

Praktische Erfahrungen mit der Webarchivierung:

Das Baden-Württembergische Online-Archiv (BOA)

97. Deutscher Bibliothekartag Mannheim 2008
„Wissen bewegen - Bibliotheken in der
Informationsgesellschaft“

Themenkreis 9: Kulturelles Erbe
Forum Langzeitarchivierung:
Aktuelle Entwicklungen und
praktische Anwendungsfelder

Stefan Wolf, Bibliotheksservice-Zentrum
Baden-Württemberg, Konstanz

Die Zukunft
ist nicht in Stein
gemeisselt!

Dinge bewegen sich!

Wie können wir den
Wandel handhaben?



BOA: Baden-Württembergisches Online-Archiv

Websites und Publikationen im Internet aus oder über Baden-Württemberg werden im "Baden-Württembergischen Online-Archiv" (BOA) gesammelt, erschlossen, der Öffentlichkeit zur Verfügung gestellt und für die Zukunft archiviert. Träger des Projekts sind die Badische Landesbibliothek in Karlsruhe, die Württembergische Landesbibliothek in Stuttgart, das Landesarchiv Baden-Württemberg und das Bibliotheksservice-Zentrum Baden-Württemberg.

Viele dieser Informationen sind heute über das Internet erhältlich, die früher als Druckschriften in den Landesbibliotheken oder als Behördenschriftgut im Landesarchiv gesammelt wurden. BOA stellt diese Dokumente stabil und dauerhaft zur Verfügung.

BOA steht in der Entwicklung. Bitte haben Sie Verständnis, dass noch nicht alle Funktionen in vollem Umfang realisiert sind. Aus technischen Gründen liegen die Online-Ressourcen der beiden Landesbibliotheken und des Landesarchivs Baden-Württembergs für eine Übergangszeit in getrennten Depotsystemen.

Den Zugang zu den Dokumenten des Landesarchivs finden Sie hier:

[BOA Landesarchiv](#)

Den Zugang zu den Dokumenten der Badischen Landesbibliothek und der Württembergischen Landesbibliothek finden Sie hier:

[BOA Landesbibliotheken](#)

Ansprechpartner:

- BLB: [Christoph Höck](#)
- BSZ: [Stefan Wolf](#)
- Landesarchiv: [Christian Keitel](#), [Kai Naumann](#)
- WLB: [Angelika Schütt-Hohenstein](#)

- 2002: Workshop Langzeitverfügbarkeit digitaler Dokumente, DNB:
„Die Teilnehmer des Workshops sehen großen Handlungsbedarf bei der Sicherung der Langzeitverfügbarkeit digitaler Objekte. Sie sind der Überzeugung, dass die sich dabei ergebenden Aufgaben als Gemeinschaftsaufgabe arbeitsteilig, kooperativ und ökonomisch gelöst werden müssen.“
- Startschuß für Nestor, aber auch für BOA:
In enger Kooperation von BLB, WLB und BSZ soll ein Depot- und Portalsystem aufgebaut werden, dessen Ziel die langfristige Archivierung und Verfügbarkeit von Online-Dokumenten ist. Dies geschieht in enger Kooperation mit dem Programm der DNB zur Langzeitarchivierung.
- Entscheidung für Eigenentwicklung auf Open-Source-Basis
Betrieb seit Mitte 2003 für Badische LB und Württembergische LB
- Anfang 2006: Landesarchiv Baden-Württemberg ist dritter Kooperationspartner

