

Felix Geisler, Wiebke Dannehl, Christian Keitel und Stefan Wolf  
**Zum Stand der Webarchivierung in  
Baden-Württemberg**  
Web archiving – the present situation in  
Baden-Württemberg

DOI 10.1515/bd-2017-0051

**Zusammenfassung:** Der Beitrag beschreibt den derzeitigen Stand der Webarchivierung auf Landesebene in Baden-Württemberg. Die wesentlichen gesetzlichen Grundlagen für die Sammlung und Archivierung von Webseiten und zugehörigen Einzeldokumenten sind vorhanden. Zugleich ist die inhaltliche Selektion und technische Realisierung des Angebots eine umfangreiche Aufgabe, der sich mehrere Landeseinrichtungen gemeinsam stellen.

**Schlüsselwörter:** Webarchivierung, Elektronisches Pflichtexemplar, Baden-Württemberg, Digitales Kulturgut, Digitales Archiv, Forschungsdaten

**Abstract:** The article describes the present situation of web archiving at the level of the Land of Baden-Württemberg. The essential legal basis for collecting and archiving websites and related single documents exists. At the same time, the selection of contents and technical realization of the offer is an extensive task that is shared among several state institutions.

**Keywords:** web archiving, electronic deposit copy, Baden-Württemberg, digital cultural assets, digital archives, research data

---

**Dr. Felix Geisler:** geisler@blb-karlsruhe.de

**Wiebke Dannehl:** dannehl@wlb-stuttgart.de

**Prof. Dr. Christian Keitel:** christian.keitel@la-bw.de

**Stefan Wolf:** stefan.wolf@bsz-bw.de

# 1 Einleitung

Die Webarchivierung wird in Baden-Württemberg kooperativ von den Landesbibliotheken und dem Landesarchiv übernommen.<sup>1</sup> Basis sind für die Landesbibliotheken zum einen das Gesetz über die Ablieferung von Pflichtexemplaren an die Badische Landesbibliothek in Karlsruhe und die Württembergische Landesbibliothek in Stuttgart sowie die Anordnung der Landesregierung über die Abgabe amtlicher Veröffentlichungen an Bibliotheken.<sup>2</sup> Für das Landesarchiv greift das Landesarchivgesetz.<sup>3</sup> Die Definition der zu sammelnden Medien ist recht weit gefasst: „Digitale Publikationen sind Medienwerke in unkörperlicher Form, die in öffentlichen Netzen dargestellt werden.“ bzw. „Unterlagen im Sinne von Absatz 1 sind insbesondere Schriftstücke, Akten, Karteien, Karten, Pläne, Bild-, Film- und Tonmaterialien sowie sonstige Informationsträger und maschinenlesbar auf diesen gespeicherte Informationen und Programme.“

Die Landesbibliotheken sehen ihren Auftrag vor allem in der Sammlung von Netzpublikationen mit Druckschriftencharakter. Daneben sammeln die Landesbibliotheken einzelne Webseiten zu landeskundlichen Themen. Das Landesarchiv übernimmt die laufende Archivierung der Webseiten von Behörden, Gerichten und Firmen, die mehrheitlich dem Land gehören.

---

**1** Beinert, Tobias; Schoger, Astrid: Vernachlässigte Pflicht oder Sammlung aus Leidenschaft? Zum Stand der Webarchivierung in deutschen Bibliotheken. In: Zeitschrift für Bibliothekswesen und Bibliographie 62 (2015), 3-4, S. 172–183. Die gemeinsame Zuständigkeit von Bibliothek und Archiv für Behördenwebseiten kann auch international beobachtet werden. 2008/9 trennten sich die National Archives des Vereinigten Königreichs mit ihrem UK Government Web Archive vom UK Web Archive. Letzteres hat den Slogan „Preserving UK Websites“, ersteres den Wahlspruch „The National Archives is preserving UK government information published on the web“. <http://www.nationalarchives.gov.uk/webarchive/> und <https://www.webarchive.org.uk/ukwa/>. [Zugriff: 15.03.2017].

