

MyCoRe stellt sich vor

http://www.mycore.de/



Frank Lützenkirchen

Universität Duisburg-Essen luetzenkirchen@ub.uni-duisburg-essen.de



MyCoRe: Entstehung

- Ursprung: **miless**, Multimedialer Lehr- und Lernserver Essen, seit Ende 1997 entwickelt
- Erste Nachnutzer: Jena (Urmel, DB Thüringen),
 Freiburg (Freimore), Leipzig/IBM (Bach Digital)
- Zusätzliche **Anforderungen** entstehen, z.B. konfigurierbares Datenmodell, Anpassbarkeit
- Gemeinsame Entwicklung eines Software-Kerns auf Java- und XML/XSL Basis, als Open Source (GNU GPL), der gemeinsame Funktionen zur schnelleren Entwicklung mächtiger lokaler Dokumenten- und Publikationsserver enthält
- MyCoRe = My Content Repository
 MyCoRe = Miless Community Core



MyCoRe: Beteiligte

Sehr aktiv an der Entwicklung beteiligt zur Zeit

RZ Leipzig, RZ/UB Jena, RZ/UB Rostock, RZ Hamburg, UB Duisburg-Essen

 Anwender von miless / MyCoRe, weitere, teils ehemalige Entwickler (Projekte, spezielle Module, Diplomarbeiten):

UB Münster, RZ Freiburg, UB Brauschweig, FU Berlin, RZ Bochum, UB Paderborn, UB Düsseldorf, UB Halle, TU München, RZ Bonn

Kooperation mit IBM

IBM Content Manager Programmierung Kontakt zum Entwickler-Labor, HowTo Support



MyCoRe: Organisation

- Geschäftsstelle: info@mycore.de
 AnsprechpartnerIn für Interessenten, Anwender und Entwicker, Koordination, Redaktionelle Arbeiten, Öffentlichkeitsarbeit, Tests
- Mailing-Listen: mycore-developer, mycore-user
- Telefonkonferenzen: ca. alle 14 Tage
- Entwicklertreffen: ca. 2-3 mal jährlich
- CVS Versionsmanagementsystem
- Webserver www.mycore.de
- Sourceforge Bug Tracking System
- Wiki: Notizen, Ideen, Entwürfe



www.mycore.de





CVS Repository





MyCoRe Binärdistribution

- Kennenlernen von MyCoRe & DocPortal über einfaches Installationsprogramm (IzPack)
- Vollständige DocPortal Anwendung mit bereits geladenen Inhalten, Dateien und Benutzern
- Integrierte Datenbank (HSQLDB), integrierter
 Web Application Server (Jetty)





MyCoRe 1.3 Beta Download

Gepackt als ZIP- oder tar.gz-Distribution,
 Download über SourceForge oder mycore.de

 Alternativ direkt über einen CVS Client aus dem Versionsmanagementsystem cvs -d anoncvs@server.mycore.de/cvs co mycore

