnis gesetzt werden, am günstigsten im Anwender-Profil.

```
export MYCORESAMPLE_HOME=~/.mycoresample  
oder  
MYCORESAMPLE_HOME=mycoresample
```

Die Konfiguration der Beispielanwendung MyCoReSample ist in zwei Verzeichnissen zu finden. Im Verzeichnis `$DOCPORTAL_HOME/config` sind alle Dateien untergebracht, die Sie für eine erste oder einfache Installation der Anwendung nicht ändern müssen. Die Voreinstellungen entsprechen dem Standard und sollten ohne Probleme funktionieren.

Im Verzeichnis `$MYCORESAMPLE_HOME/config` hingegen sind die zu verändernden Konfigurationen hinterlegt. Kopieren Sie als erstes die Datei `mycore.properties.private.template` nach `mycore.properties.private`. Dieses File enthält für den Anfang alle einzustellenden Werte, die an Ihre spezielle Systemumgebung angepasst werden müssen. Damit dies privaten Einstellungen nicht durch ein Update überschrieben werden, wurde auf die Template-Variante zurückgegriffen.

Achten Sie bei Updates stets darauf, ob sich im template etwas geändert hat!

Die anschließenden Abschnitte des Kapitels beschäftigen sich mit der Bearbeitung der Konfigurationsdatei und behandeln jeweils einen Komplex wie JDBC, Logger, usw. Es ist aber sehr empfehlenswert, wenn Sie sich die ganze Datei sorgsam durchsehen!

Die MyCoReSample Gesamtanwendung baut sich aus den drei Komponenten MyCoRe-Kern, DocPortal und MyCoReSample auf. Beim Zusammenbau der Anwendung wird die jeweils niedrigere Komponente durch gleichnamige Dateien einer höheren Komponente überlagert. Das bedeutet, ggf. können z. B. Standardanpassungen aus DocPortal mit denen aus dem MyCoReSample überlagert werden. Ein Beispiel hierfür sind die Suchmaskendefinitionen (SearchMask...).

Die MyCoRe Beispiel-Anwendung verwendet die folgenden Dateien aus dem MyCoRe Kern.

- das File `$MYCORE_HOME/bin/build.properties`
- alle *.jar Files aus dem Verzeichnis `$MYCORE_HOME/lib`
- alle Stylesheets aus `$MYCORE_HOME/stylesheets`
- Konfigurationsdaten und zusätzliche Dateien aus den Modulen

3.4.1.1 CM8 Store konfigurieren

Dieser Abschnitt ist nur für Sie interessant, wenn Sie sich für den IBM Content Manager als Backend entschieden haben.

Die Konfiguration geht von einer Standardinstallation des IBM Content Managers aus. Sie sollten nur folgende Einträge anpassen müssen:

- `MCR.persistence_cm8_password`
- `MCR.persistence_cm8_textsearch_indexdir`

- `MCR.persistence_cm8_textsearch_workingdir`

3.4.1.2 eXist Store Konfigurieren

Dieser Teil wird in Kapitel 4.2 beschrieben. Standardmäßig benutzt das MyCoreSample den JDOM Store, welcher nicht weiter konfiguriert werden muss.

3.4.1.3 JDBC-Treiber konfigurieren

Im MyCoRe-Projekt werden ein Teil der Organisations- und Metadaten in klassischen relationalen Datenbanken gespeichert. Um die Arbeit mit verschiedenen Anbietern möglichst einfach zu gestalten, wurde die Arbeit mit dieser Datenbank gegen die JDBC-Schnittstellen programmiert. **Achtung, bei der Nutzung von MySQL unter Linux dürfen Sie nicht den System-Benutzer mit dem su-Kommando wechseln, MySQL wird dann im falschen User-Kontext ausgeführt und bringt ggf. Fehler.**

In der Konfigurationsdatei `mycore.properties.private` legen Sie im Parameter **MCR.persistence_sql_driver** fest, welcher JDBC-Treiber verwendet werden soll. Weiterhin müssen Sie die Variable **MCR.persistence_sql_database_url** anpassen, die die JDBC URL für Verbindungen zu Ihrer Datenbank festlegt. Achten Sie bei Verwendung von MySQL darauf, dass der richtige Datenbank-User (per Default **mcradmin**) angegeben ist. Sollte es zu Problemen beim Zugriff auf MySQL kommen, versuchen Sie die Adresse **127.0.0.1** durch **localhost** zu ersetzen. Das Passwort sollte auf Grund der in Kapitel 2 getätigten Konfiguration leer bleiben.

Weiterhin können Sie die minimale und maximale Anzahl der gleichzeitigen Verbindungen zur Datenbank in den beiden Parametern **MCR.persistence_sql_init_connections** und **MCR.persistence_sql_max_connections** festlegen.

3.4.1.4 Die SQL-Tabellennamen

Alle Namen der verwendeten SQL-Tabellen sind im darauf folgenden Abschnitt gelistet. Für eine erste Inbetriebnahme des MyCoReSamples müssen diese nicht geändert werden.

3.4.1.5 User-Management

Auch die Einträge für das User-Management können vorerst so bleiben. Das MyCoreSample liefert eine ganze Reihe von Privilegien, Gruppen und Benutzern mit, welche im Sample zur Darstellung der Funktionalitäten benötigt werden.

3.4.1.6 Remote-Access

MyCoRe bietet die Möglichkeit, mehrere Systeme einer Anwendungsgruppe zusammen zuschalten und die MyCoRe-Oberfläche wie ein Portal zu betreiben. Im Simplen Fall können Sie also Ihr Testsystem auch über den Remote-Zugang abfragen. Dazu müssen Sie mindestens den Hostnamen unter **MCR.remoteaccess_selfremote_host** und die Portnummer unter **MCR.remoteaccess_selfremote_port** entsprechend Ihrer Systeminstallation eintragen. Es ist sinnvoll hier den vorgesehenen Tomcat oder WebSphere-Port zu nehmen, da dies den Umweg über einen Apache-Zugriff erspart. Für die Standardinstallation haben wir den Tomcat-Port **8291** vorgesehen. Die Inbetriebnahme von Tomcat wird weiter hinten beschrieben.

3.4.1.7 Speicherung von Objekten und Dokumenten konfigurieren

Neben den Metadaten sind im MyCoReSample auch eine Reihe von Dokumenten und Bildern zum Laden in die Beispielanwendung abgelegt. Je nach Eintrag in der Datei *ContentStoreSelectionRules.xml* werden zur Speicherung der Objekte und Dokumente entsprechende Stores herangezogen.

Sollten Sie sich für den IBM Content Manager entschieden haben, so kopieren sie die Datei nach *\$MYCORESAMPLE_HOME/config* und passen Sie sie entsprechend an. Der Syntax ist selbsterklärend. Für die Nutzung von CM8 sind keine weiteren Anpassungen nötig, die Textindizierung wird bei richtiger Konfiguration des IBM Content Managers automatisch durchgeführt.

3.4.1.7.1 File System Store

Standardmäßig wird alles in ein Plattenverzeichnis abgelegt, welches mit dem Parameter **MCR.IFS.ContentStore.FS.BaseDirectory** definiert wird. MyCoReSample benutzt ein Verzeichnis *\$MYCORESAMPLE_HOME/filestore*, welches automatisch angelegt wird.

3.4.1.7.2 Lucene Store

Die auf der freien Komponente Lucene basierende Variante der Textindizierung basiert auf Tools wie OpenOffice, GhostScript und XPDF. Sie ist standardmäßig in der Konfigurationsdatei *ContentStoreSelectionRules.xml* vordefiniert. es können *.pdf, *.doc, *.txt und *.ps Objekte indiziert werden.

Lucene benötigt einen Speicherbereich für die Ablage der Indizes. Hier sollte das Verzeichnis *\$MYCORESAMPLE_HOME/lucenestore* genutzt werden, welches automatisch angelegt wird. Die Eigentlichen Objekte sollten auch nach *\$MYCORESAMPLE_HOME/filestore* gespeichert werden. Bearbeiten Sie also die Werte von *MCR.IFS.ContentStore.Lucene.IndexDirectory* und *MCR.IFS.ContentStore.Lucene.BaseDirectory* entsprechend.

3.4.1.8 Editor

Hinsichtlich der Konfiguration des Editors ist nur zu beachten, dass das im Parameter **MCR.Editor.FileUpload.TempStoragePath** angegebene Verzeichnis auch angelegt ist.

3.4.1.9 Workflow

Im Konfigurationsabschnitt Workflow sind eine ganze Reihe von Einstellungen angegeben. Für den Anfang sollten Sie nur die Werte für die Parameter **MCR.editor_..._directory**, **MCR.editor_author_..._mail** und **MCR.editor_mail_sender** anpassen.

3.4.1.10 Appletsignatur

Zum Signieren des FileUpload-Applets benötigen Sie einen Schlüssel, welchen Sie im weiteren Verlauf der Installation erzeugen müssen. Die Parameter hierfür können Sie übernehmen, nur das Verzeichnis der Keys sollten Sie mit dem Parameter **SIGN.KeyStore** anpassen. Das Verzeichnis wird automatisch erzeugt.

3.4.1.11 WCMS (Web Content Management System)

Docportal bzw. MyCoReSample hat bereits fest ein Modul zum dynamischen Verwalten der Webanwendung des Dokumentenservers enthalten.

Folgende Konfigurationsänderungen müssen gemacht werden:

1. Kopieren von `mycoresample/modules/module-wcms/aif/config/mycore.properties.wcms.template` nach `mycoresample/modules/module-wcms/aif/config/mycore.properties.wcms`
2. Dummy-Variable in `mycoresample/modules/module-wcms/aif/config/mycore.properties.wcms` `$MycoreSample_Home` durch realen Pfad zum MyCoReSample ersetzen
3. Dummy-Variable in `mycoresample/modules/module-wcms/aif/config/mycore.properties.wcms` `$AppsServerDocumentationRoot` durch realen Pfad zum webapps-Verzeichnis von Tomcat oder WebSphere ersetzen.

Die Default-Kennung für das WCMS ist

Kennung: `admin`

Passwort: `wcms`

Eine ausführlichere Beschreibung zum WCMS unter anderem für das Ändern der Passwörter befindet sich unter *mycore/modules/module-wcms/documentation/Documentation.pdf*

3.4.1.12 Debug konfigurieren

Innerhalb des MyCoRe-Projektes wird zum Erzeugen aller Print-Ausgaben für das CommandLineTool und/oder die Stdout-Logs das externe Paket **log4j** des Apache-Jakarta-Projektes benutzt¹³. Dieses ist mittlerweile ein Quasistandard und ermöglicht eine gezielte Steuerung der Informationen, welche man erhalten möchte. In der Grundkonfiguration in `mycore.properties.private` ist der Output-Level INFO eingestellt. Eine zweite Standardvorgabe ist DEBUG, diese ist auskommentiert und kann alternativ bei Problemen genommen werden. **log4j** bietet jedoch darüber hinaus noch viele weitere Möglichkeiten, die Sie bitte der Dokumentation zu diesem Produkt entnehmen. Ergänzend sei auch auf das MyCoRe-Konfigurationsfile `mycore.properties.logger` hingewiesen.

3.4.2 Initialisieren des MyCoReSamples

Nun gilt es, das MyCoreSample mit Leben zu füllen. Gehen Sie nacheinander folgende Arbeitsschritte durch. Sollte ein Fehler auftreten, beseitigen Sie diesen, bevor sie weiter arbeiten.

1. Wenn Sie es nicht schon getan haben, sollten Sie nun die SQL Datenbank anlegen.

- `mysql`
- `create database mycore;`
- `quit`

oder alternativ

¹³<http://jakarta.apache.org/log4j/docs/index2.html>

- `db2`
 - `create database mycore`
 - `quit`
2. Prüfen Sie mit dem nachfolgenden Kommando, ob alle Systemumgebungs-Variablen richtig gesetzt sind. Weiterhin wird geprüft, ob eine Verbindung zur SQL-Datenbank möglich ist.
- ```
ant info
```
3. Erzeugen Sie die CommandLineTools mit Hilfe des Kommandos
- ```
ant create.unixtools bzw. ant create.dostools
```
4. Laden Sie die Privilegien, Gruppen und Benutzer für die Beispielanwendung mit
- ```
ant create.users
```
5. Der nächste Schritt ist das Anlegen der Daten-Stores. In der freien Variante werden hier nur die SQL-Tabellen inklusive der Indexe angelegt, für den Rest kommt eine Ausschrift, dass das Ablegen weiterer Stores nicht unterstützt wird. Ignorieren Sie diese. Sollten Sie einen IBM Content Manager einsetzen, so werden auch alle benötigten ItemTypes mit angelegt.
- ```
ant create.metastore
```
6. Ein erneuter Einsatz von `ant info` gestattet Ihnen, zu prüfen, ob alle Tabellen angelegt wurden. War das Anlegen erfolgreich werden nun alle Tabellen aufgelistet.
7. Laden Sie abschließend noch alle mitgelieferten Klassifikationen. Diese werden für ein reibungsloses Funktionieren der Anwendung sowie zum Einspielen der Beispieldaten benötigt.

```
ant create.class
```

Sie sollten nun ein leeres, aber vollständiges MyCoreSample haben, in welche per CommandLineTool Daten eingestellt werden könnten. Wie dies geht, wird weiter hinten beschrieben. Auch die Inbetriebnahme der Web-Anwendung ist weiter hinten detailliert beschrieben.

3.4.3 Erzeugen der Skripte `mycore.sh` / `mycore.cmd`

Wenn Sie dieser Anleitung gefolgt sind, sollten Sie jetzt bereits alle erforderlichen Skripte für die Arbeit auf der Kommandozeile verfügbar haben. Es gib jedoch einige Fälle in denen das Basis-Skript `mycore.sh` bzw. `mycore.cmd` noch einmal erzeugt werden muss, damit alle Classpath-Einträge richtig sind. Das kann z. B. sein, wenn bei einem Update ein *.jar-File in `$MYCORE_HOME/lib` ausgetauscht wurde. Führen Sie in diesem Falle nachfolgendes Kommando aus.

```
ant scripts
```

Dieser Aufruf generiert die Shell-Skripte `bin/mycore.sh` (Unix) bzw. `bin/mycore.cmd` (Windows) neu.

3.4.4 Aufruf der MyCoRe-Kommandozeile

Starten Sie die MyCoRe-Kommandozeile, auch CommandLineInterface genannt, durch Aufruf von `bin/mycore.sh` (Unix) bzw. `bin/mycore.cmd` (Windows). Eine kurze Übersicht aller Befehle erhalten Sie durch Eingabe von **help**. Sie verlassen die MyCoRe-Kommandozeile durch Eingabe von **quit** oder **exit**. Mit `help [Kommandoteil]` erhalten Sie einen kurzen Hilfetext.

Eine ausführliche Dokumentation enthält Abschnitt 5.1.2 auf Seite 39.

Sie können natürlich auch die Aufrufe in beliebige Skripte usw. einbinden, um eine Batch-Verarbeitung zu organisieren. Das Laden der Beispieldaten erfolgte auf diese Weise.

3.5 Erzeugen und Konfigurieren der Web-Anwendung

3.5.1 Erzeugen der Web-Anwendung

Zuerst müssen Sie **einmal** Schlüssel für die Applets in der Anwendung erzeugen. Dies geschieht durch die Eingabe des Kommandos

```
ant create.genkeys
```

Nun kann die Web-Applikation gebaut werden.

```
ant webapps
```

baut die MyCoRe Sample Web Application im Verzeichnis *webapps* zusammen. Alternativ können Sie auch ein Web Application Archive (war) erzeugen, indem Sie

```
ant war
```

aufrufen.

Das MyCoRe Build-Script kopiert beim Erzeugen der Web Applikation auch alle externen, erforderlichen *.jar-Dateien Ihrer verwendeten Datenbank-Systeme (IBM Content Manager / DB2, MySQL, eXist) in das Verzeichnis *WEB-INF/lib*, entsprechend den Vorgaben Ihrer Konfiguration in *build.properties*. Beachten Sie dazu bitte die Hinweise in der Ausgabe beim Erzeugen der Web Application.