browserbasierte Anwendung

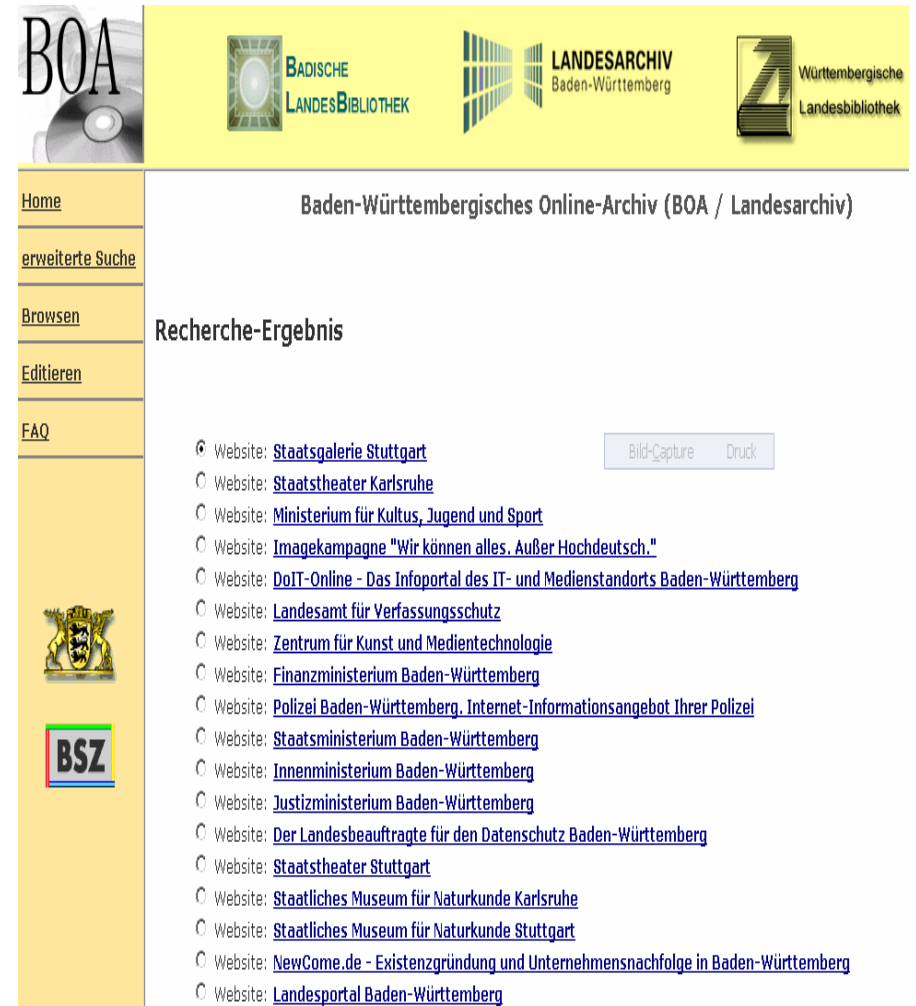
hierarchiefähig

Httrack als Offline-Browser

Rechtsgrundlage: Archiv- und Pflichtexemplargesetz

LA sammelt 64 Webseiten von Landeseinrichtungen regelmäßig

WLB und BLB sammeln im Rahmen Landesbibliographie und Pflichtexemplar



BOA

BADISCHE LANDESBIBLIOTHEK

LANDESARCHIV Baden-Württemberg

Württembergische Landesbibliothek

Home

erweiterte Suche

Browsen

Editieren

FAQ

Baden-Württembergisches Online-Archiv (BOA / Landesarchiv)

Recherche-Ergebnis

Website: [Staatsgalerie Stuttgart](#)

Website: [Staatstheater Karlsruhe](#)

Website: [Ministerium für Kultus, Jugend und Sport](#)

Website: [Imagekampagne "Wir können alles. Außer Hochdeutsch."](#)

Website: [DoIT-Online - Das Infoportal des IT- und Medienstandorts Baden-Württemberg](#)

Website: [Landesamt für Verfassungsschutz](#)

Website: [Zentrum für Kunst und Medientechnologie](#)

Website: [Finanzministerium Baden-Württemberg](#)

Website: [Polizei Baden-Württemberg, Internet-Informationsangebot Ihrer Polizei](#)

Website: [Staatsministerium Baden-Württemberg](#)

Website: [Innenministerium Baden-Württemberg](#)

Website: [Justizministerium Baden-Württemberg](#)

Website: [Der Landesbeauftragte für den Datenschutz Baden-Württemberg](#)

Website: [Staatstheater Stuttgart](#)



Website: [Staatliches Museum für Naturkunde Karlsruhe](#)

Website: [Staatliches Museum für Naturkunde Stuttgart](#)

Website: [NewCome.de - Existenzgründung und Unternehmensnachfolge in Baden-Württemberg](#)

Website: [Landesportal Baden-Württemberg](#)

Bild-Capture Druck



- **Homepages**

Harvesting mit Offline-Browser (httrack)

Beispiel: BOA (LA): [Finanzministerium BW](#)



Übernahme, Vollständigkeit, Archivierung

- **amtliche Publikationen**

elektronische Ausgabe ersetzt die gedruckte

Beispiel: [Straßenbaubericht](#) / RP Freiburg



Geschäftsgang, Präsentation, Archivierung

- **Pflichtexemplare** kommerzieller Verleger

Verpflichtung zur Ablieferung ist gegeben

Beispiel: [Springer](#) – bei DNB archiviert

➤ Urheberrecht, Benutzung, Archivierung

- **Hochschulschriften**

Eigenproduktion der Hochschule, Open Access, Zahl vs. Grösse

Beispiel: [OPUS](#)

➤ Einwerbung, Verbreitung, Archivierung

- **Digitalisate**

produziert nach eigenen Gegebenheiten
(Format, Struktur) / jetzt DFG: Praxisregeln

Beispiel: [Naunhofer Nachrichten](#)