mycore MyCoRe Kernsystem
 docportal Beispielanwendung, typischer

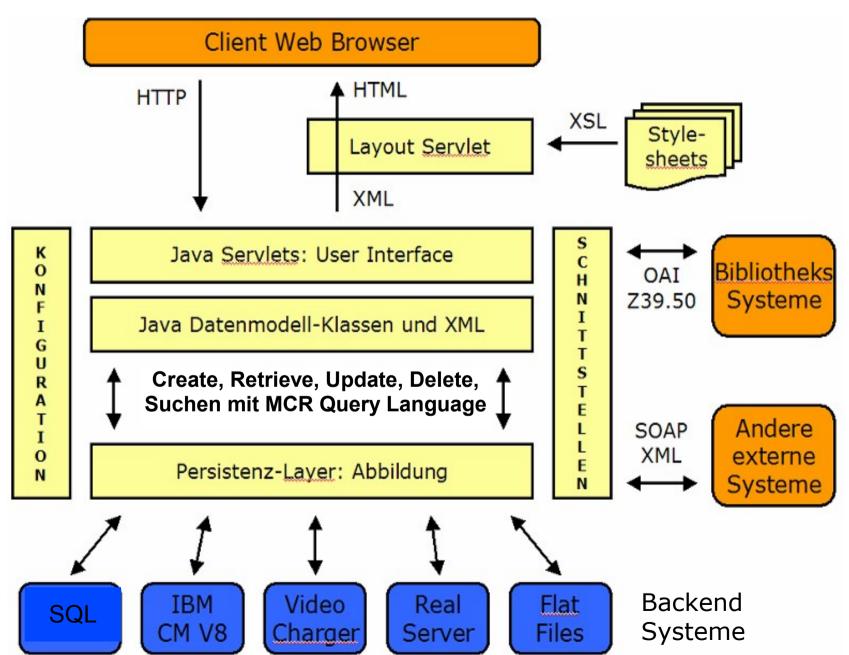
Dublin Core Dokumentenserver

content Zusätzliche Beispielinhalte

Overview.pdf
 Überblick Funktionalität und Leistungsumfang
 UserGuide.pdf
 Installation und Konfiguration von DocPortal
 ProgGuide.pdf
 Dokumentation für Anwendungsentwickler



MyCoRe Architektur





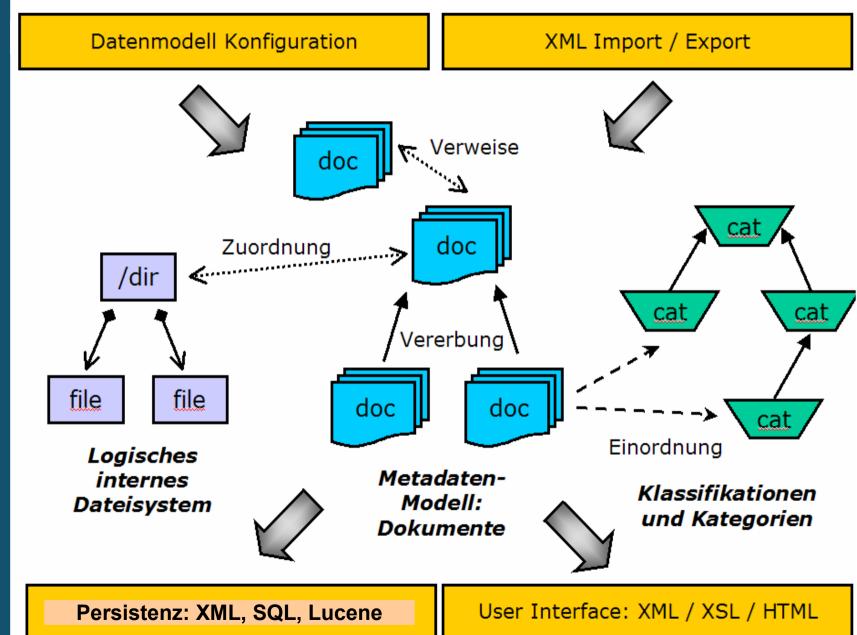
Backend-Systeme: Datenhaltung

- Erforderlich: Relationale Datenbank

 a) direkte Anbindung über Java JDBC / SQL:
 MySQL, IBM DB2, HSQLDB (bereits integriert)
 b) indirekte Anbindung über Hibernate:
 unterstützt nahezu alle gängigen RDBMS
- Erforderlich: Speicherort für **Dateien und Metadaten**:
 IBM Content Manager, SQL BLOBs, *lokales Dateisystem*,
 entferntes Dateisystem (FTP, SFTP, WebDAV, CIFS)
- Optional: Streaming Server (Helix, VideoCharger)
- Weitere Komponenten werden mitgeliefert, z.B.
 Suchmaschine: Apache Lucene / Hibernate / SQL
- Erforderlich: Java **Web Application Server** z.B. Tomcat, IBM WebSphere, Jetty (integriert) zusätzlich empfohlen: Web Server (z.B. Apache)
- Java ab Version 1.4, Apache Ant, CVS Client
- OS: Linux, Windows, Solaris, AIX, Mac OS