3.5.2 Konfiguration des Web Application Server

3.5.2.1 Tomcat

Die grundlegende Installation von Tomcat wurde bereits beschrieben. Nun soll auf dieser Basis das die WEB-Anwendung des MyCoReSamples installiert werden. Dabei ist an dieser Stelle nur ein einfaches Szenario auf der Basis der Tomcat-Grundinstallation beschrieben. Für die Konfiguration komplexerer Modelle, z. B. mehrere Applikationen nebeneinander, gibt es weiter hinten in diesem Dokument eine ausführliche Anleitung. Im CVS des MyCoRe-Projektes ist für das MyCoReSample ein Default abgelegt.

1. Holen Sie das Template für die Tomcat-Konfiguration aus den CVS

```
cvsmycore.sh checkout mycoresample-tomcat
```

bzw.

```
cvsmycore checkout mycoresample-tomcat
```

2. Prüfen Sie die Konfiguration im Verzeichnis *conf* und ändern Sie die eingetragenen Host-Namen nach Ihren Gegebenheiten. Vor allem werfen Sie einen Blick in die Datei *server.xml*.
3. Nun können Sie die Tomcat-Anwendung mit dem unter *bin* stehenden Script **mycoresample.sh** starten. Eventuell müssen Sie vorher die Rechte ändern und das Skript ausführbar machen.
4. Unter der URL `http://<host>:8291/` sollte nun die Anwendung laufen.

3.5.2.2 Websphere

In der gesonderten Dokumentation zur Installation von MyCoRe mit dem Backend IBM Content Manager v8.2 wird die Installation der MyCoRe-Webanwendung detailliert beschrieben. Die Dokumentation befindet sich im `mycore`-Verzeichnis unter

`documentation/ContentManager/InstallMyCoRe_CM82.pdf`

oder auf dem MyCoRe-Webseiten <http://www.mycore.de/content/main/documentation.xml>

3.6 Laden der Beispieldaten

MyCoRe stellt eine umfangreiche Sammlung von Beispieldaten bereit. Diese wurde aus Performance-Gründen in ein extra Paket gepackt. Sie können diese aus dem CVS beziehen und in Ihre fertige Anwendung laden. Die Beispieldaten sind in mehrere Blöcke gegliedert, welche einzeln geladen werden können und bestimmte Aspekte des MyCoRe-Projektes verdeutlichen sollen (Bilder, Volltexte, Audio, Video, Web-Seiten). Das Laden der Daten in das MyCoReSample geht wie folgend:

1. Download der Daten direkt neben die Verzeichnisse *mycore*, *docportal* und *mycoresample* mit

```
cvsmycore.sh checkout content
```

bzw.

```
cvsmycore checkout content
```

2. Wechseln Sie in das gewünschte Datenverzeichnis, z. B. *defaultsample*
3. Testen Sie die Umgebung mit

```
ant info
```

4. Laden Sie die Daten mit

```
ant load
```

5. Sie Können die Daten auch wieder aus Ihrem System entfernen. Dies geht mit

```
ant remove
```

6. Die Daten sind nun auch direkt in Ihrer Web-Anwendung verfügbar.

4. Vom Sample zum eigenen Dokumenten-Server

4.1 Allgemeines

Nachdem das MyCoReSample bei Ihnen nun läuft und erste Erfahrungen damit gesammelt wurden, soll nun auf dieser Grundlage aufbauend ein produktiver Dokumenten-Server aufgesetzt werden. Hier gilt es, das MyCoRe Modell in eine praxisorientierte Anwendung umzusetzen. Sicher werden bei jedem Anwender weitere zusätzliche Anforderungen auftreten, welche innerhalb des Beispiels oder der nachfolgenden Ausführungen keine Beachtung fanden. Die MyCoRe-Gruppe ist gern bereit, allgemein gültige Ergänzungen in das Kern-Projekt mit aufzunehmen.

Für die Erstellung einer Dokument- und Multimedia-Server-Anwendung sind im groben die folgenden Schritte erforderlich. Modifizieren Sie das Beispiel schrittweise hin zu Ihrer eigenen Applikation. Prüfen Sie in Zwischenschritten immer, dass die gewünschte Funktionalität noch erhalten geblieben ist.

- Kopieren des MyCoReSamples (alles unter `$MYCORESAMPLE_HOME`) in einen gesonderten Verzeichniszweig (z. B. `~/docserv`. Es ist von Vorteil diesen Pfad auch in einer Environment Variable abzulegen (z. B. `export DOCSERV_HOME=~/docserv`).
- Legen Sie für Ihre Anwendung die Namen für die Datenbank-Tabellen, XML- und IFS-Stores fest und ändern Sie die Konfigurationsdateien `$DOCSERV_HOME/config/mycore.properties.private` entsprechend.
- Kopieren Sie sich die Privileg-, Group- und User-Daten aus `$DOCPORTAL_HOME/config/user` und modifizieren Sie die User-Daten entsprechend ihren Anforderungen. Entfernen Sie die nicht benötigten Anweisungen aus `$DOCSERV_HOME/build.xml`, so dass nur noch die von Ihnen gewünschten Benutzer geladen werden. Laden Sie das User-System.
- Erstellen Sie sich alle benötigten Klassifikationen (z. B. eine aller Ihrer Einrichtungen) und laden Sie diese.
- Erzeugen Sie sich einen Satz Testdaten (pro MCRObjekt-Typ mindestens eine Datei) und laden Sie diese mit dem CommandLineTools. Nun können Sie auf CommandLine-Ebene schon mittels `mycore.sh` bzw. `mycore.cmd` Anfragen an das System stellen und erste Test durchführen.
- Legen Sie nun eine URL für ihren Server fest und setzen Sie einen Web-Server (WebSphere oder Apache/Tomcat) auf. Dieser Server sollte als eigenständige Einheit arbeiten und vom MyCoreSample entkoppelt sein. Am Besten ist, unter **mcradmin** eine Tomcat-Instanz auf einem freien Port anzulegen.
- Installieren Sie Ihre Anwendung im Web-Server und modifizieren Sie schrittweise die Präsentation nach Ihren Bedürfnissen.
- Testen und integrieren Sie die in diesem Kapitel beschriebenen weiterführenden Funktionalitäten, welche nicht im MyCoReSample enthalten sind.

Habe Sie all die Schritte bewältigt, sollte Ihnen nun ein ansprechender Dokument-Server zur Verfügung stehen. Sollten Sie andere Applikationen, wie Sammlungen usw. aufbauen wollen, konsultieren Sie bitte auch das Programmer Guide, wo auf derartige MyCoRe-Erweiterungen näher eingegangen wird.

4.2 Die Installation der freien XML:DB eXist

4.2.1 Allgemeines

eXist ist eine frei verfügbare XML:DB, welche die entsprechenden Interfaces implementiert. Für MyCoRe wurde ein auf diesen Schnittstellen basierendes Search-Backend implementiert. So ist die Nutzung von eXist direkt möglich. Für Produktionsumgebungen, welche auf Linux basieren ist es sinnvoll, das im MyCoReSample verwendete JDOM-Search-Backend gegen das von eXist auszutauschen, da die JDOM-Applikation mit ca. 200 Dokumenten ihre Grenze erreicht hat. Wie Sie das backend austauschen können, ohne einen Datenverlust zu erleiden, soll dieses Kapitel zeigen.

4.2.2 Vorbedingungen

Für das folgende Szenario ist darauf zu achten, dass der Tomcat bzw. WebSphere Server nicht auf die Ports 8080 und 8081 hören, da es ansonsten zu einem Konflikt mit der hier vorgestellten Installation kommt, weil der eXist servlet Container ebenfalls standardmäßig auf den port 8080 hört. Den Port für Tomcat ändern sie in der *tomcatinstalldir/conf/server.xml*. Wir haben uns in der Praxis entschlossen, mit eXist auf Ports mit den Nummern 8090, 8091 ... auszuweichen.

4.2.3 Installation

1. Download der aktuellen Version von eXist¹⁴. Zum Einsatz kam Version eXist-1.0b2 .
2. Entpacken Sie die Distribution in ein entsprechendes Verzeichnis, z. B. unter */home/mcradmin* (eXist-installdir).
3. Entfernen Sie zur Nutzung des Stand-Alone-Servers den Kommentar aus der Zeile `uri=xmldb:exist://localhost:8081` in der Datei */home/mcradmin/eXist-1.0b2/client.properties*
4. Kommentieren Sie folgende Zeile aus.
`uri=xmldb:exist://localhost:8080/exist/xmlrpc`
Diese Einstellung wird zum Beispiel für die direkte Einbindung in Tomcat verwendet, wenn man über xmlrpc auf eXist zugreifen möchte.
5. Ändern Sie ggf. die Portnummern.
6. Unter Linux müssen noch die Shell-Scripts ausführbar gemacht werden.
`chmod 755 <eXist-installdir>/bin/*.sh`
7. Starten Sie nun den Server mit `<eXist-installdir>/bin/server.sh` bzw. `<eXist-installdir>/bin/server.bat`
8. Anschließend können Sie auch den Client mit `<eXist-installdir>/bin/client.sh` bzw. `<eXist-installdir>/bin/client.bat` starten.
9. Hier sollten Sie dem **admin**- und **guest**-User ein Password spendieren.
10. Erzeugen Sie mit Hilfe des eXist-Client eine Collection mit dem Namen, der in der Konfiguration angegeben ist (in der Regel **mycore** oder für eine Dokumenten-Server **docserv**) und passen Sie Gruppe und Benutzer an.

¹⁴ <http://exist-db.org>

4.2.4 Konfigurationsänderungen in den MyCoRe-Teilen

Um das eXist-Backend als aktives zu integrieren, müssen Sie ein paar Änderungen an Ihrer MyCoRe-Installation vornehmen. Diese sollen in der folgenden Liste Schritt für Schritt beschrieben sein.

1. Bearbeiten Sie die Datei `$MYCORE_HOME/bin/build.properties` und kommentieren Sie ganz unten die Nutzung der JDOM-Variante aus.
2. Entfernen Sie die Kommentare in den Zeilen für eXist (Linux oder Windows). Achten Sie darauf, dass die Einträge hinsichtlich der Verzeichnisnamen richtig sind.
3. Löschen Sie die Datei `$MYCORE_HOME/lib/mycore-for-jdom.jar`
4. Löschen Sie das Verzeichnis `$MYCORE_HOME/classes`
5. Bauen Sie ein neues *.jar-File mit `ant jar`
6. Passen Sie die Datei `$DOCSERV/config/mycore.properties.private` im Abschnitt für eXist an und ergänzen Sie, wenn Sie nicht den Gast-Benutzer für die MyCoRe-Collection ausgewählt haben, folgende Zeilen


```
MCR.persistence_xmldb_user=?????? (Ihr eXist-Benutzer)
MCR.persistence_xmldb_passwd=?????? (das passende Passwort)
```
7. Erzeugen Sie ein neues `mycore.sh / mycore.cmd` mittels `ant scripts`

4.2.5 Füllen des Search-Backends für eXist

Nachdem Sie nun das System für die Arbeit mit eXist vorbereitet haben, gilt es nun noch das backend mit Daten zu füllen. MyCoRe speichert alle Daten (XML-Dateien) in einer SQL-Tabelle als BLOBs. Dies schafft eine hohe Performance bei der Auslieferung der Dateien. Zur Suche werden die Daten dann aufbereitet und noch einmal in eine Suchmaschine gepackt, in diesem Fall ist das jetzt eXist. Es ist also möglich auf einfache Art, aus den SQL-Tabellen die Daten in den entsprechenden Search-Store wieder zu laden. Dies geht wie folgend:

1. `$DOCSERV_HOME/bin/RepairExist.sh`
2. Starten Sie nun eine Test-Suche, indem Sie das Skript `DocQuery.sh` bzw `DocQuery.cmd` aktivieren.

Nun sollte Ihr System auf die Benutzung von eXist umgestellt sein. Damit die aktuelle Konfiguration übernommen wird, müssen sie ebenfalls die Webanwendung neu erstellen mit dem Aufruf `ant webapps`. Danach können Sie Tomcat starten, um auch interaktiv zu testen.

4.3 Die Zusammenarbeit mit anderen MyCoReSample-Installationen

Das MyCoRe-System ist so konzipiert, dass hinsichtlich der Metadaten gleichartige Installationen miteinander arbeiten können und von einer gemeinsamen Oberfläche (Frontend) abgefragt werden können. Hierzu müssen die Remote-Instanzen definiert werden. Auch die eigene Installation kann über diesen Weg abgefragt werden. Voraussetzung ist die im Abschnitt 'Erzeugen und Konfigurieren der Web-Anwendung' beschriebene Installation eines Web Application Servers, welcher für die Remote-Zugriffe via Servlets zuständig ist.

4.3.1 Die eigene Installation

Die Konfiguration für den Zugriff auf die eigene Installation finden Sie im File `mycore.properties.private` in dem Abschnitt `MCR.remoteaccess_selfremote`. Hier muss im Normalfall nur die Hostadresse und ggf. der Port geändert werden, alle anderen Angaben sollten übernommen werden können.

```
MCR.remoteaccess_selfremote_host = myhost.de
MCR.remoteaccess_selfremote_port = 8080
```

Nun sollten Sie Ihre Anwendung auch über die Remote-Schiene abfragen können.

4.3.2 Weitere Server benachbarter Einrichtungen

Wenn Sie MyCoRe-Installationen anderer Community-Mitglieder integrieren wollen, ist folgendes zu tun:

1. Kopieren Sie den `MCR.remoteaccess_selfremote...` Abschnitt in der Konfiguration in einen Abschnitt `MCR.remoteaccess_otherhost...` für den neuen Serverzugang.
2. Tragen Sie die korrekten Netzzugänge ein.
3. Ergänzen sie den Hostalias in der Konfigurationszeile `MCR.remoteaccess_hostaliases`

Wenn Sie nun die Suchmaskenkonfiguration unter `$DOCSERV/config/SearchMask...` noch um die Zeilen für einen weiteren entfernten Rechner (z. B. `otherhost`) erweitern, ist das System in Ihre Suche integriert.

4.3.3 Standard-Server des MyCoRe-Projektes

Von den Entwicklern des MyCoRe-Projektes werden exemplarisch einige MyCoRe-Sample-Installationen bereitgehalten. Diese sind im Konfigurationsfile `mycore.properties.remote` notiert und sollten in der Regel verfügbar sein. Sie repräsentieren eine Auswahl der verschiedenen Persistence-Layer. Auch die Auswahl für die Suche in diesen Instanzen ist bereits in das Sample integriert und solle nach dem erfolgreichen Start der Web Applikation aktiv sein.

Alias	URL	Port	Standort
mcrLpzMS	http://mycoresamplelinux.dl.uni-leipzig.de/	8291	Uni Leipzig

Tabelle 5: Feste Test-Instanzen für das MyCoRe-Beispiel

4.4 Einbindung virtueller Host-Namen mit Hilfe des Apache-Web-Server

Dieses Kapitel bezieht sich auf die SuSE 9.2 Distribution. Für andere Linux-Systeme sind ggf. kleine Änderungen erforderlich.

Standardmäßig ist der Apache2 ohne Einbindung der Proxy-Module in den Installations-CD's enthalten. Soll die Proxy-Funktionalität genutzt werden, dann ist die Neucompilierung der Quellen von Apache2 erforderlich. Der Quellcode des Apache2 liegt auf <http://httpd.apache.org> für ein Download bereit. Die aktuelle Version ist **httpd-2.0.54**.

Für die Übersetzung des Apachen2 sind noch die *apr* und *apr-util* Komponenten erforderlich. Diese

sind nicht Standardmäßig in den Installations-CD's von SuSE enthalten. Die Versionen `apr-1.1.1` und `apr-util-1.1.2` stehen unter <http://httpd.apache.org> als Tar/Zip-Files zur Verfügung. Im Quellverzeichnis von **httpd-2.0.54** sind die `apr` und `apr-util` Quellen der Version 0.9.6 enthalten.