Produktion, Präsentation, Archivierung

- **Bilder**

Eigenproduktion für viele Nutzungen

Beispiel: Staatsgalerie Stuttgart



Transfer, Verkauf, Dokumentation

- **Tondateien**

Bewahrung oraler Tradition

Beispiel: Deutsches Volksliedarchiv



Digitalisierung, Transfer, Sicherung

browserbasierte Anwendung

hierarchiefähig

Httptrack als Offline-Browser

Rechtsgrundlage: Archiv- und Pflichtexemplargesetz

LA sammelt 64 Webseiten von Landeseinrichtungen regelmäßig

WLB und BLB sammeln im Rahmen Landesbibliothek und Pflichtexemplar

Welche Vorgaben sind implizit?



BOA

BADISCHE LANDESBIBLIOTHEK

LANDESARCHIV Baden-Württemberg

Württembergische Landesbibliothek

Home

erweiterte Suche

Browsen

Editieren

FAQ

Baden-Württembergisches Online-Archiv (BOA / Landesarchiv)

Recherche-Ergebnis

Website: [Staatsgalerie Stuttgart](#)

Website: [Staatstheater Karlsruhe](#)

Website: [Ministerium für Kultus, Jugend und Sport](#)

Website: [Imagekampagne "Wir können alles. Außer Hochdeutsch."](#)

Website: [DoIT-Online - Das Infoportal des IT- und Medienstandorts Baden-Württemberg](#)

Website: [Landesamt für Verfassungsschutz](#)

Website: [Zentrum für Kunst und Medien](#)

Website: [Filmkommission Baden-Württemberg](#)

Website: [Polizei Baden-Württemberg - Internet-Info-Angebot Ihrer Polizei](#)

Website: [Staatsministerium Baden-Württemberg](#)

Website: [Innenministerium Baden-Württemberg](#)

Website: [Justizministerium Baden-Württemberg](#)

Website: [Der Landesbeauftragte für den Datenschutz Baden-Württemberg](#)

Website: [Staatstheater Stuttgart](#)

Website: [Staatliches Museum für Naturkunde Karlsruhe](#)

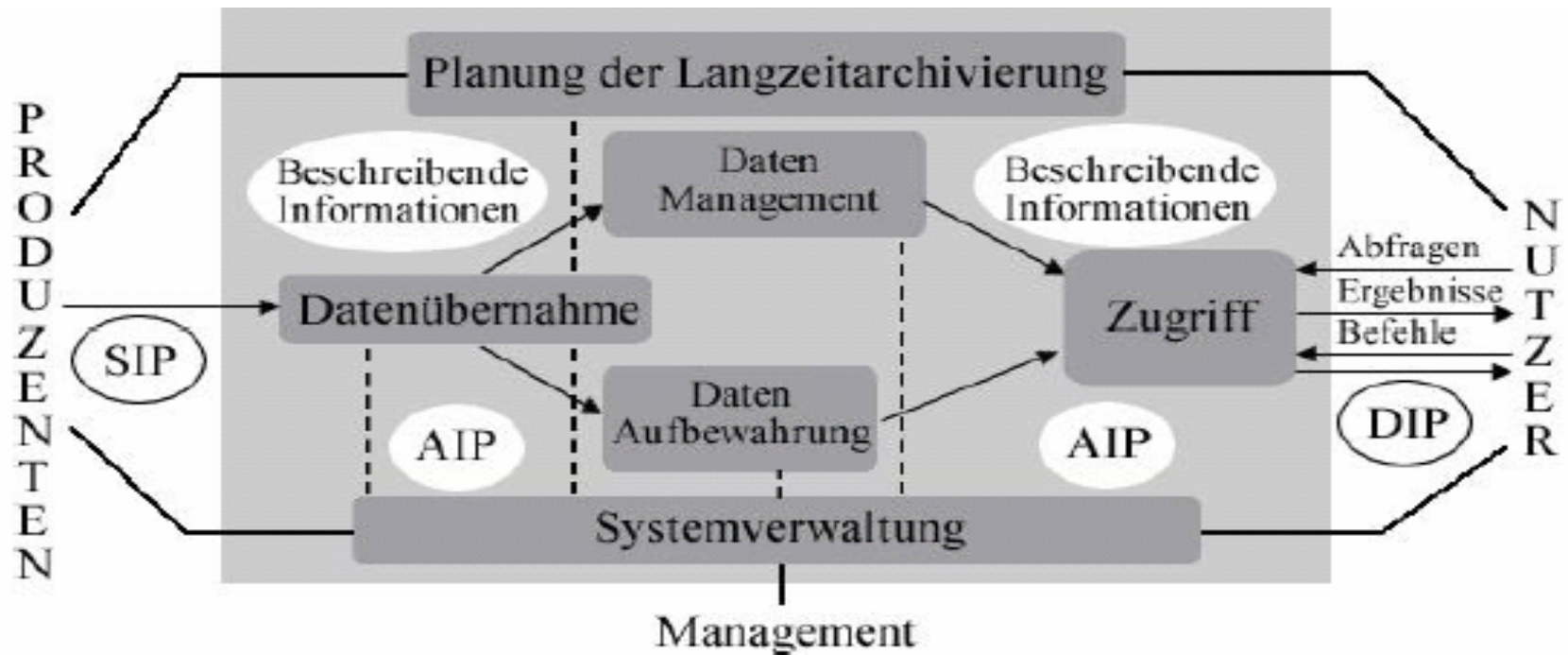
Website: [Staatliches Museum für Naturkunde Stuttgart](#)