**2** Gesetz über die Ablieferung von Pflichtexemplaren an die Badische Landesbibliothek in Karlsruhe und die Württembergische Landesbibliothek in Stuttgart. URL <http://www.landesrecht-bw.de/jportal/?quelle=jlink&query=PfliExplAbtG+BW&psml=bsbawueprod.psml&max=true&aiz=true> [Zugriff: 08.03.2017]. Anordnung der Landesregierung über die Abgabe amtlicher Veröffentlichungen an Bibliotheken. URL <http://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/bestandsaufbau/amtsdruckschriften/abgabeerlasse/baden-wuerttemberg/> [Zugriff: 08.03.2017].

**3** Gesetz über die Pflege und Nutzung von Archivgut (Landesarchivgesetz – LArchG). <http://www.landesrecht-bw.de/jportal/?quelle=jlink&query=ArchivG+BW&psml=bsbawueprod.psml&max=true&aiz=true> [Zugriff: 08.03.2017].

## 2 Das Baden-Württembergische Online-Archiv

Die Technik für die Webarchivierung durch die Landesbibliotheken und das Landesarchiv wird vom Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) zur Verfügung gestellt.<sup>4</sup> Eingesetzt wird die Software SWBcontent, eine JAVA-basierte browsergestützte Workflow-Lösung auf Open-Source-Basis, die die für die Webseitenarchivierung spezialisierten Werkzeuge und Standards integriert. SWBcontent ist seit 2004 Basis des Baden-Württembergischen Online-Archivs (BOA), das von den Landesbibliotheken zusammen mit dem BSZ aufgebaut wurde. Das Landesarchiv kam 2005 als Partner hinzu und erhielt eine eigene Instanz innerhalb des Baden-Württembergischen Online Archivs (BOA).

Die Durchführung von Web-Downloads sowohl einzelner PDF-Dokumente als auch ganzer Webseiten erfolgte zunächst durch die Integration des Offline-Browsers HTTrack, seit 2012 ist Heritrix, ein Webcrawler, in BOA verfügbar. Downloads erfolgen im WARC-Format, das seit 2009 einen ISO-Standard darstellt (ISO 28500:9000). Die Präsentation der archivierten Webinhalte wird mit der vom Internet Archive im Open Source bereitgestellten Wayback Machine realisiert. Die Vergabe und Registrierung von URNs erfolgt automatisiert, ein Statistik-Tool ist angeschlossen. Eine freie und auch eine beschränkte Bereitstellung der Inhalte ist möglich, bevorzugt ist allerdings in jedem Fall die freie Bereitstellung im Web.

Seit dem 1. Januar 2017 ist das BSZ Mitglied im International Internet Preservation Consortium (IIPC). Durch diesen Schritt möchte es die Vernetzung mit anderen Akteuren auf dem Gebiet der Webarchivierung vorantreiben und leichteren Zugriff auf neueste Software-Tools erhalten.

## 3 Erschließung

Alle laufend in BOA archivierten Webseiten werden in die Zeitschriftendatenbank (ZDB) aufgenommen. Netzpublikationen mit Druckschriftencharakter werden je nach Erscheinungsweise in der ZDB oder dem Südwestdeutschen Bibliotheksverbund (SWB) nachgewiesen. Nach dem Anhängen von Exemplardaten gelangen die Einträge in die OPACs und Resource Discovery Systeme der beiden Landesbibliotheken. Publikationen mit Landesbezug werden zudem in der Landesbi-

---

<sup>4</sup> AG Langzeitarchivierung des Ministeriums für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg: Langzeitarchivierung digitaler Objekte in Baden-Württemberg: Ein Schichtenmodell der Kompetenzen, Funktionen, Dienstleistungen und Schnittstellen. URL <https://swop.bsz-bw.de/frontdoor/index/index/docId/478> [Zugriff: 08.03.207].

bliographie Baden-Württemberg Online inklusive des betreffenden Links zum archivierten Objekt aufgenommen. Einträge des Landesarchivs werden zusätzlich im Online-Findmittelsystem des Landesarchivs Baden-Württemberg (OLF21) beschrieben. In OLF21 werden Webseiten nicht in einem Sammelbestand, sondern zusammen mit den anderen Unterlagen der jeweiligen Behörde „provenienzgerecht“ nachgewiesen.