4.4.1 Installation von httpd-2.0.54 mit Einbindung von mod_proxy

Entpacken des `httpd-2.0.54.tar.gz` in ein Arbeitsverzeichnis.

- `tar -xf httpd-2.0.54.tar`
- `cd httpd-2.0.54`
- `./configure --enable-proxy --enable-proxy-connect --enable-proxy-ftp --enable-proxy-http`
- `make`
- `make install` (installiert neuen Apache2 standardmäßig unter `/usr/local/apache2`)

Soll nun diese neue Version des Apache2 immer bei einem Neustart aktiviert wird, muss das Skript `/usr/local/apache2/bin/apachectl` nach `/etc/init.d/apache2` kopiert werden, oder es ist ein entsprechender Link zu setzen.

Zur Kontrolle der Übersetzung können Sie mittels des Kommandos

`/usr/local/apache2/bin/httpd -l` die Einbindung der Proxy-Module testen. Die Auflistung muss die beiden Module `apr` und `apr-util` anzeigen.

4.4.2 Die Verbindung von Tomcat5 und Apache2

Die Verbindung zwischen dem Apache2 und Tomcat5 wird in den Konfigurationsfiles `/usr/local/apache2/httpd.conf` und der `server.xml` von der Tomcat-Anwendung konfiguriert. Es wird ein virtueller Host in der `httpd.conf` definiert.

```
<VirtualHost mycoresample.dl.uni-leipzig.de:80>

    ProxyPass / http://mycoresample.dl.uni-leipzig.de:8291/
    ProxyPassReverse / http://mycoresample.dl.uni-leipzig.de:8291/

    ...

</VirtualHost>
```

Abbildung 2: Ausschnitt der `httpd.conf`

Die folgenden Änderungen basieren auf den im Laufe der Installation benutzten Tomcat5 Konfiguration, wie Sie in Kapitel 3 beschrieben ist.

```

<Service name="MYCORESAMPLE-Standalone">

    <!-- Define a non-SSL Coyote HTTP/1.1 Connector on port 8080 -->
    <Connector port="8291"
        maxThreads="200" minSpareThreads="25" maxSpareThreads="150"
        enableLookups="false" redirectPort="8292" acceptCount="800"
        debug="0" connectionTimeout="2000000"
        buffersize="67440000" socketBuffer="-1"
        URIEncoding="UTF-8"
        proxyName="mycoresample.dl.uni-leipzig.de" proxyPort="80"
        disableUploadTimeout="true" />

    ...
</Service>

```

Abbildung 3: Änderungen in der server.xml

Nach dem Neustart von Tomcat5 und Apache2 sollte das System nun über die virtuelle Adresse ansprechbar sein.

4.5 Nutzung der OAI Schnittstelle

4.5.1 Grundlagen

Die Open Archives Initiative¹⁵ hat 2001 ein offenes Protokoll für das Sammeln (Harvesting) von Metadaten vorgestellt. Dies geschah vor dem Hintergrund, dass gängige Suchmaschinen im WWW für die wissenschaftliche Nutzung wegen der i.d.R. unüberschaubaren Treffermenge und der fehlenden Qualität der angebotenen Treffer kaum nutzbar sind. Das **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)** liegt mittlerweile in der Version 2.0 vor. Das OAI-PMH dient zur Kommunikation zwischen **Data Providern** und **Service Providern**. Unter einem Data Provider versteht man hierbei ein Archivierungssystem, dessen Metadaten von einem (oder mehreren) Service Provider(n) abgerufen werden, der diese als Basis zur Bildung von Mehrwertdiensten benutzt (z. B. der Suche über viele Archive gleichzeitig).

Zum besseren Verständnis der weiteren Dokumentation führe ich hier die wichtigsten Definitionen kurz an:

- Ein **Harvester** ist ein Client, der OAI-PMH Anfragen stellt. Ein Harvester wird von einem Service Provider betrieben, um Metadaten aus Repositories zu sammeln.
- Ein **Repository** ist ein über das Netzwerk zugänglicher Server, der OAI-PMH Anfragen verarbeiten kann, wie sie im Open Archives Initiative Protocol for Metadata Harvesting 2.0 vom 2002-06-14 beschrieben werden¹⁶. Ein Repository wird von einem Data Provider betrieben, um Harvestern den Zugriff auf Metadaten zu ermöglichen.

Der für MyCoRe und Miles implementierte OAI Data Provider ist zertifiziert und erfüllt den OAI-PMH 2.0 Standard.

¹⁵ <http://www.openarchives.org/>

¹⁶ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

4.5.2 Der OAI Data Provider

MyCoRe bietet ein extrem flexibles Klassifikations-/Kategoriensystem. Der MyCoRe OAI Data Provider benutzt eine frei definierbare Teilmenge dieser Klassifikationen zur Strukturierung der Metadaten gemäß der Definition von **Sets** in OAI 2.0. An den Harvester werden also nur Metadaten ausgeliefert, die in diesen Klassifikationen erfasst sind, wobei die Klassifikationen eine Set-Struktur erzeugen. Zur weiteren Einschränkung kann eine zusätzliche Klassifikation (restriction classification) angegeben werden, in der die Elemente klassifiziert sein müssen, die für den OAI Data Provider aber nicht strukturbildend ist.

Sollen weitere Daten über OAI zugänglich gemacht werden, so bietet der OAI Data Provider die Möglichkeit, unter verschiedenen Namen mehrere Servlet-Instanzen zu betreiben, wobei eine Instanz jeweils ein OAI-Repository darstellt.

4.5.3 Installation

Zur Einbindung des OAI Data Providers müssen Eintragungen in den Deployment Descriptor des Servletcontainers und in die mycore.properties erfolgen.

4.5.4 Der Deployment Descriptor

Für jedes OAI-Repository muss eine Servlet-Instanz in den Deployment Descriptor nach folgendem Muster eingetragen werden:

```
<servlet id="OAIDataProvider">
  <servlet-name>
    OAIDataProvider
  </servlet-name>
  <servlet-class>
    org.mycore.services.oai.MCROAIDataProvider
  </servlet-class>
</servlet>
<servlet-mapping>
  <servlet-name>
    OAIDataProvider
  </servlet-name>
  <url-pattern>
    /servlets/OAIDataProvider
  </url-pattern>
</servlet-mapping>
```

Abbildung 4: Einbindung des OAI-Data-Providers in web.xml

4.5.5 Die mycore.properties.oai

Bei den einzurichtenden Properties ist zwischen *instanzunabhängigen* und *instanzabhängigen* Properties zu unterscheiden. Instanzunabhängige Properties sind hierbei für jedes angebotene OAI-Repository gültig, instanzabhängige Properties beziehen sich auf das jeweilige OAI-Repository.

4.5.6 Instanzunabhängige Properties

- `MCR.oai.adminemail=admin@uni-irgendwo.de` **(notwendig)** Der Administrator der OAI-Repositories.
- `MCR.oai.resumptiontoken.dir=/mycore/temp` **(notwendig)** Ein Verzeichnis, in welches der OAI Data Provider Informationen über Resumption Token ablegt.
- `MCR.oai.resumptiontoken.timeout=48` **(notwendig)** Die Zeit (in Stunden), für die die Informationen über die Resumption Token nicht gelöscht werden. Da das Löschen nur erfolgt, wenn auf ein OAI-Repository zugegriffen wird, können die Dateien evtl. auch länger aufgehoben werden.
- `MCR.oai.maxreturns=50` **(notwendig)** Die maximale Länge von Ergebnislisten, die an einen Harvester zurückgegeben werden. Überschreitet eine Ergebnisliste diesen Wert, so wird ein Resumption Token angelegt.
- `MCR.oai.queryservice=org.mycore.services.oai.MCROAIQueryService` **(notwendig)** Die Klasse, die für das Archiv das Query-Interface implementiert. Für Miless ist dies `OAIService`.
- `MCR.oai.metadata.transformer.oai_dc=MyCoReOAI-mycore2dc.xsl` **(notwendig)** Das Stylesheet, das die Transformation aus dem im Archiv benutzten Metadatenschema in das für OAI benutzte OAI Dublin Core Metadatenschema durchführt. Wenn sich das im Archiv benutzte Metadatenschema ändert, muss dieses Stylesheet angepasst werden. Optional können weitere Stylesheets angegeben werden, die einen Harvester mit anderen Metadatenformaten versorgen, hierbei muß aber auch jeweils ein Namensraum und ein Schema angegeben werden, z. B.
`MCR.oai.metadata.transformer.xmetadiss=MyCoReOAI-mycore2xmetadiss.xsl`
`MCR.oai.metadata.namespace.xmetadiss=http://www.ddb.de/standards/xMetaDiss/`
`MCR.oai.metadata.schema.xmetadiss=http://www.ddb.de/standards/xmetadiss/xmetadiss.xsd`
 Diese Stylesheets benutzen als Eingabe das Ergebnis des ersten Stylesheets.

4.5.7 Instanzabhängige Properties

Bei instanzabhängigen Properties wird der im Deployment Descriptor verwendete Servletname zur Unterscheidung für die einzelnen Repositories verwendet.

- `MCR.oai.repositoryname.OAIDataProvider=Test-Repository` **(notwendig)** Der Name des OAI-Repositories.
- `MCR.oai.repositoryidentifier.OAIDataProvider=mycore.de` **(notwendig)** Der Identifikator des OAI-Repositories (wird vom Harvester abgefragt).
- `MCR.oai.setscheme.OAIDataProvider=DocPortal_class_00000006`
`DocPortal_class_00000005` **(notwendig)** Die MyCoRe-Klassifikation, die zur Bildung der Struktur des OAI-Repositories verwendet wird.

- `MCR.oai.dini-mapping.doc-type.text=FORMAT0001` **(optional)** Mapping der Klassifikations-CategoryIDs auf international übliche Bezeichnungen in jeweils einer eigenen Zeile. Hier erscheint im OAI-Result „doc-type:text“ statt „FORMAT0001“. Beachte hierzu die DINI-Empfehlungen zu OAI:
<http://www.dini.de/documents/OAI-Empfehlungen-Okt2003-de.pdf>
- `MCR.oai.restriction.classification.OAIDataProvider=DocPortal_class_00000006` **(optional)** Die MyCoRe-Klassifikation, die zur Beschränkung der Suche verwendet wird.
- `MCR.oai.restriction.category.OAIDataProvider=dk01` **(optional)** Die MyCoRe-Kategorie, die zur Beschränkung der Suche verwendet wird.
- `MCR.oai.friends.OAIDataProvider=miami.uni-muenster.de/servlets/OAIDataProvider` **(optional)** Unter dieser Property können weitere (bekannte und zertifizierte) OAI-Repositories angegeben werden, um den Harvestern die Suche nach weiteren Datenquellen zu vereinfachen.

4.5.8 Test

Um zu testen, ob das eigene OAI-Repository funktioniert, kann man sich des Tools bedienen, das von der *Open Archives Initiative* unter <http://www.openarchives.org> zur Verfügung gestellt wird. Unter dem Menüpunkt **Tools** wird der **OAI Repository Explorer** angeboten.

4.5.9 Zertifizierung und Registrierung

Ebenfalls auf der oben angegebenen Website findet sich unter dem Menüpunkt **Community** der Eintrag **Register as a data provider**. Dort kann man anfordern, das eigene Repository zu zertifizieren und zu registrieren. Die Antwort wird an die in den Properties eingetragene Email-Adresse geschickt.

5. Weiterführende Informationen zum Aufbau von MyCoRe-Anwendungen

5.1 XML-Syntax des Datenmodells

In diesem Abschnitt wird der Syntax der einzelnen XML-Daten-Dateien und der MyCoRe-Standard-Datentypen näher beschrieben. Die Kenntnis der Syntax benötigen Sie, um eigene Datensätze für Ihren Dokumenten-Server zu erstellen. Eine umfassende Beschreibung der zugehörigen Klassen finden Sie im Programmer Guide. In den folgenden Abschnitten wird lediglich auf die XML-Syntax der Daten eingegangen.

5.1.1 Die MCRObjektID

Die Identifikation eines jeden MyCoRe Objektes erfolgt grundsätzlich über eine eindeutige ID. Die ID kann per Hand vergeben oder auch automatisch via API generiert werden. Diese hat für alle Objekte einen einheitlichen Aufbau, dessen Inhalt für jedes Projekt und jeden Datentyp festzulegen ist:

$$ID = \text{"projektkürzel_type_nummer"}$$

projektkürzel	Dieses Element ist für ein Projekt und/oder eine Einrichtung / Datengruppe festzulegen, zum Beispiel <i>UBLPapyri</i> oder <i>MyCoReDocument</i> . In MyCoRe wird es teilweise auch zur Identifikation einzelner Konfigurationsdaten mit genutzt.
type	Das Element beschreibt den Datenmodelltyp, d. h. der <i>type</i> verweist auf die zugehörige Datenmodell-Konfiguration, zum Beispiel <i>datamodel-author</i> oder <i>datamodel-document</i> . In MyCoRe wird es oft zur Identifikation einzelner Konfigurationsdaten im Zusammenhang mit der Verarbeitung dieses Datenmodells genutzt.
nummer	Ist eine frei wählbare positive Integerzahl. Diese Zahl kann in Verantwortung des Projektmanagers per Hand oder automatisch vergeben werden. Bei der Projektdefinition wird die Größe des Zahlenbereiches festgelegt. Es hat sich als sinnvoll erwiesen, nicht weniger als 8 Ziffern einzuplanen.

Tabelle 6: Aufbau der MCRObjektID

Im MyCoRe-Projekt sind zwei MCRObjektID-Typnamen reserviert und dürfen nicht für anderweitige Objekte genutzt werden. Der Typ **class** steht für Klassifikationen, der Typ **derivate** wird für Multimediaobjekte verwendet.

Es sei noch einmal ausdrücklich darauf hingewiesen, dass die MCRObjektID eine zentrale Rolle im ganzen MyCoRe-Projekt spielt. Über sie werden alle Daten identifiziert und referenziert. Es sind daher die vorgegebenen Regeln streng einzuhalten. Da es derzeit für den Datentyp zum anhängen nur eine *type*-Bezeichnung gibt, kann es beim Design eines Projektes hilfreich sein, sich für eine Gruppe von Projektkürzeln zu entscheiden, z. B. *DOLAuthor_author_...*, *DOLDocument_document_...* usw. So kann jedem Datenmodell eine dedizierte Derivate-Gruppe zugeordnet werden z. B. *DOLAuthor_derivate_...* oder *DOLDocument_derivate_...*. Diese Trennung ist nicht zwingend, hat sich aber bei der Verwaltung großer Datenmengen als günstig erwiesen. Manchmal ist es sogar sinnvoll, hierzu noch mehrere Projektkürzel für ein Datenmodell

zu verwenden, je nach Umfang des Datenbestandes und der Sicherungs- und Reparatur-Strategien des Projektes.

5.1.2 Das Klassifikationen-Datenmodell

Wie bereits erwähnt dienen Klassifikationen der einheitlichen Gliederung bestimmter Fakten. Sie sorgen dafür, dass eine einheitliche Schreibweise für bestimmte Begriffe verwendet wird. Diese Einzelbegriffe werden als Kategorien bezeichnet. Innerhalb einer Kategorie kann der Begriff in verschiedenen Sprachen aufgezeichnet sein. Die eindeutige Zuordnung zu einer Kategorie erfolgt über einen Bezeichner. Dieser besteht aus der Klassifikations- und Kategorie-ID und muss eindeutig sein.