MyCoRe Datenmodell





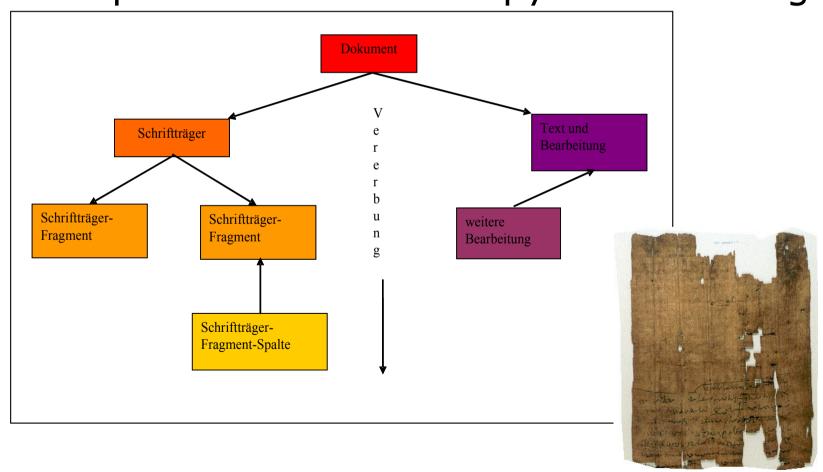
Konfigurierbares Datenmodell

- Metadatenmodell ist frei konfigurierbar
- **Objekttypen** sind frei definierbar, z.B. *Dokument, Person, Institution*
- Je Objekttyp: Konfiguration der Felder: Name, Typ, Wiederholbarkeit
 z.B. Titel, MCRMetaLangText, 1-n mal
- Vordefinierte Basisdatentypen, z.B. LangText, ISO8601Date, Number, Boolean, Link, Classification, ...
- Erweiterung um eigene Datentypen möglich (z.B. GIS-Koordinaten, ...)



Datenmodell: Beziehungen

- Hierarchien mit Metadaten-Vererbung
- Verweise zwischen Objekttypen
- Beispiel: Datenmodell Papyri-Anwendung





Datenmodell: Klassifikationen

- Klassifikation: Hierarchie von Kategorien oder flache Liste, z.B. Dewey Dezimal-Klassifikation, Publikationstyp, Organisationsstruktur der Universität, Fachklassifikationen wie PACS, MSC
- Objekte (Dokumente, Personen) werden Klassifikationskategorien zugeordnet
- Einfacher Import über XML-Dokumente
- Suche in Kategorien oder Hiearchien
- Browsen über Kategorien
- Online-Klassifikations-Editor



Datenmodell: Klassifikationen

```
←□ [ 22 Dok.]Allaemeines
        4 Dok.]Wissenschaft und Kultur allgemein
     __17 Dok.]Schrift, Buch, Bibliothek, Information und Dokumentation
        O Dok.]Nachschlagewerke, Bibliographien
        O Dok. 1Adreßbücher. Telefonbücher
       O Dok.]Kalender
       O Dok. 1Publizistik

■[ ____0 Dok.]Kinder- und Jugendliteratur.

       0 Dok.]Comics, Cartoons, Karikaturen
🖅 [ 18 Dok.]Philosophie, Psychologie
 ■ 4 Dok.]Grenzgebiete der Wissenschaft und Esoterik
    6 Dok.]Philosophie

■[ __16 Dok.]Psychologie

        5 Dok.]Religion, Theologie
    28 Dok. Sozialwissenschaften
[ _405 Dok.]Mathematik, Naturwissenschaften
      19 Dok. Natur, Naturwissenschaften allgemein
     26 Dok.]Mathematik
    _337 Dok.]Informatik, Datenverarbeitung
     24 Dok.]Physik, Astronomie
    11 Dok.]Chemie
        6 Dok.]Geowissenschaften

■I 24 Dok.]Biologie

□ 180 Dok.]Angewandte Wissenschaften, Medizin, Technik
7 Dok.]Kunst, Kunstgewerbe, Photographie, Musik, Spiel, Sport
      10 Dok.]Sprach- und Literaturwissenschaften, Belletristik
      20 Dok. IGeographie, Geschichte
       0 Dok.]Zusätzliche Gruppen
```



