

# Archivierung von elektronischen Dokumenten und Websites in Rheinland-Pfalz

Ein Projekt von HBZ Köln und RLB Koblenz

Nestor-Workshop, Göttingen 1./2.6.2004

Lars Jendral, RLB Koblenz



HBZ



RHEINISCHE  
LANDESBIBLIOTHEK

## Edoweb-Sammelauftrag

- Flüchtige elektronische Inhalte wie z.B. Websites und Newsletter, die sogenannten Netzpublikationen sollen dauerhaft vorgehalten werden
- Erschließung und Bereitstellung von landeskundlich relevanten elektronischen Objekten und Inhalten aus der Region
- Vorbereitung für eine Langzeitarchivierung der archivierten Objekte durch Festlegung von relevanten Metadatenätzen



## Edoweb-Historie

- August 2002: Vereinbarung zwischen RLB und HBZ zum Aufbau eines Archivservers für elektronische Pflichtexemplare und landeskundliche Websites
- Januar 2003: Beginn des Projektbetriebs auf Basis des Dokumentenverwaltungssystem OPUS
- Dezember 2003: Freischaltung für die Öffentlichkeit als edoweb:  
<http://www.edoweb-rlp.de>



- **RLB-Dokserver** für „klassische“ eDokumente
- Dokumente werden erschlossen wie gedrucktes Material
- **RLB-Webserver** für Websites
- Dateien einer Website werden mittels des Offline-Browsers Htrack eingesammelt
- Metadaten werden übernommen und nach Dublin Core-Schema aufbereitet



## RLB-Dok und RLB-Web gemeinsam:

- weitergehende Erschließung durch DDC, Systematikhauptgruppen der Landesbibliographie sowie Geo-Schlüssel
- geltende Urheberrechtssituation erfordert, die Genehmigung des Rechteinhabers einzuholen: wird per Formmail abgewickelt
- jedes Dokument erhält eine URN als persistent identifier



## Benutzung

- Suche in bibliographischen Metadaten
- Thematisches Browsen
- Volltextsuche (geplant)



## Weiterentwicklung des Systems:

- ⇒ Einsatz von Suchmaschinentechologien erfordert geringeren Katalogisierungsaufwand und
- ⇒ hält Personalaufwand für neuen Sammelauftrag in vertretbaren Grenzen
- ⇒ Deshalb: keine Beschränkung allein auf qualitativ hochwertige und intellektuell erschlossene Websites



## Stand Mai 2004

- Dokserver: 70 Dokumente, 1108 unbearbeitet in Warteposition
- Webserver: 350 Dokumente, 200 bearbeitet im temporären Bereich (Schwerpunkte: Verbandsgemeinden, Weltkulturerbe Mittelrhein)
- Sommer 2004: Archivserver für elektronische Amtsdruckschriften aus Rheinland-Pfalz
- Bestrebungen von Bund und Ländern, Pflichtexemplargesetze auf elektronische Publikationen auszuweiten



## Vorläufige Ergebnisse

- Beim Einsammeln von dynamisch erzeugten Websites und solchen Sites, die mit Hilfe eines Content Management Systems (CMS) erstellt werden, besteht Forschungs- und Handlungsbedarf
- Verwendetes Webarchivsystem muß hierarchische Strukturen abbilden können
- Exakt gefasste Sammelrichtlinien bedürfen weiterer Praxiserfahrungen



## Aussicht

- Erneutes, automatisches Einsammeln bereits erfasster Websites in regelmäßigen Abständen
- Einbindung der Metadaten via OAI-Service-Provider „MEIND (Metadata on Internet Documents)“ in neue Version der Digitalen Bibliothek
- Entwicklung gemeinsamer Standards für technische Metadaten zur Langzeitarchivierung mit DDB, HBZ, BSZ und baden-württembergischen Landesbibliotheken




Opus-Gatherer - Netscape

Datei Bearbeiten Anzeigen Gehe Lesezeichen Extras Fenster Hilfe

http://ribweb.opus.hbz-nrw.de/cgi-bin/useragent/hittrack/opus\_gatherer.cgi

Opus-Gatherer



**Website Gatherer der Rheinischen Landesbibliothek**

RHEINISCHE  
LANDESBIBLIOTHEK


## Konfiguration

Sie wollen eine Website archivieren. Dazu müssen Sie uns erstmal Information über die Website geben:

- Geben Sie die URL der Website ein:

**ACHTUNG!** Es können nur Websites, nicht einzelne HTML-Seiten heruntergeladen werden  
 Es werden nur Dateien, die in der Hierarchie unterhalb dieser URL stehen, geholt.

- Gibt es Dateien oder Verzeichnisse unterhalb dieser URLs, die Sie nicht archivieren wollen? Geben Sie die URL zu diesem Verzeichnis oder zu dieser Datei an:
- Geben Sie die Endungen (z.B., ".doc") von Dateien an, die Sie nicht in das Archiv aufnehmen wollen. Geben Sie die Endungen ohne Anführungszeichen und durch Leerzeichen getrennt:
- Wieviele Ebenen tief soll archiviert werden? Keine Angabe = keine Beschränkung
- Verlangt die Site einen Benutzernamen und Passwort?  
 Benutzername:   
 Passwort:

 Dokument: Done (0,22 Sek.)





## Ändern eines Eintrags in TEMP oder Übernahme in OPUS

### Zusätzliches Autorenfeld anfordern

Originaltitel:

Englischer Titel:

Deutscher Titel:

Urheber:

Sonst. Beteiligter:

Sonst. bet. Inst.:

Schlagwörter SMD:

SW unkontr. deutsch:

SW unkontr. englisch:

Abstract Originalsprache:

Braubach: Verbandsgemeinde am Fuße der Marksburg

Fremdenverkehrsamt Braubach

Rhein-Lahn-Kreis

SMD

deutsch



FLDL

RHEINISCHE  
LANDESBIBLIOTHEK

Abstract in einer weiteren Sprache:

Erstellungsjahr:   
Dokumentart:   
DDC-Sachgruppe:   
Sprache:   
Quelle:   
Identifizier:   
email:   
Gültig bis:  (Format 31.01.2000, 0 für unbeschränkt)  
RPB - Systematik:   
Geografische Systematik:

Volltext: [469-index.html](#)  
Publicationsdatum: 28.11.2003  
Datum letzte Änderung: 08.03.2004  
RLB-Webarchiv-Ident-Nr.: 469  
Status:




HBZ



 [Suche in den bibliographischen Daten über die Dokumente](#)

 [Volltextsuche in den Dokumenten selbst](#)

 [Browsen - Überblick nach Fächern und Dokumentarten](#)

[Home](#) | [Suchen](#) | [Veröffentlichen](#) | [Hilfe](#) | [Viewer](#)

Fragen und Anregungen an [www.kam@rlb.de](mailto:www.kam@rlb.de), 11.04.03