Klassifikationen werden im MyCoRe-Sample als extra XML-Datei erstellt, in die Anwendung importiert und in Form einer Datenbank gespeichert. Dies ist für den Nutzer transparent und erfolgt mittels Schnittstellen. Der Zugriff auf die Daten erfolgt dann durch den oben genannten Bezeichner. Die Klassifikations-ID ist eine MCRObjectID mit dem Typ class. Die Kategorie-ID ist dagegen frei wählbar. Sie darf mehrstufig sein, jede Stufe spiegelt eine Hierarchieebene wieder. Die Stufen in der ID werden mit einem Punkt voneinander getrennt, 'Uni.URZ'. Das wiederum gestattet eine Abfrage nach allen untergeordneten Stufen bzw. Sub-Kategorien wie 'Uni.*'. **Achtung, sollten Sie Zahlen als Kategorie-IDs mit verwenden, so planen Sie entsprechende führende Nullen ein, andernfalls wird das Suchergebnis fehlerhaft! Weiterhin ist es sehr zu empfehlen, dieser Zahlenfolge einen Buchstaben voran zusetzen, damit die ID nicht als Zahl interpretiert wird (z. B. beim Content Manager 8.2).**

Im ID Attribut einer category ist der eindeutige Bezeichner anzugeben. Das darunter befindliche label Tag bietet die Möglichkeit, eine Kurzbezeichnung anzugeben. Mehrsprachige Ausführungen sind erlaubt. Dasselbe gilt für das Tag description. Beide Werte werden als Strings aufgefasst. Eine category kann wiederum category Tags beinhalten.

```

<?xml version="1.0" encoding="UTF-8" ?>
<mycoreclass
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="MCRClassification.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  ID="..." >
  <label xml:lang="..." text="..." description="..." />
  ...
  <categories>
    <category ID="...">
      <label xml:lang="..." text="..." description="..." />
      ...
      <category ID="...">
        <label xml:lang="..." text="..." description="..." />
        ...
      </category>
    <category ID="...">
      <label xml:lang="..." text="..." description="..." />
      ...
    </category>
  </category>
</categories>
</mycoreclass>

```

Abbildung 5: XML-Syntax eines Klassifikations-Objektes

5.1.3 Das Metadatenmodell

Die zu speichernden Daten des Beispiels teilen sich in unserem Modell in Metadaten und digitale Objekte. Dies gilt auch für die vom Anwender entwickelten Applikationen. Digitale Objekte sind Gegenstand des Abschnitts **'IFS und Content Store'**. Unter Metadaten verstehen wir in MyCoRe alle beschreibenden Daten des Objektes, die extern hinzugefügt, separat gespeichert und gesucht werden können. Dem gegenüber stehen Daten welche die digitalen Objekte selbst mitbringen. In diesem Abschnitt werden nur erstere behandelt.

Um die Metadaten besser auf unterschiedlichen Datenspeichern ablegen zu können, wurde ein System von XML-Strukturen entwickelt, das es gestattet, neben den eigentlichen Daten wie Titel, Autor usw. auch Struktur- und Service-Informationen mit abzulegen. Die eigentlichen Nutzerdaten sind wiederum typisiert, was deren speicherunabhängige Aufzeichnung erheblich vereinfacht. Es steht dem Entwickler einer Anwendung jedoch frei, hier bei Bedarf weitere hinzuzufügen. Im Folgenden soll nun der Aufbau der Metadatenobjekte im Detail beschrieben werden. Zum Verständnis des Samples sei hier auch auf den vorigen Abschnitt verwiesen. Die Metadaten werden komplett in XML erfasst und verarbeitet. Für die Grundstrukturen und Standardmetadatentypen

werden seitens MyCoRe bereits XMLSchema-Dateien mitgeliefert.

5.1.3.1 XML-Syntax eines Metadatenobjektes

```
<?xml version="1.0" encoding="UTF-8" ?>
<mycoreobject
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="....xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  ID="..."
  label="..." >
  <structure>
    ...
  </structure>
  <metadata xml:lang="de">
    ...
  </metadata>
  <service>
    ...
  </service>
</mycoreobject>
```

Abbildung 6: XML-Syntax eines Metadaten-Objektes

- Für **xsi:noNamespaceSchemaLocation** ist das entsprechende XMLSchema-File des Metadatentyps anzugeben (document.xsd)
- Die **ID** ist die eindeutige MCRObjektID.
- Der **label** ist ein kurzer Text-String, der bei administrativen Arbeiten an der Datenbasis das Identifizieren einzelner Datensätze erleichtern soll. Er kann maximal 256 Zeichen lang sein.
- Innerhalb der XML-Datenstruktur gibt es die Abschnitte **structure**, **metadata** und **service** zur Trennung von Struktur-, Beschreibungs- und Wartungsdaten. Diese Tag-Namen sind reserviert und **dürfen NICHT anderweitig verwendet werden!**

5.1.3.2 XML-Syntax des XML-Knotens structure

Im XML-Knoten **structure** sind alle Informationen über die Beziehung des Metadatenobjektes zu anderen Objekten abgelegt. Es werden derzeit die folgenden XML-Daten unter diesem Knoten abgelegt. Die Tag-Namen **parents/parent**, **children/child** und **derobjects/derobject** sind reserviert und **dürfen NICHT anderweitig verwendet werden!** Alle Sub-Knoten haben einen Aufbau wie für MCRMetaLinkID beschrieben.

In **parents** wird ein Link zu einem Elternobjekt gespeichert, sofern das referenzierende Objekt

Eltern hat. Ob dies der Fall ist, bestimmt die Anwendung. Das Tag dient der Gestaltung von Vererbungsbäumen und kann durch den Anwender festgelegt werden. Siehe auch 'Programmers Guide', Abschnitt Vererbung. Die Werte für xlink:title und xlink:label werden beim Laden der Daten automatisch ergänzt.

Die Informationen über die **children** hingegen werden durch das MyCoRe-System beim Laden der Daten **automatisch** erzeugt und **dürfen nicht per Hand geändert werden**, da sonst das Gesamtsystem nicht mehr konsistent ist. Werden die Metadaten eines Kindes oder eines Baumes von Kindern gelöscht, so wird in diesem Teil des XML-Knotens der Eintrag durch die Software entfernt.

Dasselbe gilt auch für den XML-Unterknoten derobjects. In diesem Bereich werden alle Verweise auf die an das Metadatenobjekt angehängten digitalen Objekte gespeichert. Jeder Eintrag verweist mittels einer Referenz auf ein Datenobjekt vom Typ mycorederivate, wie es im nachfolgenden Abschnitt 'IFS und Content Store' näher erläutert ist.

```
<structure>
  <parents class="MCRMetaLinkID">
    <parent xlink:type="locator" xlink:href="...mcr_id..." />
  </parents>
  <children class="MCRMetaLinkID">
    <child xlink:type="locator" xlink:href="...mcr_id..."
      xlink:label="..." xlink:title="..." />
    ...
  </children>
  <derobjects class="MCRMetaLinkID">
    <derobject xlink:type="locator" xlink:href="...mcr_id..."
      xlink:label="..." xlink:title="..." />
    ...
  </derobjects>
</structure>
```

Abbildung 7: XML-Syntax eines structure XML-Knotens

5.1.3.3 XML-Syntax des XML-Knotens metadata

Der Abschnitt **metadata** des MyCoRe-Metadatenobjektes nimmt alle Beschreibungsdaten des eigentlichen Datenmodells auf. Diese werden ihrerseits in vordefinierten Datentyp-Strukturen mit festgelegter Syntax abgelegt. Die Liste der Einzelelemente und die Reihenfolge der Typen ist dabei quasi beliebig in Anordnung und Länge. Wichtig ist nur, dass alle Datentypen bestimmte gemeinsame Eigenschaften haben. Es ist auch jederzeit möglich, weitere Typen den Projekten der Anwender hinzuzufügen (siehe dazu das Dokument MyCoRe Programmer Guide).

Die Metadaten bestehen aus einer Ansammlung von Informationen rund um das multimediale Objekt. Vorrangig wird dieser Teil in der Suche abgefragt. Jedes Metadatum (auch Metadaten-Tag)

enthält im class Attribut den Namen des MCRMeta-Typs bzw. der gleichnamigen MCRMeta-Java Klasse. Daneben gibt es noch ein Attribut heritable, in dem festgelegt wird, ob diese Metadaten vererbbar sein sollen. Weiterhin können noch die Attribute parasearch für die Einbindung in die parametrische Suche und textsearch für die Volltext-Suche über das gesamte Metadatenobjekt angegeben werden. Es sind jeweils die booleschen Werte true oder false möglich. Die mit der Vererbung verbundenen Mechanismen sind in dieser Dokumentation weiter hinten beschrieben.

Für MyCoRe wurden einige Basismetadatentypen festgelegt, mit denen die Mehrzahl der bisher in Betracht kommenden Anwendungen gelöst werden können. Die einzelnen Daten treten dabei als Liste auf, in denen mehrere Elemente des gleichen Typs erscheinen können, beispielsweise ein Titel in verschiedenen Sprachen. Jedes Listenelement hat wiederum per Default ein **type** Attribut und eine gemäß W3C spezifizierte Sprache im Attribut **xml:lang**. Die Angabe der Sprache im Tag metadata ist für alle eingeschlossenen Metadatentypen der Default-Wert. Die Liste der aktuell unterstützten Sprach-Codes entnehmen Sie bitte der Java-Quelldatei

~/mycore/sources/org/mycore/common/MCRDefaults.java

Für interne Zwecke wurde ein weiteres Attribut inherited eingeführt. Dieses ist NICHT durch den Anwender zu verändern! Es wird gesetzt, wenn das betreffende Metadatum von einem Elternteil geerbt wurde (siehe Vererbung). Diese Information ist für die Datenpräsentation sehr hilfreich.

```
<metadata xml:lang="...">
  <... class="MCRMeta..." heritable="..." parasearch="..."
    textsearch="...">
    ...
  </...>
  ...
</metadata>
```

Abbildung 8: XML-Syntax eines metadata XML-Knotens

Für das MyCoRe-Beispiel mit einem Dublin Core Datenmodell werden bereits einige Metadatentypen verwendet, welche dem MyCoRe-Kern beigelegt sind. Die Syntax der einzelnen Typen wird in den nachfolgenden Absätzen genau beschrieben.

5.1.3.4 MyCoRe Metadaten-Basistypen

In MyCoRe gibt es eine Reihe von vordefinierten XML-Datenstrukturen zur Abbildung bestimmter mehr oder minder komplexer Daten. Diese Strukturen bilden die MyCoRe-Datentypen, welche von der Dateneingabe bis hin zur Validierung und Datenpräsentation für einen einheitlichen Umgang mit den Daten sorgen. Dabei ist zwischen einfachen, recht atomaren Typen und anwendungsspezifischen komplexen Typen zu unterscheiden. Eine Auflistung finden Sie in nachfolgender Tabelle.

Einfache Typen	Komplexe Typen
MCRMetaBoolean	MCRMetaAddress
MCRMetaClassification	MCRMetaInstitutionName
MCRMetaDate	MCRMetaPersonName
MCRMetaISBN	MCRMetaIFS

Einfache Typen	Komplexe Typen
MCRMetsLangText	
MCRMetaLink	
MCRMetaLinkID	
MCRMetaNBN	
MCRMetaNumber	
MCRMetaXML	

Tabelle 7: MyCoRe-Basisdatentypen

5.1.3.4.1 XML-Syntax des Metadatentyps MCRMetaAddress

Der Basistyp MCRMetaAddress beinhaltet eine Liste von postalischen Anschriften in der Ausprägung eines XML-Abschnittes. Dabei wird berücksichtigt, dass die Anschrift in verschiedenen Sprachen und in international gängigen Formen gespeichert werden soll. Die einzelnen Subtags sind dabei selbsterklärend. Die Angaben zu **type** und **xml:lang** sind optional, ebenso die unter subtag liegenden Tags, jedoch muss mindestens eines ausgefüllt sein. Alle Werte werden als Text betrachtet. Das optionale Attribut textsearch hat keinen Effekt bei diesem Typ.

```
<tag class="MCRMetaAddress" heritable="..." parasearch="...">
  <subtag type="..." xml:lang="...">
    <country>...</country>
    <state>...</state>
    <zipcode>...</zipcode>
    <city>...</city>
    <street>...</street>
    <number>...</number>
  </subtag>
  ...
</tag>
```

Abbildung 9: XML-Syntax des Metadatentyps MCRMetaAddress

```
<addresses class="MCRMetaAddress" heritable="false" parasearch="true">
  <address type="Work" xml:lang="de">
    <country>Deutschland</country>
    <state>Sachsen</state>
    <zipcode>04109</zipcode>
    <city>Leipzig</city>
    <street>Augustuspaltz</street>
    <number>10/11</number>
  </address>
  ...
</addresses>
```

Abbildung 10: Beispiel des Metadatentyps MCRMetaAddress

5.1.3.4.2 XML-Syntax des Metadatentyps MCRMetaBoolean

Der Basistyp MCRMetaBoolean beinhaltet eine Liste von Wahrheitswerten mit zugehörigen **type** Attributen. Das optionale Attribut **textsearch** hat keinen Effekt bei diesem Typ. Folgende Werte sind zulässig:

- für **true** - 'true', 'yes', 'wahr' und 'ja'
- für **false** - 'false', 'no', 'falsch' und 'nein'

```
<tag class="MCRMetaBoolean" heritable="..." parasearch="...">
  <subtag type="..." xml:lang="...">
    ...
  </subtag>
  ...
</tag>
```

Abbildung 11: XML-Syntax des Metadatentyps MCRMetaBoolean

```
<publishes class="MCRMetaBoolean" heritable="true" parasearch="true">
  <publish type="Ausgabe_1" xml:lang="de">ja</publish>
  <publish type="Ausgabe_2" xml:lang="de">nein</publish>
  ...
</publishes>
```

Abbildung 12: Beispiel des Metadatentyps MCRMetaBoolean

5.1.3.4.3 XML-Syntax des Metadaten-Basistyps MCRMetaClassification

```
<tag class="MCRMetaClassification" heritable="..." parasearch="...">
  <subtag classid="..." categid="..." />
  ...
</tag>
```

Abbildung 13: XML-Syntax des Metadaten-Basistyps MCRMetaClassification

Der Basistyp MCRMetaClassification dient der Einbindung von Klassifikationen¹⁷ und deren Kategorien in die Metadaten. Beide Identifizierer zusammen beschreiben einen Kategorieeintrag vollständig. Dabei ist für die categid eine, ggf. mehrstufige, Kategorie-ID einzutragen. Die classid muss vom Typ MCRObjektID sein. Das optionale Attribut textsearch hat keinen Effekt bei diesem Typ. **Bitte beachten Sie die Hinweise zur Gestaltung der Kategorie-IDs im vorigen Kapitel!**

```
<origins class="MCRMetaClassification" heritable="false"
  parasearch="true">
  <origin classid="MyCoReDemoDC_class_1" categid="Unis.Leipzig.URZ" />
  ...
</origins>
```

Abbildung 14: Beispiel des Metadaten-Basistyps MCRMetaClassification

5.1.3.4.4 XML-Syntax des Metadaten-Basistyps MCRMetaDate

Der Basistyp MCRMetaDate beschreibt eine Liste von Datumsangaben welche zusätzlich mit einem type Attribut versehen werden können. Das Darstellungsformat muss der angegebenen Sprache oder der ISO 8601 Notation folgen. Innerhalb von MyCoRe werden dann alle Datumsangaben in das ISO 8601 Format umgewandelt. Das optionale Attribut textsearch hat keinen Effekt bei diesem Typ.

```
<tag class="MCRMetaDate" heritable="..." parasearch="...">
  <subtag type="..." xml:lang="...">
    ...
  </subtag>
  ...
</tag>
```

Abbildung 15: XML-Syntax des Metadaten-Basistyps MCRMetaDate

¹⁷siehe voriges Kapitel

```
<dates class="MCRMetaDate" heritable="false" parasearch="true">
  <date type="heute" xml:lang="de">15.10.2003</date>
  <date type="morgen" xml:lang="us">2003/16/10</date>
  ...
</date>
...
</dates>
```

Abbildung 16: Beispiel des Metadaten-Basistyps MCRMetaDate

5.1.3.4.5 XML-Syntax des Metadatentyps MCRMetaInstitutionName

Der Basistyp MCRMetaInstitutionName beinhaltet eine Liste der Namen einer Firma oder Einrichtung oder eines Bereiches der selben. Dabei soll berücksichtigt werden, dass die Name in verschiedenen Sprachen und in international gängigen Formen gespeichert werden sollen. Über das Attribut **type** ist eine zusätzliche Differenzierung der verschiedenen Namen möglich.