Datenmodell: Klassifikationen

Dewey Dezimal-Klassifikation (Insgesamt 526 Dokumente nach dieser Klassifikation eingeordnet)

[_21 Dok.] 000 Informatik, Informationswissenschaft, allgemeine Werke	
[6 Dok.] 100 Philosophie und Psychologie	
[299 Dok.] 500 Naturwissenschaften und Mathematik [299 Dok.] 500 Naturwissenschaften und Mathematik	
☐ [8 Dok.] 550 Geowissenschaften	
☐ [_91 Dok.] 570 Biowissenschaften; Biologie ☐ [_91 Dok.] 570 Biowissenschaften; Biologie	
⊕ [2 Dok.] 800 Literatur	

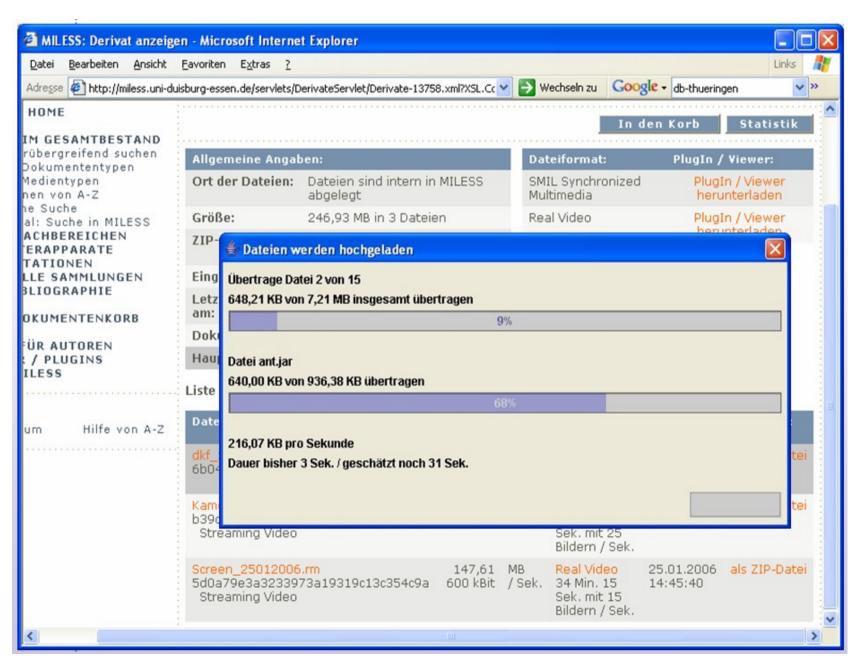


Internal Filesystem (IFS)

- Dokumente können nicht nur "flache" Dateien, sondern auch ganze Verzeichnisstrukturen enthalten (auch mehrere 1000, > 1 GByte)
- IFS verwaltet Datei- und Verzeichnisdaten, z.B. Dateityp, Größe, MD5 Prüfsumme, ... in eigenen Datenbanktabellen
- FileContentStores speichern Dateiinhalte:
 - Beliebige Dateisysteme: lokal oder entfernt angebunden über FTP, SCP/SFTP, CIFS / Samba, WebDAV (Apache Commons VFS), IBM Content Manager 8
- Upload: Über HTML Formular oder Applet
- Download: ZIP-Generierung on-the-fly
- Integrierte Volltextsuche, Image Viewer