Website: [NewCome.de - Existenzgründung und Unternehmensnachfolge in Baden-Württemberg](#)

Website: [Landesportal Baden-Württemberg](#)

- Das BSZ ist Dienstleister für Bibliotheken, Archive und Museen - Langzeitarchivierung ist eine Kernaufgabe
- Das BSZ entwickelt Software und Installationen nach gemeinsam definierten und priorisierten Anforderungen
- BSZ stellt die Installation zur Verfügung; Auswahl und Beschreibung der Dokumente sowie die Einhaltung der rechtlichen Schranken liegen bei den Kooperationspartnern

- Einhaltung von Standards und Schnittstellen
 - BSI, METS/LMER, METS/PREMIS, KOLIBRI...
- Mehrfachnutzung bzw. Automatische Erhebung von Metadaten
 - RAK/MAB in <http://www.boa-bw.de>, JHOVE
- Nutzung von Open Source Bausteinen
 - JHOVE, Httrack, MySQL...
- Präsentation: DFG-Viewer
- IT-Grundschatz nach BSI-Standard
- Bedarfsgerechter Ausbau von Soft- und Hardware
 - Leitungskapazität, Festplatten, Bandsicherung, dislozierte Speicherung



SIP Submission Information Package = die digitalen Ressourcen, die die aufbewahrenden Institutionen übernehmen

AIP Archival Information Package = vom Langzeitarchiv mit Metadaten ergänzte digitalen Medien. In dieser Form werden die digitalen Dokumente langfristig aufbewahrt

DIP Dissemination Information Package = in dieser Form werden die digitalen Medien je nach rechtlichen Bedürfnissen generiert und zur Verfügung gestellt

- Bsp.: Ingest (nach OAIS)
 - Können wir, weil wir die zukünftige Migration bestimmter Dateiformate heute noch nicht garantieren können, die Annahme dieser Dateiformate ausschließen (z.B. beim Harvesting von Websites)?
 - Fehler, die hier gemacht werden, werden teuer.
 - PDF → PDF/A, Videoformate...
- Verwendung des Offline-Browser Htrack
 - Einstellungen in Content-Management-Systemen
 - Redundanz der Speicherung in Zeitschnitten

- wir übernehmen, speichern und sichern, dokumentieren Dateiformate, vergeben Persistente Identifier, bilden Prüfsummen, halten die Struktur fest, versionieren Objekte, fügen Metainformation hinzu (SIP → AIP), prüfen die Rekonstitution aus dem Bandarchiv, bauen eine Exit-Strategie auf ...
 - auch dort, wo wir nicht auf eigenbestimmte Formate und Strukturen setzen können
 - Eine endgültige Lösung für ein „Dark Archive“ steht noch aus!
 - Letter of intent für KOPAL

einfache	↔	komplexe Strukturen
selbst bestimmte	↔	fremd bestimmte Formen
Vorhandensein	↔	Fehlen von Metadaten (zu Technik, Inhalt, Struktur)
Nutzung	↔	Archiv

- Einbettung in produzierende Geschäftsgänge
- Einbettung in bearbeitende Geschäftsgänge
- Einbettung in nutzende und präsentierende Geschäftsgänge

➤ Die Aufgabe ist komplex, die Mittel sind klein!

"Langzeit" bedeutet für die Bestandserhaltung digitaler Ressourcen nicht die Abgabe einer Garantieerklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, vom Informationsmarkt verursachten Wandel bewältigen können.

(Hans Liegmann)

Lösungen sind

- ➔ bedarfsgerecht
- ➔ paradigmatisch
- ➔ pragmatisch
- ➔ gemeinsam getragen
- ➔ nachvollziehbar

➔ VERANTWORTBAR

Bleiben Offene Fragen?

Wenn nicht jetzt – dann jederzeit gerne unter

stefan.wolf@bsz-bw.de
07531/883050



Danke für die
Aufmerksamkeit!