- **name** beinhaltet den vollständigen Namen (Pflicht)
- **nickname** das Pseudonym (z. B. UBL) (optional)
- **property** den rechtlichen Stand, GmbH (optional)

Das optionale Attribut textsearch bei diesem Typ bewirkt nur die Speicherung des Namens als Metadaten-Volltext.

```
<tag class="MCRMetaInstitutionName" heritable="..." parasearch="..."
textsearch="...">
  <subtag type="..." xml:lang="...">
    <fullname>...</fullname>
    <nickname>...</nickname>
    <property>...</property>
  </subtag>
  ...
</tag>
```

Abbildung 17: Syntax des Metadaten-Basistyps MCRMetaInstitutionName

5.1.3.4.6 XML-Syntax des Metadatentyps MCRMetaISBN

Diese Metadatentyp ist ganz speziell zur Speicherung einer ISBN gedacht. Er gestattet nur eine Kardinalität.

```
<tag class="MCRMetaISBN" heritable="..." parasearch="...">
  <subtag>ISBN</subtag>
</tag>
```

Abbildung 18: Syntax des Metadaten-Basistyps MCRMetaISBN

5.1.3.4.7 XML-Syntax des Metadatentyps MCRMetaLangText

Der Basistyp MCRMetaLangText dient der Speicherung einer Liste von Textabschnitten mit zugehöriger Sprachangabe. Das Attribut **textsearch** bewirkt, dass alle Text-Values in einen gemeinsamen Textindex des Metadatenobjektes abgelegt werden. Über das **form** Attribut kann noch spezifiziert werden, in welcher Form der Text geschrieben ist.

```
<tag class="MCRMetaLangText" heritable="..." parasearch="..."
textsearch="...">
  <subtag type="..." xml:lang="..." form="...">
    ...
  </subtag>
  ...
</tag>
```

Abbildung 19: XML-Syntax des Metadaten-Basistyps MCRMetaLangText

```
<titles class="MCRMetaLangText" heritable="true" parasearch="true"
textsearch="true">
  <title type="maintitle" xml:lang="de" form="plain">
    Mein Leben als MyCoRe-Entwickler
  </title>
  ...
</titles>
```

Abbildung 20: Beispiel des Metadaten-Basistyps MCRMetaLangText

5.1.3.4.8 XML-Syntax der Metadatentypen MCRMetaLink und MCRMetaLinkID

Der Basistyp MCRMetaLink wurde geschaffen, um eine Verknüpfung verschiedener MyCoRe-Objekte untereinander zu realisieren. Außerdem können hier genauso Verweise auf beliebige externe Referenzen abgelegt werden. Der Typ MCRMetaLink ist eine Implementation des W3C XLink Standards¹⁸. Auf dieser Basis enthält der MyCoRe-Metadatentyp zwei Arten von Links - eine Referenz und einen bidirektionalen Link. Bei beiden werden jedoch in MCRMetaLink nicht alle Möglichkeiten der XLink Spezifikation ausgeschöpft, da dies für die in MyCoRe benötigten Funktionalitäten nicht erforderlich ist.

¹⁸ siehe 'XLM Linking Language (XLink) Version 1.0'

Im Referenztyp ist das Attribut **xlink:type='locator'** immer anzugeben. Die eigentliche Referenz wird im **xlink:href** Attribut notiert. Dabei kann die Referenz eine URL oder eine MCRObjectID sein. Daneben können noch weitere Informationen im **xlink:label** angegeben werden, die Rolle einer verlinkten Person. Der Referenztyp kommt im MyCoRe-Sample bei der Verlinkung von Dokumenten und Personen zum Einsatz. Um den Update-Aufwand in Grenzen zu halten, wurde die genannte Verbindung als Referenz konzipiert. Somit weiß das referenzierte Objekt im Sample nichts über den Verweis.

Alternativ dazu besteht die Möglichkeit eines bidirektionalen Links. Dieser wird sowohl in der Link-Quelle wie auch im Link-Ziel eingetragen. Der Typ ist in diesem Fall **xlink:type='arc'**. Weiterhin sind die Attribute **xlink:from** und **xlink:to** erforderlich. Optional kann noch ein Titel in **xlink:title** mitgegeben werden. Das optionale Attribut **textsearch** hat keinen Effekt bei diesem Typ.

Der Basistyp MCRMetaLinkID entspricht im Aufbau dem MCRMetaLink. Der einzige Unterschied ist, dass die Attribute **xlink:href**, **xlink:from** und **xlink:to** nur mit MCRObjectIDs belegt werden dürfen.

```
<tag class="MCRMetaLink" heritable="..." parasearch="...">
  <subtag xlink:type="locator" xlink:href="..." xlink:label="..."
xlink:title="..."\>
  <subtag xlink:type="arc" xlink:from="..." xlink:to="..."
xlink:title="..."\>
  ...
</tag>
```

Abbildung 21: XML-Syntax des Metadaten-Basistyps MCRMetaLink

```
<urls class="MCRMetaLink" heritable="false" parasearch="...">
  <url xlink:type="locator" xlink:href="http://www.zoje.de"
xlink:label="ZOJE" xlink:title="Eine externe URL"\>
  <url xlink:type="arc" xlink:from="mcr_object_id_1"
xlink:to="mcr_object_id_2" xlink:title="Link zwischen Objekten"\>
  ...
</urls>
```

Abbildung 22: Beispiel des Metadaten-Basistyps MCRMetaLink

5.1.3.4.9 XML-Syntax des Metadattentyps MCRMetaNBN

Diese Metadattentyp ist ganz speziell zur Speicherung einer NBN gedacht. Er gestattet nur eine Kardinalität.

```
<tag class="MCRMetaNBN" heritable="..." parasearch="...">
  <subtag>NBN</subtag>
</tag>
```

Abbildung 23: Syntax des Metadaten-Basistyps MCRMetaNBN

5.1.3.4.10 XML-Syntax des Metadatentyps MCRMetaNumber

Der Basistyp MCRMetaNumber ermöglicht das Speichern und Suchen von Zahlenwerten. Die Zahlendarstellung kann je nach Sprache, d. h. im Deutschen mit Komma und im Englischen mit Punkt, angegeben werden. Weiterhin sind die zusätzlichen Attribute **dimension** und **measurement** möglich. Beide Attribute sind optional, ebenso wie das Default-Attribut **type**. Das optionale Attribut textsearch hat keinen Effekt bei diesem Typ.

```
<tag class="MCRMetaNumber" heritable="..." parasearch="...">
  <subtag xml:lang="..." dimension="..." mesurement="...">
    ...
  </subtag>
  ...
</tag>
```

Abbildung 24: XML-Syntax des Metadaten-Basistyps MCRMetaNumber

```
<masse class="MCRMetaNumber" heritable="false" parasearch="true">
  <mass xml:lang="de" dimension="Breite" mesurement="cm">
    12,1
  </mass>
  <mass xml:lang="en" type="neu" dimension="width" mesurement="ft">
    12.2
  </mass>
  ...
</masse>
```

Abbildung 25: Beispiel des Metadaten-Basistyps MCRMetaNumber

5.1.3.4.11 XML-Syntax des Metadatentyps MCRMetaPersonName

Der Basistyp MCRMetaPerson beinhaltet eine Liste von Namen für natürliche Personen. Dabei wird berücksichtigt, dass die Namen in verschiedenen Sprachen und international gängigen Formen auftreten können. Das Attribut **type** dient der Differenzierung der verschiedenen Namen einer

Person, Geburtsname, Synonym, Kosenamen usw. **firstname** repräsentiert den/die Vornamen, **callname** den Rufnamen, **surname** den Familiennamen, **academic** den akademischen Titel und **peerage** den Adelstitel und **prefix** Namenszusätze wie 'von', 'de' usw. **fullname** enthält nochmal den automatisch zusammengesetzten Namen.. Das optionale Attribut **textsearch** hat den Effekt, dass alle textlichen Werte des Namens in das allgemeine Feld zur Textsuche des Metadatenobjektes übernommen werden.

```
<tag class="MCRMetaPersonName" heritable="..." parasearch="...">
  <subtag type="..." xml:lang="...">
    <firstname>...</firstname>
    <callname>...</callname>
    <surname>...</surname>
    <fullname>...</fullname>
    <academic>...</academic>
    <peerage>...</peerage>
    <prefix>...</prefix>
  </subtag>
  ...
</tag>
```

Abbildung 26: XML-Syntax des Metadaten-Basistyps MCRMetaPersonName

```

<tag class="MCRMetaPersonName" heritable="true" parasearch="false">
  <subtag type="geburtsname" xml:lang="de">
    <firstname>Lisa Marie</firstname>
    <callname>Lisa</callname>
    <surname>Schnell</surname>
    <fullname>Schnelle, Lisa</fullname>
  </subtag>
  <subtag type="familiennamen" xml:lang="de">
    <firstname>Lisa Marie</firstname>
    <callname>Lisa</callname>
    <surname>Schmidt</surname>
    <fullname>Dr. phil. Freifrau von Schnelle, Lisa</fullname>
    <academic>Dr. phil.</academic>
    <peerage>Freifrau</peerage>
    <prefix>von</prefix>
  </subtag>
  ...
</tag>

```

Abbildung 27: Beispiel des Metadaten-Basistyps MCRMetaPersonName

5.1.3.4.12 XML-Syntax des Metadatentyps MCRMetaXML

Der Basistyp MCRMetaXML wurde zusätzlich als Container für einen beliebigen XML-Baum in das Projekt integriert. Dieser wird in den Textknoten des Subtags gestellt und kann dort theoretisch beliebig groß sein. Achten Sie aber darauf, dass entsprechend viel Speicherplatz in dem XML-SQL-Store vorgesehen wird. Die Tag-Attribute **parasearch** und **textsearch** haben keine Wirkung.

```

<tag class="MCRMetaXML" heritable="..." parasearch="...">
  <subtag type="..." >
    ...
  </subtag>
  ...
</tag>

```

Abbildung 28: XML-Syntax des Metadaten-Basistyps MCRMetaXML

```

<tag class="MCRMetaPerson" heritable="true" parasearch="false">
  <subtag type="tabl" >
    <table width="100"><tr><td>Col A</td><td>Col B</td></tr></table>
  </subtag>
  ...
</tag>

```

Abbildung 29: Beispiel des Metadaten-Basistyps MCRMetaXML

5.1.3.5 XML-Syntax des XML-Knotens service

Für die Einrichtung eines Workflow und um die Wartung großer Datenmengen zu vereinfachen, wurde der XML-Abschnitt service in das Metadatenobjekt integriert. Hier sind Informationen wie Datumsangaben und Flags für die Verarbeitung im Batch-Betrieb enthalten. Auch das **SimpleAccess-System** benutzt die Flags zur gezielten Speicherung von Zustandsinformationen. **Achtung, die Tag-Namen sind fest vorgegeben und dürfen nicht anderweitig verwendet werden!**

Die Datumsangaben servdates verhalten sich analog zu denen in MCRMetaDate. Folgende Möglichkeiten für das Attribut type sind vorgesehen. Weitere Typen sind jedoch integrierbar.

- **acceptdate** - Datum aus dem Dublin Core, kann frei verwendet werden.
- **createdate** - Das Erzeugungsdatum des Objektes, dieser Wert wird **automatisch** beim Anlegen des Objektes erzeugt und **bleibt immer erhalten!**
- **modifydate** - Das Datum des letzten Update, dieser Wert wird **automatisch** beim Update des Objektes erzeugt und **bleibt immer erhalten!**
- **submitdate** - Datum aus dem Dublin Core, kann frei verwendet werden.
- **validfromdate** - Datum aus dem Dublin Core, kann frei verwendet werden.
- **validtodate** - Datum aus dem Dublin Core, kann frei verwendet werden.

Im servflags Teil können kurze Texte untergebracht werden. Die Anzahl der servflag Elemente ist **theoretisch** unbegrenzt.

```
<service>
  <servdates class="MCRMetaDate">
    <servdate type="...">...</servdate>
    ...
  </servdates>
  <servflags class="MCRMetaLangText">
    <servflag>...</servflag>
    ...
  </servflag>
</service>
```

Abbildung 30: XML-Syntax des service XML-Knotens

5.1.4 Das Speichermodell für die Multimediatdaten (IFS)

Im bisherigen Verlauf dieses Kapitels wurden nur die beschreibenden Daten des multimedialen Objektes erläutert. Dieser Abschnitt beschäftigt sich damit, wie die eigentlichen Objekte dem Gesamtsystem hinzugefügt werden können. Im MyCoRe Projekt wurde zur Ablage der digitalen Objekte das Konzept des **IFS** entwickelt. Hier ist es möglich, über spezielle Konfigurationen festzulegen, in welchen Speicher (Store) die einzelnen Dateien gespeichert werden sollen.

Das Laden von Objekten erfolgt mittels einer Metadaten-Datei, welche alle Informationen über die zu speichernde(n) Datei(en) und ihre Beziehung(en) zu den Metadaten enthält. Die zu speichernden multimedialen Objekte werden im Weiteren als Derivate, also Abkömmlinge, bezeichnet, da ein Objekt in mehreren Formen, Grafikformaten, auftreten kann. Die Struktur der XML-Datei für Derivate ist fest vorgegeben, alle Felder, die nutzerseitig geändert werden können, sind unten beschrieben.