Datei-Upload mit Java-Applet





Audio-/Video-Streaming

- Helix Universal Server, IBM VideoCharger
- Streaming von AVI, MPEG 1, MPEG2, Real,
 Windows WMV, Apple Quicktime, WAV, MP3
- Audio-/Video-Streaming-Server sind für den Endnutzer transparent integriert
 - Benutzer lädt Video in MyCoRe hoch
 - MyCoRe erkennt anhand des Dateityps, dass das Video auf dem Streaming Server (File Content Store) gespeichert werden soll
 - Benutzer klickt auf Videodatei, Player startet
- Einzelne Szenen eines Videos lassen sich gezielt einzeln über URL wiedergeben



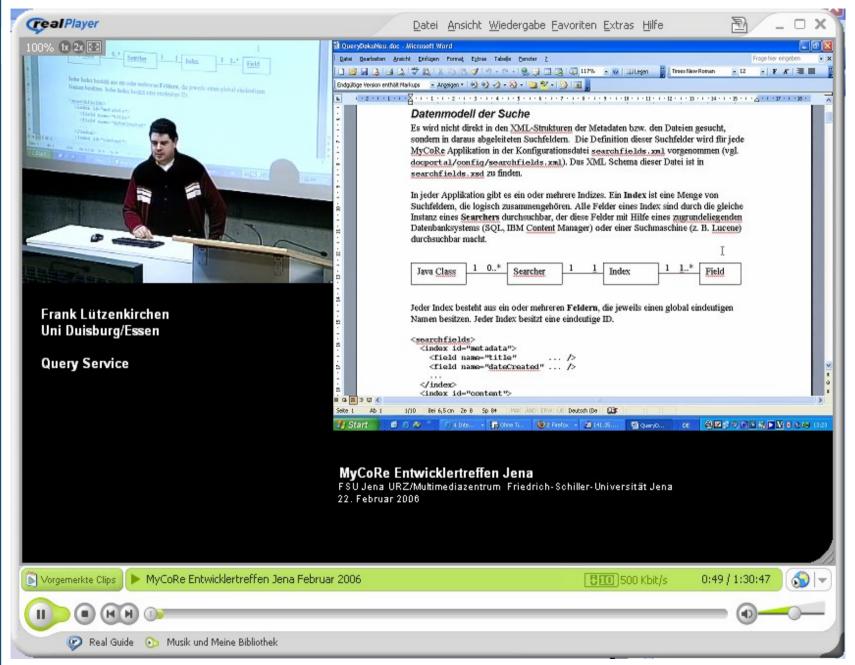
Audio-/Video-Streaming

 Übernahme technischer Metadaten (Bitrate, Framerate, Dauer etc.) vom Streaming Server

Dateiformat:	Größe der Datei:		Dateityp:	
encodings/tracheostomie-DSL-350-kBit.rm	21,49		Real Video	
Streaming Video	350 kBit		8 Min. 30 Sek. mit 25 Bildern / Sek.	
encodings/tracheostomie-Dual-ISDN-106-kBit.rm	6,57		Real Video	
Streaming Video	106 kBit		8 Min. 30 Sek. mit 25 Bildern / Sek.	
encodings/tracheostomie-Modem-034-kBit.rm	2,14	MB	Real Video	
Streaming Video	34 kBit	/ Sek.	8 Min. 30 Sek. mit 25 Bildern / Sek.	
encodings/tracheostomie-MPEG1-1300-kBit.mpeg	82,71		MPEG 1 Video	
Streaming Video	1,3 MBit		8 Min. 29 Sek. mit 25 Bildern / Sek.	