Eine grobe klassifikatorische Erschließung der archivierten Objekte erfolgt mit den DDC-Sachgruppen der deutschen Nationalbibliographien (DNB) und im Bereich der Landesbibliotheken auch mit der Systematik der Landesbibliographie Baden-Württemberg.

## 4 Webarchivierung bei den Landesbibliotheken

Die Auswahl der Inhalte für die Webarchivierung bei den Landesbibliotheken orientiert sich einerseits am Pflichtexemplargesetz und an der Anordnung der Landesregierung über die Abgabe amtlicher Veröffentlichungen an Bibliotheken (sogenannte *Amtsdruckschriften-Verordnung*)<sup>5</sup> und andererseits am landeskundlichen Inhalt der für eine Sammlung in Frage kommenden Webseiten und -dokumente. Bei Publikationen, die unter die *Amtsdruckschriften-Verordnung* oder das Pflichtexemplargesetz fallen, werden die Urheber über die Archivierung und Bereitstellung informiert. Gegebenenfalls wird die Erlaubnis zur uneingeschränkten Bereitstellung der Inhalte über das Internet eingeholt. Bei Inhalten, die nicht unter die *Amtsdruckschriften-Verordnung* oder das Pflichtexemplargesetz fallen, wird bei den Urhebern die Genehmigung zur Archivierung und Zugänglichmachung im Netz angefragt. Für die *Amtsdruckschriften* gilt zusätzlich: „Sofern die Veröffentlichung in elektronischer Form erscheint, erfolgt die Abgabe grundsätzlich nur in dieser Form.“ Dies erspart den Landesbibliotheken die Aufbewahrung der parallelen Printausgaben.

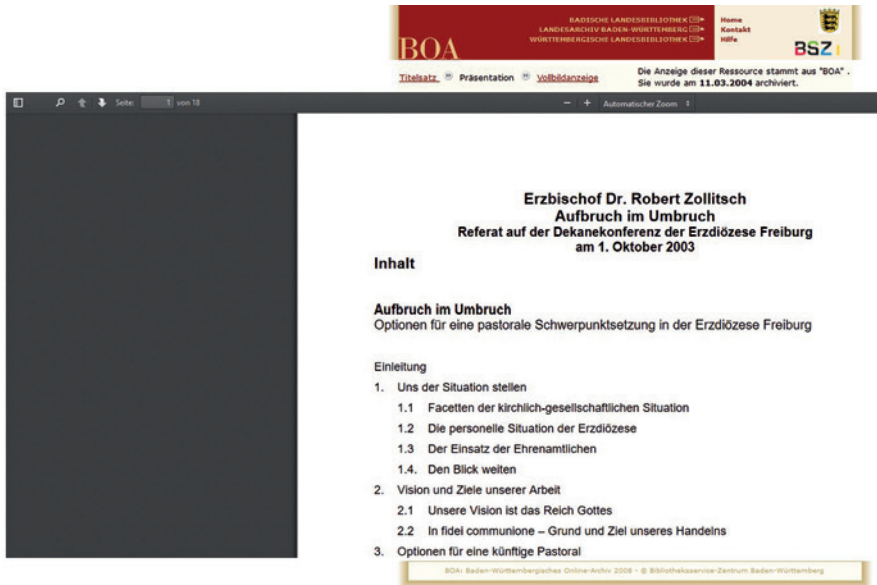
---

<sup>5</sup> Wie Anm. 4.