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<mycorederivate
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="....xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  ID="..."
  label="..."
>
  <derivate>
    <linkmetas class="MCRMetaLinkID">
      <linkmeta xlink:type="locator" xlink:href="..." />
    </linkmetas>
    <internals class="MCRMetaIFS">
      <internal
        sourcepath="..."
        maindoc="..."
      />
    </internals>
  </derivate>
  <service>
    ...
  </service>
</mycoreobject>

```

Abbildung 31: XML-Syntax des Derivate-Datenmodells

- Für **xsi:noNamespaceSchemaLocation** ist die entsprechende XML Schema-Datei anzugeben (Derivate.xsd)
- Die **ID** ist die eindeutige MCRObjektID.
- Der **label** ist ein kurzer Text-String, der bei administrativen Arbeiten an der Datenbasis das Identifizieren einzelner Datensätze erleichtern soll. Er kann maximal 256 Zeichen lang sein.
- Die Referenz in **linkmeta** ist die MCRObjektID des Metadatensatzes, an den das/die Objekte angehängt werden sollen.
- Das Attribut **sourcepath** enthält die Pfadangabe zu einer Datei oder zu einem Verzeichnis, welches als Quelle dienen soll. Aus diesen Dateien kann nun eine Datei ausgewählt werden, welche den Einstiegspunkt bei HTML-Seiten darstellen soll. Bei einzelnen Bildern ist hier noch einmal der Dateiname anzugeben. Ist nichts angegeben, so wird versucht Dateien wie index.html usw. zu finden.

5.2 Die Verwendung der Kommandozeilenschnittstelle

Mit Hilfe der Kommandozeilenschnittstelle können Sie administrative Aufgaben für ihre MyCoRe-Anwendung auf Kommandozeilenebene durchführen. Dies ermöglicht auch die Abarbeitung von Scripts zu Massenverwaltung der eingestellten Daten. So können Sie etwa Objekte (Dokumente, Derivate, Personen, u.s.w.) oder Klassifikationen in das Repository einlesen, aktualisieren und löschen, Suchen durchführen und Objekte in Dateien exportieren. Diese Funktionalitäten sind insbesondere bei einer Migration eines bestehenden Systems zur Sicherung der Daten sehr sinnvoll.

Die nachfolgenden Abschnitte sollen Ihnen eine Übersicht der möglichen Kommandos geben. Die einzelnen Kommandogruppen werden dabei intern in den Kommandokern importiert und stehen dem Benutzer zur Verfügung. Die mit dem MyCoRe-System ausgelieferten Kommandos können beliebig erweitert werden. Konsultieren Sie hierzu das MyCoRe Programmer Guide.

5.2.1 Basiskommandos des Kommandozeilensystems

Das Kommandozeilensystem enthält eine kleine Zahl an Basiskommandos. Diese sollen im folgenden Abschnitt beschrieben werden. Es wird ergänzt durch eine Vielzahl von weiteren Kommandos für die einzelnen Bereiche der Arbeit mit MyCoRe. Es besteht auch die Möglichkeit, eigene Anwendungskommandos als Java-Klassen zu entwickeln und zu integrieren. Das Kommandozeilen-Tool ist `mycore.sh` und steht nach dem Erzeugen mittels `ant create.scripts` im `bin`-Verzeichnis der Anwendung. In den unten stehenden Beschreibungen sind alle erforderlichen Parameter in der Form `{n}` notiert. `n` gibt dabei die Nummer des Parameters an.

Basiskommandos sind:

- `help` – zeigt alle möglichen Kommandos, auch die von der Anwendung hinzugefügten.
- `process {0}` – führt das im Parameter 0 angegebene System-Shell-Kommando aus.
- `! {0}` – führt das im Parameter 0 angegebene System-Shell-Kommando aus.
- `whoami` – zeigt den aktuellen Benutzer an.
- `login {0}` – startet einen Benutzerwechsel-Dialog für den Benutzer `{0}`.
- `change to user {0} with {1}` – Wechselt den Benutzer `{0}` mit dem Passwort `{1}`.
- `exit` – beendet das Kommandozeilensystem.
- `quit` – beendet das Kommandozeilensystem.

5.2.2 Kommandos zur Arbeit mit der Benutzerverwaltung

Eine Gruppe der verfügbaren Kommandos der Kommandozeilenschnittstelle ermöglicht die Verwaltung von Benutzern, Gruppen und Privilegien. Diese Kommandos werden im folgenden vorgestellt. Oft werden bei den Kommandos XML-Dateien mit Definitionen von Benutzern, Gruppen und Privilegien erwartet. Die Syntax dieser XML-Beschreibungen finden Sie im Programmer Guide. Es werden derzeit nicht alle Geschäftsprozesse der Benutzerverwaltung in den folgenden Kommandos abgebildet. Der Schwerpunkt liegt auf einem Management-Interface für administrativen Zugriff. Die GUI der Benutzerverwaltung (geplant für Version 1.1) wird die Möglichkeit einer vollständigen Bearbeitung aller Geschäftsprozesse bieten.

5.2.2.1 Allgemeine Verwaltungskommandos

Die folgenden Kommandos sind allgemeiner Natur.

init superuser – Dieses Kommando wird nur bei der Installation und Konfiguration des MyCoRe-Systems einmalig verwendet. Dabei werden Daten über den zu verwendenden Administrations-Account und den Gast-Account aus den Konfigurationsdateien gelesen, die grundlegenden Privilegien installiert sowie das Benutzersystem initialisiert.

check user data consistency – Dieses Kommando dient zur Kontrolle der Konsistenz des Benutzersystems. Alle Verbindungen zwischen Benutzern und Gruppen sowie zwischen Gruppen untereinander werden kontrolliert und Unregelmäßigkeiten, die eventuell durch den Import von Daten (siehe weiter unten) entstanden sind, werden ausgegeben.

set user management to read only mode

set user management to read/write mode – Mit diesen Kommandos können die Daten der Benutzerverwaltung eingefroren werden. Dies sollte vor dem Exportieren von Daten in XML-Dateien geschehen, damit sich nicht während des Exports Daten ändern oder Objekte angelegt werden.

5.2.2.2 Kommandos zum Arbeiten mit XML-Dateien

Diese Kommandos dienen dem Anlegen und Ändern von Privilegien, Gruppen und Benutzern aus XML-Dateien heraus.

create user data from file {0}

create group data from file {0} – Diese Kommandos erwarten eine XML-Datei als Parameter. In der Datei müssen ein oder mehrere Definitionen von Benutzern oder Gruppen existieren, die dann in das System integriert werden. Ein Benutzerpasswort muss im Klartext in der definierenden XML-Datei vorliegen (für die Syntax siehe den Programmer Guide). Ist die Passwortverschlüsselung eingeschaltet (siehe *mycore.properties.user*), so wird das Passwort bei der Ablage in der Datenbank automatisch verschlüsselt. Bei der Erzeugung der Objekte wird die Korrektheit der Eingaben bezüglich vorhandener Regeln überprüft. So wird z. B. getestet, ob sich eine Gruppe durch schon vorhandene Gruppen im System implizit selbst enthält oder ob IDs doppelt vergeben wurden.

import user data from file {0}

import group data from file {0} – Diese Kommandos verhalten sich ähnlich den vorhergehenden Befehlen, mit dem Unterschied, dass Daten ohne Logiküberprüfung eingelesen werden und Benutzerpasswörter bei eingeschalteter Passwortverschlüsselung verschlüsselt in den XML-Dateien vorliegen müssen. Die Import-Befehle werden üblicherweise benutzt, wenn Objekte zuvor in XML-Dateien exportiert wurden und wieder eingelesen werden sollen.

update user data from file {0}

update group data from file {0}

update privileges data from file {0} – Mit diesen Befehlen werden bereits vorhandene Benutzer, Gruppen und Privilegien aktualisiert. Dabei ist zu bedenken, dass „update“ im Sinne von „festsetzen auf neue Werte“ zu verstehen ist, die Objekte also nach dem update genau die Definitionen haben, die in den XML-Dateien festgelegt werden. Einige der Attribute können allerdings nicht verändert werden, z. B. die Erzeuger-Accounts oder das

Datum der Erzeugung. Sollen diese Daten unbedingt verändert werden, dann müssen die Objekte vorher gelöscht und neu angelegt werden.

`save all users to file {0}`

`save all groups to file {0}`

`save all privileges to file {0}`

`save user {0} to file {1}`

`save group {0} to file {1}` – Mit diesen Kommandos werden alle oder einzelne Objekte der Benutzerverwaltung in XML-Dateien gespeichert. Passwörter von Benutzern werden bei eingeschalteter Verschlüsselung verschlüsselt abgelegt. Die so entstanden Dateien können beispielsweise mit den import-Kommandos wieder geladen werden.

`encrypt passwords in user xml file {0} to file {1}` – Passwortverschlüsselung kann durch einen Konfigurationsparameter in der Datei *mycore.properties.user* aktiviert oder deaktiviert werden. Dieses Kommando wird benötigt, wenn man ein bestehendes System mit nicht eingeschalteter Verschlüsselung auf ein System mit Verschlüsselung migrieren will. Dabei verfährt man folgendermaßen: Zunächst werden alle Benutzer des alten Systems mit dem Kommando (siehe oben) `save all users to file` in eine XML-Datei exportiert. Daraufhin wendet man `encrypt passwords in user xml file {0} to file {1}` auf diese Datei an und erhält damit verschlüsselte Passwörter in den XML-Dateien. Mit dem Kommando (siehe oben) `update user data from file` können diese Daten in das System reintegriert werden. Danach muss die Kommandozeilenschnittstelle geschlossen und die Verschlüsselung in *mycore.properties.user* eingeschaltet werden.

5.2.2.3 Kommandos zum direkten Arbeiten mit Objekten der Benutzerverwaltung

`delete user {0}`

`delete group {0}` – Durch Angabe des Benutzer- oder Gruppennamens werden die Objekte mit diesen Kommandos aus dem System entfernt (und abhängige Objekte aktualisiert).

`list all users`

`list user {0}`

`list all groups`

`list group {0}`

`list all privileges` – Die Kommandos dienen dem Auflisten der Objekte der Benutzerverwaltung und sind selbsterklärend.

`set password for user {0} to {1}` – Mit Hilfe dieses Befehls kann das Passwort eines Benutzers direkt über die Kommandozeile gesetzt werden. Voraussetzung ist, dass die notwendigen Privilegien vorliegen.

`enable user {0}`

`disable user {0}` – Mit Hilfe dieser Kommandos können einzelne Benutzer temporär deaktiviert und wieder aktiviert werden. Ist ein Benutzer disabled, so kann er oder sie sich nicht mehr am System anmelden.

`add group {0} as member to group {1}`

`remove group {0} as member from group {1}`

`add user {0} as member to group {1}`

remove user {0} as member from group {1} – Mit diesen Kommandos kann direkt auf die Mitgliederlisten von Gruppen zugegriffen werden, indem Mitglieder (sowohl Gruppen als auch Benutzer) hinzugefügt oder gelöscht werden können.

add to group {0} the privilege {1}

remove from group {0} the privilege {1} – Die beiden Kommandos gestatten ein Hinzufügen oder Entfernen von Privilegien zu einer existierenden Gruppe. Da entsprechend Privileg muss dem System bekannt sein.

5.2.2.4 Das Sichern und Restaurieren der Benutzerverwaltungsdaten

Während der Initialisierung eines MyCoRe-Systems werden ein Administrationsaccount und ein Gastzugang eingerichtet zusammen mit den zugehörigen primären Gruppen (siehe Kommando *init superuser*). Dadurch ist das Sichern und Reimportieren der gesamten Daten der Benutzerverwaltung mit etwas mehr Handarbeit verbunden, weil der Administrationsaccount und Gastzugang zwar mit gesichert werden, aber vor einer Restauration der Daten z. B. nach einem Crash der SQL-Datenbank neu initialisiert werden müssen. Das bedeutet, dass sie bereits vorhanden sind und ein *import user data from file* deswegen nicht geht. Andererseits können sich die Daten dieser beiden Benutzer natürlich auch verändert haben, so dass die alten Daten wieder hergestellt werden müssen. Der folgende Ablauf führt zum Ziel. Dabei stehen <superuser> und <superuser-group> bzw. <guest> und <guest-group> für die in mycore.properties.private eingetragenen Parameter für den Administrations- und Gastzugang. In der MyCoRe-Kommandozeile werden die folgenden Befehle durchgeführt:

```
MyCoRe:> save user <superuser> to file <superuser.xml>
MyCoRe:> save user <guest> to file <guest.xml>
MyCoRe:> save group <superuser-group> to file <superuser-group.xml>
MyCoRe:> save group <guest-group> to file <guest-group.xml>
MyCoRe:> save all users to file <all-users.xml>
MyCoRe:> save all groups to file <all-groups.xml>
MyCoRe:> save all privileges to file <privileges.xml>
```

Die Benutzer <superuser> und <guest> sowie die zugehörigen Gruppen müssen aus den Dateien <all-users.xml> bzw. <all-groups.xml> manuell entfernt werden. Dann können alle Daten in einer neu erstellten SQL-Datenbank folgendermaßen importiert werden:

```
MyCoRe:> init superuser
MyCoRe:> update privileges data from file <privileges.xml>
MyCoRe:> import user data from file <all-users.xml>
MyCoRe:> import group data from file <all-groups.xml>
MyCoRe:> update user data from file <superuser.xml>
MyCoRe:> update user data from file <guest.xml>
MyCoRe:> update group data from file <superuser-group.xml>
MyCoRe:> update group data from file <guest-group.xml>
MyCoRe:> check user data consistency
```

5.2.3 Kommandos zur Arbeit mit den Daten

Die nachfolgende Auflistung von Kommandos gibt Ihnen eine Vielzahl von Werkzeugen zum Administrieren der Dateninhalte des MyCoRe-System an die Hand. Sie können mit Hilfe dieser Kommandos auch eine Batch-Verarbeitung via Scripts organisieren.

5.2.3.1 Kommandos zum Verwalten von Klassifikationen

- `load classification from file {0}` – es wird eine Klassifikation in Form einer XML-Definition gelesen und in das System geladen.
- `load all classifications from directory {0}` – es werden alle XML-Definitionen von Klassifikationen aus einem Verzeichnis gelesen und in das System geladen.
- `update classification from file {0}` – es wird eine Klassifikation in Form einer XML-Definition gelesen. Diese überschreibt die im System bereits geladene Klassifikation. Achtung, lassen Sie keine Kategorien einer bestehenden Klassifikation weg, das sonst Ihr System ggf. in einen inkonsistenten Zustand kommen kann!
- `update all classifications from directory {0}` – es werden alle XML-Definitionen von Klassifikationen aus einem Verzeichnis gelesen. Diese überschreiben die im System bereits geladene Klassifikationen. Achtung, lassen Sie keine Kategorien einer bestehenden Klassifikation weg, das sonst Ihr System ggf. in einen inkonsistenten Zustand kommen kann!
- `delete classification {0}` – es wird eine Klassifikation mit der im Parameter {0} angegebenen MCRObjektID gelöscht.
- `save classification {0} to {1}` – es wird eine Klassifikation mit der im Parameter {0} angegebenen MCRObjektID in eine Datei mit dem im Parameter {1} angegebenen Namen gespeichert.

5.2.3.2 Kommandos zur Verwaltung der Objekte

- `load object from file {0}`
- `load all objects from directory {0}`
- `load derivate from file {0}`
- `load all derivatives from directory {0}` – die Kommandos laden die Metadaten-Objekte oder die Derivate inklusive ihrer Bilder und Dokumente von einer Quelldatei oder einem Quellverzeichnis in das System.
- `update object from file {0}`
- `update all objects from directory {0}`
- `update derivate from file {0}`
- `update all derivatives from directory {0}` – die Kommandos laden die Metadaten-Objekte oder die Derivate inklusive ihrer Bilder und Dokumente von einer Quelldatei oder einem Quellverzeichnis in das System. Dabei werden die bestehenden Daten bis auf Strukturinformationen überschrieben. Strukturinformationen sind ggf. Daten über Vater-Kind- oder Objekt-Derivate-Beziehungen.
- `save object of {0} to directory {1}`
- `save object from {0} to {1} to directory {2}`
- `save derivate of {0} to directory {1}`
- `save derivate from {0} to {1} to directory {2}` – die Kommandos speichern ein oder mehrere Metadaten-Objekte oder Derivate in ein Verzeichnis ab. Für die Parameter {0}

und {1} bzw. {0} sind MCRObjectIDs anzugeben.

`delete object {0}`

`delete object from {0} to {1}`

`delete derivate {0}`

`delete derivate from {0} to {1}` – die Kommandos löschen einzelnen Metadaten-Objekte oder Derivate oder ganze Gruppen von selbigen. Alle Parameter müssen dabei MCRObjectIDs sein.

5.2.3.3 Anfragen an das System per Kommandozeilen

`query host {0} {1} {2}` – das Kommando gestattet eine Anfrage an die Datenbasis.

Parameter {0} ist dabei ein bekannter Hostalias oder eines der Schlüsselworte **local** oder **remote**. Parameter {1} enthält den MCRObjectID-type bzw. den Layout-type (document oder alldocs). Der letzte Parameter beinhaltet die eigentliche Anfrage im MyCoRe-XPath-Format (z. B. `/mycoreobject[@ID="DocPortal_document_00000001"]`).

`query local {0} {1}`

`query local {0} {1}` – in beiden Kommandos ist der Host schon gesetzt. Die Anfragen laufen gegen das lokale System oder gegen dessen remote-Schnittstelle.

5.2.3.4 Sonstige Kommandos

`create database for {0}` – erzeugt für die Benutzung des IBM Content Managers 8.2 den ItemType für den unter Parameter {0} angegebenen Datenmodelltyp.

`repair metadata search of type {0}`

`repair metadata search of ID {0}`

`repair derivate search` – die Kommandos lesen die XML-SQL-Tabellen und reparieren zerstörte Such-Indizes wie eXist Collections oder IBM Content Manager Item Types. Das Kommando kann auch beim Wechsel eines SearchStore angewendet werden.

`get last object ID for base {0}`

`get next object ID for base {0}`

`get next derivate ID for base {0}`

`show last object ID for base {0}`

`show next object ID for base {0}` – die Kommandos liefern die nächste freie oder die letzte MCRObjectID für eine vorgegeben MCRObjectID-Basis zurück. Die Basis besteht aus Project- und Type-String in der Form *project_type*.¹⁹

`check file {0}` – prüft eine im Parameter {0} angegeben Datei mittels XML-Parser.

5.3 Das SimpleAccessControl-System

Das **SimpleAccessControl**-System ist eine einfache Einrichtung, um die Zugriffe auf die MyCoRe-Datenobjekte zu organisieren. Das gesamte System basiert auf einfachen Prinzipien. Dabei werden zwei Gruppen von Aktionen gegen ein Objekt unterschieden – Präsentation der Daten und Bearbeiten der Daten. Als weitere Komponenten spielen die im Datenobjekt gespeicherten

¹⁹ Siehe Abschnitt zur MCRObjectID.

Zugriffsvorschriften, die aktuellen Benutzerdaten und ggf. die Sitzungsdaten eine wichtige Rolle.

Die nachfolgende Skizze zeigt den Zusammenhang der einzelnen Komponenten und ihr Zusammenspiel bei der Sicherung der Datenobjekte.

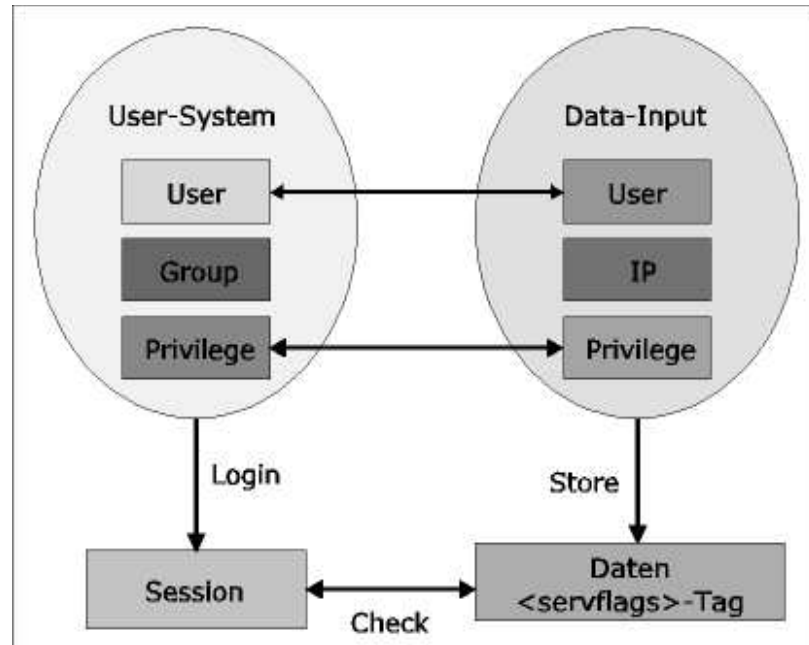


Abbildung 32: Das Prinzip des SimpleAccessControl-Systems

5.3.1 Benutzer und Privilegien

Das Benutzer-System basiert auf einer einfachen Verknüpfung von Privilegien, Gruppen und Benutzern. Gruppen können Privilegien haben – Benutzer und/oder Gruppen gehören Gruppen an. Mit diesem Prinzip können Benutzern Privilegien zugeordnet werden. Über die Verschachtelung der Gruppen werden dabei Privilegien auch vererbbar. In der Abbildung erbt die Gruppe von der untergeordneten Gruppe und vererbt ihrerseits die eigenen Privilegien an die übergeordnete Gruppe.

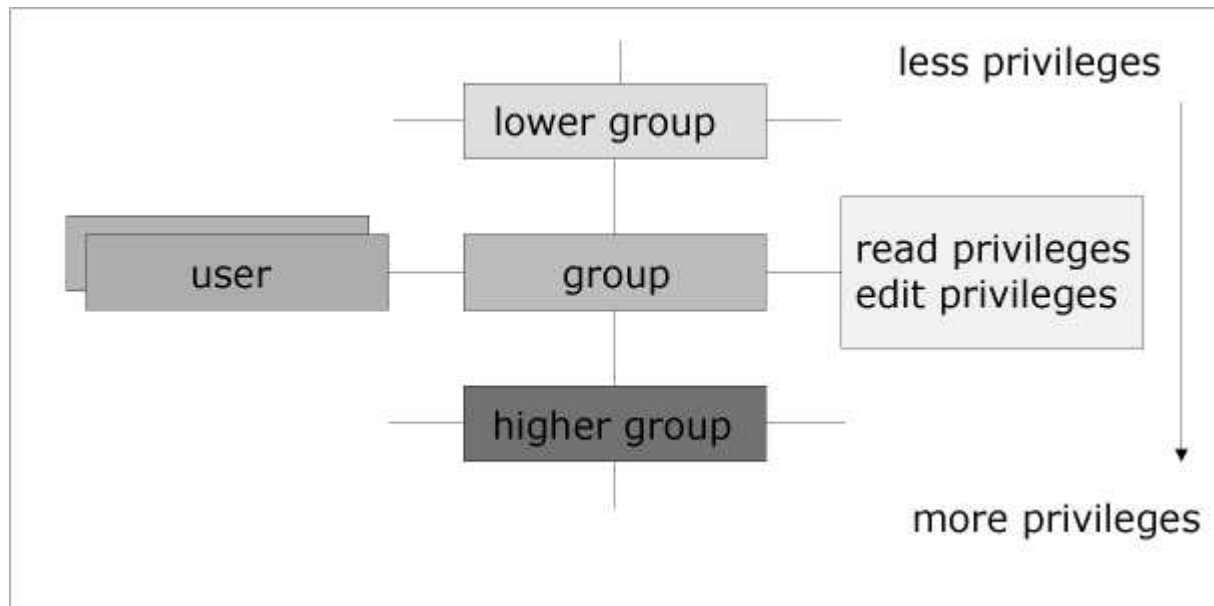


Abbildung 33: Zusammenhang von Gruppen, Benutzern und Privilegien

5.3.2 Zugriffsdaten eines Objektes

Jedes Datenobjekt hält in seinen XML-Metadaten Informationen über die Zugriffsrechte. Im Detail sind das:

1. Das Lese-Zugriffs-Privileg --> Access:...
2. Das Status-Flag --> State:...