Audio-/Video-Streaming





MyCoRe Query Service

- Abbildung von Daten auf Suchfelder
- Wahlmöglichkeit zwischen mehreren Implementierungen, basierend auf
 - Apache Lucene (Open Source, Volltextsuche)
 - SQL (Relationale Datenbank)
 - XSL im Hauptspeicher (zu Testzwecken)
 - IBM Content Manager (in Entwicklung)
- Verteilte Suche über mehrere Server via WebServices (SOAP) Schnittstelle oder Java-Schnittstelle
- Sortierung der Trefferliste, Suche verfeinern, "Sie haben gesucht nach…"

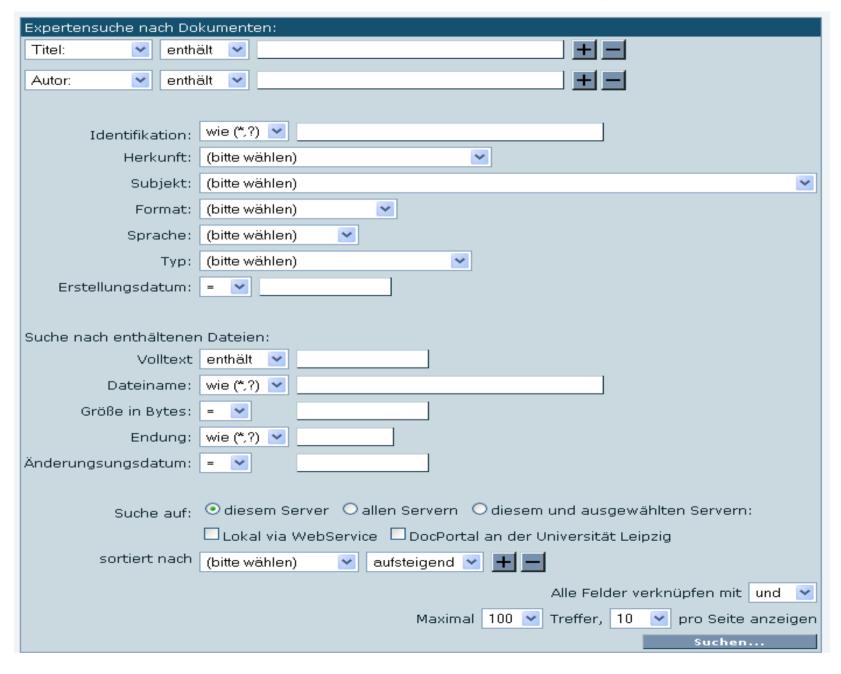


Was ist durchsuchbar?

- Metadaten von Dokumenten, Personen:
 z.B. Titel, Autor, Datum
- Metadaten von gespeicherten Dateien z.B. Dateiname, Typ, Größe
- Volltexte von gespeicherten Dateien z.B. PDF, OpenOffice, HTML
- XML innerhalb gespeicherter Dateien z.B. Excel-Tabelle als XML-Dokument manifest.xml aus E-Learning Scorm Modul
- Über **Erweiterungen** (Event Handler) extrahierte technische Metadaten, z.B. EXIF-Daten von Bilder, ID3 von MP3s



Suchmasken: frei konfigurierbar





MyCoRe Query Language: MCRQL

- **Einfache Suchausdrücke:** title contains 'Optik'
- Zusammengesetzte Suchausdrücke: Kombination mit UND/ODER/NICHT:

```
(objectType = "document") and
((title contains "Regenbogen") or
(fileSize > 500000))
```

- **Suchoperatoren**: =, <, >, <=, >=, contains, phrase (Phrasensuche), like (Suche mit Platzhaltern *,?)
- Suchfeldtypen: text, name, identifier, date, time, timestamp, boolean, integer, decimal