**Abb. 1:** Einzelseite aus dem Webauftritt der Arbeitsgemeinschaft Hauptbahnhof Stuttgart vom 06.10.2010, eingebracht in BOA durch die Württembergische Landesbibliothek Stuttgart.

Derzeit existieren im landesbibliothekarischen Teil des Baden-Württembergischen Online-Archiv knapp 10.000 Objekte, die mit insgesamt 3.391 Titelaufnahmen im SWB oder der ZDB verknüpft sind (Stand: 31. Dezember 2016). Darunter befindet sich eine Vielzahl an Broschüren, Infomaterial, Berichten und kurzen Abhandlungen, die in der Bibliothekssprache gerne unter der Bezeichnung „graue Literatur“ firmieren. Einzelne Webseiten werden nach landeskundlicher Bedeutung ausgewählt und abhängig vom Inhalt entweder einmalig, mehrmals oder in regelmäßigen Zeitschnitten gespeichert. Bei regelmäßiger Speicherung werden Zeitschnitte halbjährlich oder jährlich vorgenommen. Anlässe für das sogenannte „Event-Harvesting“ sind u. a. Ausstellungen in den Landeseinrichtungen oder politische Ereignisse (wie Wahlen, Bürgerentscheide und Großprojekte im Land).



**Abb. 2:** Einzeldokument aus dem Webauftritt des Erzbistums Freiburg vom 11.03.2004, eingebracht in BOA durch die Badische Landesbibliothek Karlsruhe.

## 5 Webarchivierung beim Landesarchiv

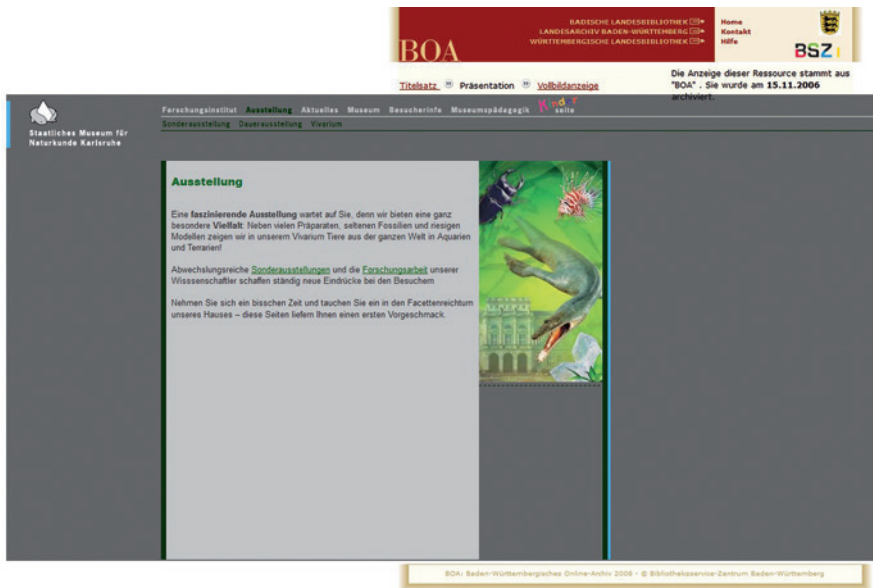
Bereits 2005 hat sich das Landesarchiv Baden-Württemberg dem bestehenden BOA-Verbund angeschlossen und 2006 mit der Archivierung von Webseiten der Landesbehörden begonnen.<sup>6</sup> Derzeit werden die Webseiten von 161 Behörden archiviert (Stand: 6. März 2017). Hinzu kommen 46 Webseiten, die von Behörden zur Flüchtlingshilfe und zum Umweltschutz, zu Kriminalität, Gesundheit oder anderen Themen bereitgestellt werden.