```

<mycoreobject ...>
...
<service>
...
<servflags class="MCRMetaLangText">
  <servflag>Access:public</servflag>
  <servflag>State:complete</servflag>
  <servflag>IP:139.18.0.0/255.255.0.0</servflag>
  <servflag>User:obelix</servflag>
  ...
</servflags>
</service>
</mycoreobject>
  
```

Abbildung 34: XML-Syntax des service XML-Knotens in Hinblick auf das SimpleAccessControl-System

3. Die IP-Daten --> IP:...
4. Die Benutzer, welche an dem Objekt gearbeitet haben --> User:...

5.3.2.1 Das Lese-Zugriffs-Privileg

Mit der ersten Zeile und der Kennung '**Access:**' wird ein Privileg angegeben, welches für den lesenden Zugriff auf die Daten erforderlich ist. Per Default ist das '**public**', welches der Benutzer **gast** hat. Da sich alle Benutzergruppen von der Gruppe **gastgroup** ableiten sollten, hat dieses Privileg jeder Benutzer. Die Lese-Privilegien sollten die Form hidden-... haben.

5.3.2.2 Das Status-Privileg

Das Status-Privileg wird derzeit noch nicht mit benutzt. Es ist jedoch vorgesehen, um über den Status des Datenobjektes zu informieren und diese Information ggf. in einem XSL-Stylesheet zu verwenden.

5.3.2.3 Das IP-Flag

Über das IP-Flag kann der Zugriff von Web-Anwendungen auf die Daten beschränkt werden. Die Anfragen werden nur gestattet, wenn diese aus einem Netzsegment kommt, welches zugelassen ist. Dazu wird die aktuelle IP-Adresse mit der Subnetzmaske **OR** verknüpft und gegen die Segmentadresse verglichen.

5.3.2.4 Die Benutzer

Bei jeder Manipulation der Daten, wird aufgezeichnet, welche Benutzer dies getan haben. Ist der Benutzer noch nicht registriert, so wird ein weiterer <servflag> -Eintrag mit der Kennung '**User:**' hinzugefügt.

5.3.3 Testmechanismen

Für die Prüfung des berechtigten Zugriffes stehen API-Funktionen in einer Java-Klasse und ein Servlet zur Arbeit mit XSLT zur Verfügung. Die Testmechanismen sind im aktuellen MyCoReSample-Code integriert.

5.3.4 Die konkrete Umsetzung im MyCoReSample

5.3.4.1 Administrative Privilegien

Diese Privilegien dienen der Organisation des Benutzersystems und sind nur für die Administratoren einer Anwendung gedacht. Im MyCoReSample sind all diese Privilegien der Gruppe **admingroup** zugeordnet, deren Benutzer der **administrator** ist.

Privileg	Gruppe	Beschreibung
user administrator	• rootgroup	Superuserrechte
create user	• admingroup	Anlegen von Benutzern
delete user	• admingroup	Löschen von Benutzern
modify user	• admingroup	Verändern von Benutzern
create group	• admingroup	Anlegen von Gruppen
delete group	• admingroup	Löschen von Gruppen
modify group	• admingroup	Verändern von Gruppen

Privileg	Gruppe	Beschreibung
modify privileges	<ul style="list-style-type: none">• rootgroup	Verändern von Privilegien
list all users	<ul style="list-style-type: none">• alle Gruppen	Anzeigen den Benutzer
list all privileges	<ul style="list-style-type: none">• admingroup	Anzeigen der Privilegien
create classifications	<ul style="list-style-type: none">• admingroup	Erstellen einer Klassifikation
delete classifications	<ul style="list-style-type: none">• admingroup	Löschen einer Klassifikation
modify classifications	<ul style="list-style-type: none">• admingroup	Verändern einer Klassifikation

Tabelle 8: Liste der möglichen administrativen Privilegien

5.3.4.2 Privilegien zur Bearbeitung der Daten

Die folgenden Privilegien sichern die Arbeit mit dem SimpleWorkflow ab. Sie müssen den entsprechenden Gruppen zugewiesen werden, welche bestimmte Aktionen ausführen sollen.

Privileg	Gruppe	Beschreibung
create-institution	<ul style="list-style-type: none"> • authorgroup1 • authorgroup2 	Erstellen der eingegebenen Institutionsdaten
delete-institution	<ul style="list-style-type: none"> • ditorgroup1 • ditorgroup2 	Löschen der eingegebenen Institutionsdaten aus dem Workflow
modif- institution	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Verändern der eingegebenen Institutionsdaten
commit-institution	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Einstellen in den Server der eingegebenen Institutionsdaten
remove-institution	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Löschen der eingegebenen Institutionsdaten aus dem System
create-author	<ul style="list-style-type: none"> • uthorgroup1 • authorgroup2 	Erstellen der eingegebenen Personendaten
delete-author	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Löschen der eingegebenen Personendaten aus dem Workflow
modify-author	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Verändern der eingegebenen Personendaten
commit-author	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Einstellen in den Server der eingegebenen Personendaten
remove-author	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Löschen der eingegebenen Personendaten aus dem System
create-document	<ul style="list-style-type: none"> • authorgroup1 • authorgroup2 	Erstellen von Dokumenten
delete-document	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Löschen der eingegebenen Dokumenten aus dem Workflow
modify-document	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Verändern der eingegebenen Dokumenten
commit-document	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Einstellen in den Server der eingegebenen Dokumenten
remove-document	<ul style="list-style-type: none"> • authorgroup1 • authorgroup2 	Löschen der eingegebenen Dokumenten aus dem System
editor	<ul style="list-style-type: none"> • editorgroup1 • editorgroup2 	Berechtigt zur Bearbeitung aller Datenobjekte, deren Autor in einer Gruppe ist, welche die Editorgruppe erweitert. (editorgroup1 --> authorgroup1)

Tabelle 9: Liste der möglichen Bearbeitungs-Privilegien

5.3.4.3 Lese-Privilegien

Diese Privilegien dienen nur der Selektion der Leseberechtigung. Per Default haben alle Objekte das Privileg **public**. Soll die Nutzung des Objektes auf eine bestimmte Gruppe von Nutzern eingeschränkt werden, so können beliebig viele Leseprivilegien **hidden-...** vergeben werden. Die

Gruppe der Nutzer die dann noch lesen dürfen, MUSS das entsprechende Privileg haben.

Privileg	Gruppe	Beschreibung
hidden-group1	• readergroup1	Lesen der noch nicht öffentlichen Dokumenten Gruppe 1
hidden-group2	• readergroup2	Lesen der noch nicht öffentlichen Dokumenten Gruppe 2.
public	• gastgroup	Lesen aller öffentlich gemachten Dokumente

Tabelle 10: Liste der möglichen Lese-Privilegien

5.3.4.4 Benutzer und Gruppen des Beispiels

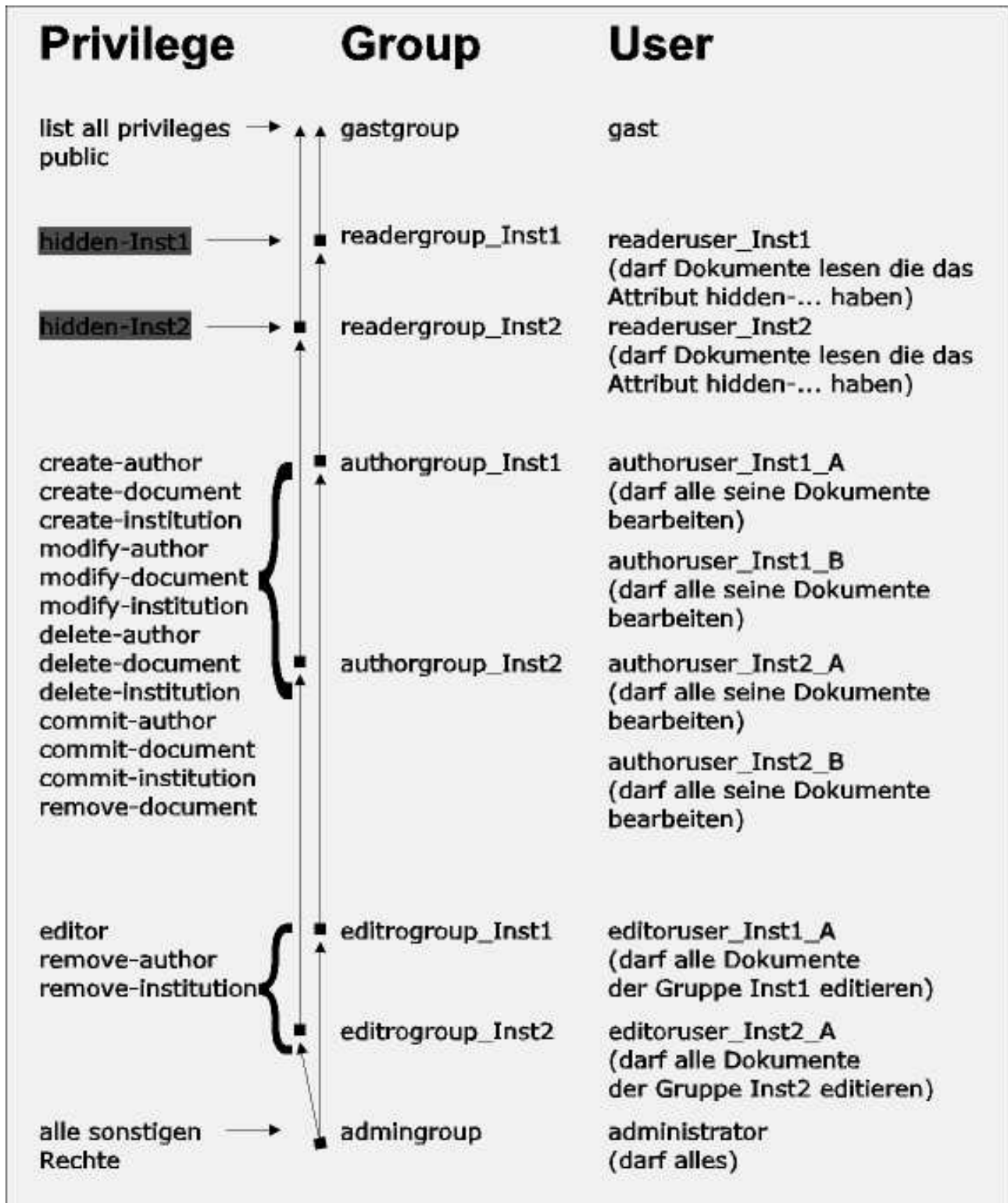
Die nachfolgende Tabelle gibt eine Übersicht über die Beispielbenutzer des MyCoReSamples. Bitte beachten Sie, dass Sie bei einer Adaption des MyCoReSamples dies Gruppen und Benutzer modifizieren bzw. löschen sollten.

User	Passwort	Gruppe
administrator	alleswirdgut	admingroup
reader1A	reader1A	readergroup1
reader2A	reader2A	readergroup2
author1A	author1A	authrogroup1
author1B	author1B	authrogroup1
author2A	author2A	authrogroup2
author2B	author2B	authrogroup2
editor1A	editor1A	editrogroup1
editor1B	editor1B	editrogroup1
editor2A	editor2A	editrogroup2
editor2B	editor2B	editrogroup2

Tabelle 11: Benutzer des MyCoReSample

5.3.4.5 Übersicht der Zusammenhänge von Benutzern, Gruppen und Privilegien im MyCoReSample

Die nachfolgende Grafik soll noch einmal die Zusammenhänge der Kaskaden von Gruppen und Benutzern am konkreten Beispiel verdeutlichen. Für eigene Anwendungen kann diese Modell adaptiert und ausgebaut werden.



5.4 Das SimpleWorkflow-System zur interaktiver Autorensarbeit

Das **SimpleWorkflow**-System wurde entwickelt, um mit einem einfachen Werkzeug die interaktive Autoren- und Editorenarbeit zu ermöglichen und damit eine sinnvolle Arbeit mit einer MyCoRe-Applikation zu ermöglichen. Es ist jedoch so konzipiert, dass es auch über eine Servlet-Schnittstelle in größere Workflow-Engines eingebunden werden kann. Einen Workflow im eigentlichen Sinne

gibt es nur sehr eingeschränkt und in einfachem Ablauf. Weiterführende organisatorische Maßnahmen waren auch nicht Ziel dieser Entwicklung.

Die Komponente wurde in einen Modul verlagert und ist somit durch andere Komponenten ersetzbar. Eine genaue Beschreibung der Details zur Integration finden Sie im Programmer Guide. Die wichtigsten Merkmale dieses Moduls sind:

- Mit dem System kann ein einfacher Eingabe- und Bearbeitungs-Dialog realisiert werden.
- Eingabe und Bearbeitung werden durch eine Rechtekontrolle mittels des **SimpleAccessControl**-System realisiert. Nur berechtigte Benutzer dürfen die Daten manipulieren.
- Die Zwischenspeicherung aller bearbeitet Daten erfolgt auf einem Plattenbereich, so dass bei Fehlern ggf. auch der Administrator direkt eingreifen kann. Daten die erfolgreich in den Server geladen wurden, werden dort wieder gelöscht.
- Das System benutzt die MyCoRe-interne Editor-Komponente.
- Das System basiert auf einer Reihe von Servlets, XML-Seiten und Stylesheets, sowie der Einbindung in die Editor-Formulare.
- Alle Funktionen werden über ein einheitliches Servlet initiiert (MCRStartEditorServlet). Die möglichen Aufrufe sind weiter unten notiert.

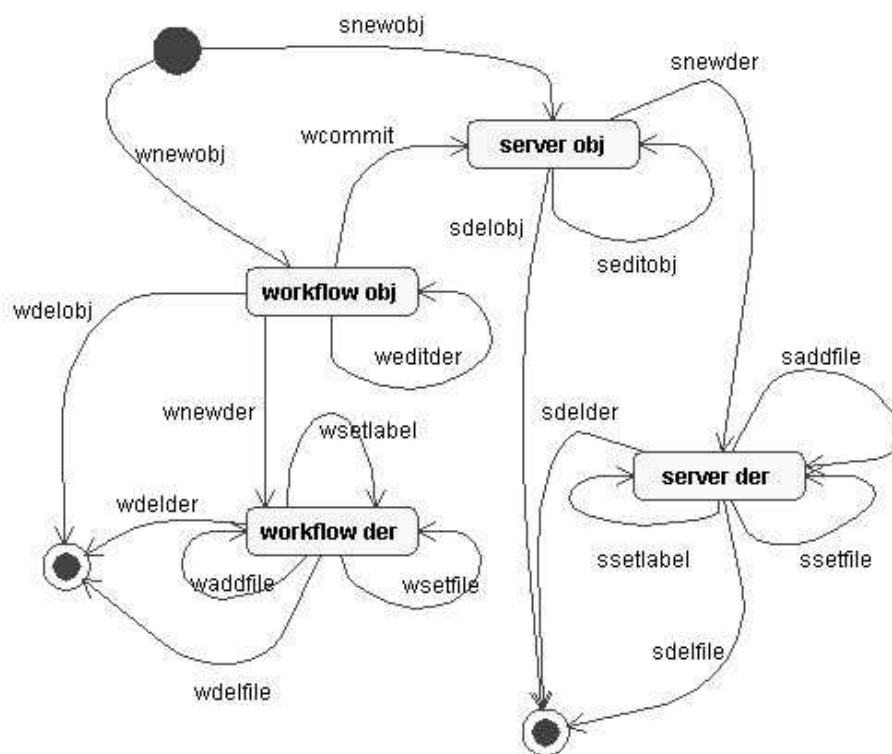


Abbildung 35: Funktionsschema des SimpleWorkflow

5.4.1 Das MCRStartEditorServlet

Dieses Servlet ist der Einstiegspunkt für die Nutzung des SimpleWorkflow-Systems. Von ihm aus werden alle Verarbeitungsprozesse angeschoben. Das Servlet seinerseits startet dann wieder Web-

Dialoge oder führt selbstständig Aktionen aus. Dabei sind die folgenden Startparameter von Interesse:

Parameter	Bedeutung
todo	Zeigt an, welche Aktion auszuführen ist.
type	Gibt den Datenmodell-Typ des Metadaten-Objektes an.
step	Gibt den Verarbeitungsschritt an (z. B. author, editor, commit).
layout	Gestattet eine verfeinerte Angabe des Verarbeitungsschrittes (ist optional).
tf_mcrid	Enthält eine MCRObjektID, welche neu hinzugekommen und/oder dem System noch nicht bekannt ist. Die Gültigkeit wird geprüft.
se_mcrid	Enthält eine MCRObjektID, welche aus einem Datensatz oder ähnlichen Quellen extrahiert wurde und gültig sein sollte.
re_mcrid	Enthält eine weitere MCRObjektID, welche aus einem Datensatz oder ähnlichen Quellen extrahiert wurde und gültig sein sollte (z. B. zugehöriger Metdatensatz).
extparm	Erweiterungsparameter, wie in einigen wenigen Fällen benutzt.

Tabelle 12: Parameter des MCRStartEditorServlets

Die nächsten Tabellen sollen nun eine Übersicht der möglichen Aktionen geben. Jede Aktion ist dabei an ein entsprechendes Privileg gebunden, welches der aktuelle Benutzer gerade haben muss. Hat er das nicht, so wird seine Aktion abgewiesen und gelangt nicht zur Ausführung. Dabei wird noch nach dem Datenmodell-**type** unterschieden, d. h. ein Benutzer muss für genau diesen **type** auch die Berechtigung haben. Die Aktionen unterscheiden sich in dem Ziel-Store, *todo=w...* steht für den Plattenbereich; *todo=s...* arbeitet mit den bereits eingestellten Server-Daten. Der Parameter **layout** ist optional und dient der Verfeinerung der möglichen Arbeitsschritte. Während alle Aktionen, die mit einem **w** beginnen auf dem Plattenbereich (workflow) arbeiten, veranlassen alle Aktionen mit **s** einen Zugriff und Änderungen im Server-System.

Aktion	todo	ID	Privileg	ruft
Anlegen neuer Metadaten	wnewobj	tf_mcrid	create- type	editor_form_ type _[layout_]author.xml
Anlegen eines Neuen Derivates	wnewder	se_mcrid	create- type	MCRStartEditorServlet?todo=waddfile
Hinzufügen neuer Dateien aus dem Upload	waddfile	se_mcrid re_mcrid	create- type	fileupload_new.xml
Bearbeiten von Metadaten	weditobj	se_mcrid	modify- type	editor_form_ type _[layout_]editor.xml
Bearbeiten des Label eines Derivate-Metadaten-Satzes	weditder	se_mcrid re_mcrid	modify- type	editor_form_derivate_editor.xml
Löschen aller Daten eines Objektes	wdelobj	se_mcrid	delete- type	editor_ type _editor.xml
Löschen eines Derivates	wdelder	se_mcrid	delete- type	editor_ type _editor.xml
Löschen einer Datei aus einem Derivate	wdelfile	se_mcrid	delete- type	editor_ type _editor.xml
Setzen der Hauptdatei in einem Derivate	wsetfile	se_mcrid	modify- type	editor_ type _editor.xml
Hochladen eines Datensatzes vom Plattenbereich zum Server	wcommit	se_mcrid	commit- type	MCRQueryServlet mit der ID mit mode=ObjectMetadata

Tabelle 13: Mögliche Aktionen mit dem MCRStartEditorServlet auf dem Plattenbereich

Aktion	todo	ID	Privileg	ruft
Bearbeiten der Metadaten	seditobj	se_mcrid	commit- type	MCRQueryServlet mit der ID mit mode=ObjectMetadata
Bearbeiten des Label der Derivate-Metadaten	seditder	se_mcrid re_mcrid	commit- type	editor_form_commit_derivate.xml
Löschen eines Datenobjekts	sdelobj	se_mcrid	remove- type	editor_deleted.xml
Löschen eines Derivates von einem Datenobjekt	sdelder	se_mcrid re_mcrid	remove- type	MCRQueryServlet mit der ID mit mode=ObjectMetadata
Hinzufügen eines neuen Derivates zu einem Datenobjekt des Servers; Zwischenablage der Daten auf dem Plattenbereich	snewder	re_mcrid	commit- type	MCRStartEditorServlet?todo=saddfile
Hinzufügen von Daten zu einem Derivate aus dem Server; Zwischenablage der Daten auf dem Plattenbereich	snewfile	se_mcrid re_mcrid	commit- type	MCRStartEditorServlet?todo=saddfile
Upload von Datenobjekten in die Zwischenablage	saddfile	se_mcrid re_mcrid	commit- type	MCRStartEditorServlet?todo=scommitder
Laden eines Derivates vom Zischenspeicher in den Server	scommitder		commit- type	MCRQueryServlet mit der ID mit mode=ObjectMetadata
Setzen Des Label in einem Derivate	ssetlabel	se_mcrid re_mcrid	commit- type	MCRQueryServlet mit der ID mit mode=ObjectMetadata
Setzen der Main File Markierung im Derivate	ssetfile	se_mcrid re_mcrid	commit- type	MCRFileNodeServlet mit der ID des Derivates

Aktion	todo	ID	Privileg	ruft
löschen einer Datei aus einem Derivate	sdelfile	se_mcrid re_mcrid	remove- <i>type</i>	MCRFileNodeServlet mit der ID des Derivates

Tabelle 14: Mögliche Aktionen mit dem MCRStartEditorServlet im Server

5.4.2 Abläufe für neue Datenobjekte

Die Abläufe für die Eingabe neuer Datensätze sind praktisch für alle Datenmodelle gleich. Lediglich die Anbindung der Derivate an die Metadaten-Objekte ist nicht immer gegeben. Das hängt allein an der Gestaltung des jeweiligen Datenmodell-Konzeptes für ein Projekt (z. B. haben Personendaten im DocPortal-Projekt keine eigenen Derivate). Wird beim SimpleWorkflow ein neues Objekt eingestellt, so befinden sich alle relevanten Daten vorerst auf einem Plattenbereich, der über die Konfiguration festgelegt wird. Erst wenn das **Commit** zum Server-System ausgeführt wurde, werden die Daten von diesem Zwischenspeicher wieder gelöscht. Jeder Datenmodell-*type* hat dabei in der Regel ein eigenes Verzeichnis innerhalb des Workflow-Plattenbereiches.

5.4.3 Abläufe für Datenobjekte aus dem Server

Wurden Datenobjekte in den Server eingebracht, so steht Benutzern, welche berechtigt sind, die Möglichkeit einer Änderung der Daten und/oder das Löschen der selben frei. Für das Bearbeiten der Daten werden diese zwischenzeitlich auf dem Plattenbereich gespeichert. Bei erfolgreicher Beendigung einer Aktion werden die temporären Daten wieder vom Plattenbereich gelöscht. Im Falle eines Fehlers kann über den Zugriff auf den Plattenbereich (Workflow) und entsprechender Aktionen der Fehler behoben werden. Alle Commits sind als Update ausgelegt, so dass ältere Versionen im Server auch bei einem Commit vom Workflow als Folge eines Fehlers überschrieben werden. Einzelheiten zu den Abläufen finden Sie im Programmer Guide.

5.4.4 Einbindung in das SimpleAccessControl-System

Im **SimpleWorkflow** kommt die Zugriffsrechteverwaltung des SimpleAccessControl-System zur Wirkung. Dabei werden folgende Schritte geprüft:

1. Hat der aktuelle Benutzer das erforderliche create-*type*, modify-*type*, delete-*type* oder commit-*type* Privileg?
2. Ist der aktuelle Benutzer als Autor (User:....) im Datensatz in <servflags> eingetragen?
3. Wenn nicht, hat der Benutzer das Privileg editor und gehört er einer Gruppe an (mittelbar oder unmittelbar), welcher der User angehört?

Sind diese Kriterien nicht erfüllt wird die Verarbeitung der Aktion mit dem Hinweis, dass der Benutzer nicht berechtigt ist, abgelehnt. Ausnahme ist der Benutzer **administrator**, welcher immer Zugriff hat.

6.Hints & Tips / Troubleshooting

6.1 Eine zweite Instanz von eXist auf einem System einrichten

Wenn Sie eine zweite Instanz von eXist auf Ihrem System (Computer) einrichten möchten, so ist es erforderlich, dass diese mit anderen Port-Nummern arbeitet. Andernfalls kommt es zum Konflikt und eXist lässt sich nicht starten. Dies ist unabhängig von der User-ID, unter der eXist laufen soll.

Folgende Arbeiten sind durchzuführen:

- Packen Sie eine Original-eXist-Distribution aus.
- Führen Sie alle Anpassungen wie oben beschrieben durch.
- Im Wurzel-Verzeichnis von eXist (\$EXIST_HOME) ist in der Datei *client.properties* der Port von 8081 auf 8091 (oder einen anderen) zu ändern. Auch im Wert der **alternateURI** ist dies erforderlich.
- In *build.xml* ist ebenfalls der Port unter dem Wert **location** auf 8090 (oder einen passenden) zu ändern.
- Um die Änderungen vollständig zu machen, sollten Sie auch die Datei *backup.properties* entsprechend anpassen.
- Im Source-Code sind die Klassen `org.exist.Server.java`, `org.exist.ServerShutdown.java`, `org.exist/exist.xsl`, `org.exist.InteractivClient.java` und `org.exist.client.InteractivClient.java` anzupassen und mit `ant` neu zu compilieren.

Nach dem Start des eXist-Servers steht dieser nun unter dem neuen Port zur Arbeit bereit.