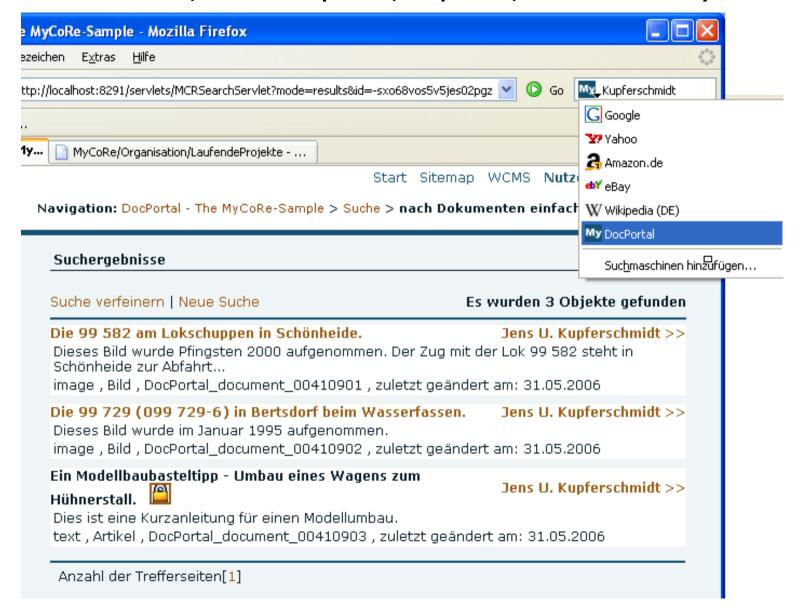
Expertensuche

Suchausdruck in MCRQL einge	eben:				
(objectType = "document") and ((title contains "Regenbogen") or (author contains "Kupferschmidt"))					
	Maximal 100 💌 Treffer, 10 💌 pro Seite anzeigen				
	Suchen				
Beispiele:					
title contains Regenbogen (objectType = "document") a "Kupferschmidt"))	nd ((title contains "Regenbogen") or (author contains				
Syntax:					
	(field2 operator2 value2) [and ()] field2 operator2 value2) [or ()]				
Field: Type:	Index:				
id identifier	metadata				
objectType identifier	metadata				
link identifier	metadata				
parent identifier					
·	metadata				
allMeta text	metadata				



Such-Plug-in für Webbrowser

für Firefox, Netscape 6, Opera, SeaMonkey





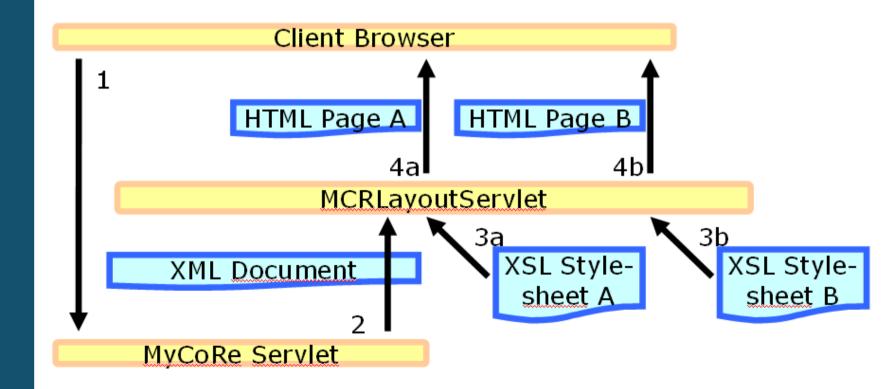
MyCoRe Benutzer-Interface

- Command Line Interface (CLI): auf dem Server, zu administrativen Zwecken, z.B. Batch-Import/Export vieler Daten via XML
- Web Interface: Ajax-Ready!
 - Java Servlets erzeugen XML-Ausgaben
 - Serverseitige Generierung von HTML-Seiten aus diesen Ausgaben über XSL Stylesheets
 - Dadurch Trennung von Daten, Layout und Logik, einfachere/getrennte Anpassbarkeit
- Alternative Implementierung:
 - Java Server Pages (JSP)
 - HTML-Seiten, die Java-Code enthalten



Web Content Management System

- WYSIWYG-Editor für statische Webseiten erlaubt einfaches Bearbeiten von HTML
- Navigationsmenü online bearbeitbar
- Wechsel von Layout-Templates und CSS





I18N: Mehrsprachigkeit

- Mehrsprachige Web-Oberfläche
- Bezeichner sind in sprachabhängigen Dateien abgelegt (Java Resource Bundle):

```
messsages_de.properties, messages_en.properties
metaData.document.title = Titel
metaData.edit = Bearbeiten
```