Nach dem Landesarchivgesetz sind Webseiten ebenso wie die anderen schriftlichen Erzeugnisse der Landesbehörden zunächst einmal Unterlagen. Und während früher alles in die Akten wanderte, fließen nun immer mehr Informationen in die Webseiten der Behörden. So werden Organigramme zwar weiterhin zur Auswertung der archivierten Akten benötigt, sie lassen sich aber dort nicht mehr

<sup>6</sup> Naumann, Kai: Gemeinsam Stark. Web-Archivierung in Baden-Württemberg, Deutschland und der Welt. In: *Archivar* 65 (2012), Nr. 01, S. 33–41.

auffinden – im Gegensatz zu den Webseiten dieser Behörden. Die Webseite bildet denn auch am ehesten das aktuelle Selbstverständnis der Behörden und deren Kommunikation mit den Bürgerinnen und Bürgern ab, wobei sich diese Kommunikation immer mehr bidirektional entwickelt. Informationen der Behörde stehen so einträchtig neben Texten und Eingaben von Bürgern und Unternehmen. Aufgrund dieser vielfachen Wechselbeziehungen wird seit 2005 die Bewertung und Archivierung von Webseiten nicht isoliert, sondern zusammen mit allen anderen Unterlagen in der Landesarchiv-internen AG Überlieferungsbildung diskutiert. Wenn die Archivierung einer Webseite beschlossen wurde, wird diese bis zu einer bestimmten Linktiefe vollständig übernommen. Isolierte PDF-Dateien werden dagegen vom Landesarchiv nicht in BOA archiviert.

Seit einigen Jahren archiviert das Landesarchiv in BOA auch einzelne private Blogs zur Ergänzung seiner sonstigen Überlieferung. Beispielsweise ergänzen die zu Stuttgart21 archivierten Blogs die staatliche Überlieferung zu dem Thema. In diesen Fällen erfolgt die Übernahme auf der Grundlage eines Archivierungsvertrags. Wiederum anders sieht es bei den Intranetseiten der Behörden aus. Sie wandern, Archivwürdigkeit vorausgesetzt, ins Digitale Magazin (DIMAG) des Landesarchivs und sind nicht im Internet zugänglich.



**Abb. 3:** Einzelseite aus dem Webauftritt des Staatlichen Museums für Naturkunde Karlsruhe vom 15.11.2006, eingebracht in BOA durch das Landesarchiv Baden-Württemberg.

## 6 Fazit

In Baden-Württemberg hat sich die Zusammenarbeit von vier Landeseinrichtungen auf dem Gebiet der Webarchivierung bewährt. Gemeinsam gelingt es, wesentliche Teile des digitalen Kulturerbes des Bundeslandes zu sichern und für die Nachwelt zu erhalten. Die parallele Sammlung von Webseiten und Einzeldokumenten und deren Erschließung im Kontext aller anderen Bestände von Bibliotheken und Archiven ermöglicht zugleich das leichte Auffinden der passenden Inhalte durch den Nutzer.



**Dr. Felix Geisler**  
Badische Landesbibliothek  
Erbprinzenstr. 15  
76133 Karlsruhe  
Deutschland  
E-Mail: [geisler@blb-karlsruhe.de](mailto:geisler@blb-karlsruhe.de)



**Wiebke Dannehl**  
Württembergische Landesbibliothek  
Konrad-Adenauer-Str. 8  
70173 Stuttgart  
Deutschland  
E-Mail: [dannehl@wlb-stuttgart.de](mailto:dannehl@wlb-stuttgart.de)



**Prof. Dr. Christian Keitel**  
Landesarchiv Baden-Württemberg  
Eugenstraße 7  
70182 Stuttgart  
Deutschland  
E-Mail: [christian.keitel@la-bw.de](mailto:christian.keitel@la-bw.de)





**Stefan Wolf**

Bibliotheksservice-Zentrum Baden-Württemberg

Felix-Wankel-Straße 4

78467 Konstanz

Deutschland

E-Mail: [stefan.wolf@bsz-bw.de](mailto:stefan.wolf@bsz-bw.de)