 Statische Webseiten enthalten sprachabhängige Abschnitte:

```
<section xml:lang="de" title="Daten bearbeiten">
...
</section>
<section xml:lang="en" title="Edit data">
...
</section>
```



MyCoRe Editor Framework

- Konfigurierbares Datenmodell erfordert konfigurierbare Eingabemasken, Suchmasken
- Such- und Eingabemasken werden über eine XML-Definitionsdatei erstellt
- MyCoRe Editor Framework erzeugt ein XML-Dokument aus den Eingaben im HTML-Formular bzw. stellt die XML-Quelldaten im Formular da
- Einfache Abbildung zwischen XML und Formular
- **Eingabevalidierung**, z.B. required, min/max, syntaktische Korrektheit von Email-Adressen...
- Einfache Wiederholbarkeit von Feldern oder ganzer Formularbereiche



MyCoRe Editor: Beispiel





Benutzer- und Rechteverwaltung

- Benutzer (Login-User)
- Benutzer gehören Gruppen an
- Zugriffsrechte auf Dokumente (lesen, bearbeiten) können auf Benutzer oder Gruppen, oder auf IP-Adressbereiche beschränkt werden
- Privilegien regeln z.B. das Recht, ein neues Dokument einstellen zu dürfen



URNs: Persistent Identifier

- Unterstützung von URNs (urn:nbn:de:...)
 vgl. www.persistent-identifier.de
- **Erzeugung**: wählbarer Algorithmus Zeitstempel, fortlaufende Nummer, ...

```
MCR.URN.SubNamespace.Essen.Prefix =
    urn:nbn:de:hbz:465-
MCR.URN.SubNamespace.Essen.NISSBuilder =
    org.mycore.services.urn.MCRNISSBuilderDateCounter
MCR.URN.SubNamespace.Essen.NISSPattern =
    yyyyMMdd-HHmmss-
erzeugt z.B. urn:nbn:de:hbz:465-20060912-102358-7
```

- Lokaler Resolver (URN -> Dokument)
- Registrierung bei DNB über OAI (Epicur)





Modul: OAI-Schnittstelle

- OAI Data Provider Implementierung:
 OAI-PMH 2.0 Protocol for Metadata Harvesting
- Abbildung beliebiger Klassifikation auf OAI Sets
- Unterstützung verschiedener Metadatenformate
 OAI DC

Dublin Core

XMetaDiss

Automatische Meldung von Dissertationen und Hochschulschriften an die Deutsche Nationalbibliothek

Epicur

Automatische Registrierung von URNs (urn:nbn:de, Persistent Identifier) bei der Deutschen Nationalbibliothek

Datenformate und Ausgaben anpassbar (XSL)



Modul: Z39.50-Server

- Implementierung eines Z39.50 Servers auf Basis von **JZKit** (Open Source)
- Implementierung: FU Berlin, Duisburg-Essen
- Abbildung von Z39.50 Anfragen in PQF (Prefix Query Format) auf MyCoRe Query Language
- Derzeit wird nur ein Teil der Z39.50 Abfragesprache unterstützt
- Abbildung von Z39.50 Attributen auf MyCoRe Query Suchfelder über Konfigurationsdatei
- Transformation von MyCoRe Metadaten ins MAB2 Endlosformat über ein XSL Stylesheet



Modul: Image Viewer

- Bildbetrachter im Browser mit serverseitiger Transformation der Bilder
- Verwendet Java Advanced Imaging API
- Geeignet z.B. für hochauflösende Digitalisate

- Größenanpassungen: Skalieren
- Bereitstellen von **Ausschnitten** (Sie wählen im Übersichtsbild einen anzuzeigenden Bereich)
- Generieren von Thumbnails
- Caching generierter Bilder
- Skalierung on-the-fly bei Bedarf



Vielen Dank für Ihre Aufmerksamkeit!

