

Zum Stand der kooperativen Webarchivierung in Thüringen. Gemeinsames Sammeln von landeskundlich relevanten Websites der Thüringer Universitäts- und Landesbibliothek und der Deutschen Nationalbibliothek

Thomas Mutschler, Thüringer Universitäts- und Landesbibliothek, Jena

Zusammenfassung

Die Thüringer Universitäts- und Landesbibliothek Jena sammelt in Kooperation mit der Deutschen Nationalbibliothek seit dem Jahr 2018 landeskundlich relevante Websites. Der vorliegende Artikel zeichnet die Entwicklung der Webarchivierung an der Thüringer Universitäts- und Landesbibliothek Jena und den Weg zum Kooperationsprojekt mit der Deutschen Nationalbibliothek aus Thüringer Sicht nach. Es wird ein Überblick geboten über das Sammelprofil des Webarchivs Thüringen, die rechtlichen Grundlagen, den Workflow des Einsammelns der Websites sowie deren Erschließung und Präsentation. Ein Fazit und ein Ausblick auf künftige Herausforderungen runden den Beitrag ab.

Abstract

Since the year 2018 the Thuringian University and State Library Jena has been cooperating with the German National Library in the context of collecting and archiving websites with a regional focus on the federal state of Thuringia. From the point of view of the Thuringian University and State Library the article describes the evolution of the „Webarchiv Thüringen“. It provides an overview of the collection profile, the legal situation, the workflow and the indexing and presentation of the collected websites. Finally, a conclusion is provided and challenges for the future are discussed.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/5632>

Autorenidentifikation: Mutschler, Thomas: GND: 1203663153

ORCID: <https://orcid.org/0000-0001-6016-8700>

Schlagwörter: Webarchivierung, Thüringer Universitäts- und Landesbibliothek, Deutsche Nationalbibliothek

Dieses Werk steht unter der Lizenz [Creative Commons Namensnennung 4.0 International](#)

1. Webarchivierung als Chance für neue Kooperationen

Aufgrund der immensen Bedeutung des Web sowohl für das öffentlich-staatliche als auch für das private Leben steht die Thüringer Universitäts- und Landesbibliothek Jena (ThULB Jena) seit einiger Zeit vor der Aufgabe, landeskundlich einschlägige Netzressourcen aus und über Thüringen zu sammeln und zu archivieren, um sie als digitale Überlieferung für künftige Forschungen verfügbar zu machen. Während in Thüringen für die Langzeitarchivierung der druckbildähnlichen digitalen Pflichtexemplare (z. B. PDFs) und der im Digitalisierungszentrum der Thüringer Universitäts- und Landesbibliothek

produzierten Digitalisate eigene Repositorien zur Verfügung stehen,¹ kooperiert die Bibliothek im Bereich der Webarchivierung seit 2018 mit der Deutschen Nationalbibliothek (DNB).²

Die Deutsche Nationalbibliothek verfügt seit einiger Zeit über eine umfassende Infrastruktur für das Einsammeln, Verzeichnen und Archivieren von Websites. Gemäß der Maxime von Elisabeth Niggemann, dass Webarchivierung Kooperation und Arbeitsteilung braucht,³ sind seit 2018 auch Regional- und Landesbibliotheken, welche über die gesetzliche Basis, aber kein eigenes Webarchiv verfügen, eingeladen, mit der Deutschen Nationalbibliothek zu kooperieren. Dieses neue Angebot weckte das Interesse der Thüringer Universitäts- und Landesbibliothek als dem deutschlandweit ersten Projektpartner der Deutschen Nationalbibliothek im Bereich der Webarchivierung. Das Ziel der Kooperation besteht im Aufbau einer landeskundlich profilierten Sammlung an archivierten Websites aus und über Thüringen, die über das Portal der Deutschen Nationalbibliothek zugänglich sind.

2. Begriffsbestimmung

Mit „Webarchivierung“ ist im Folgenden das automatisierte Einsammeln von Websites, deren Speicherung in einem digitalen Archiv sowie die Erschließung und Bereitstellung der gespeicherten Kopien – mit dem Ziel, sie als digitale Überlieferung dauerhaft verfügbar zu machen – gemeint.⁴ Websites werden von Bibliotheken oder Archiven entweder auf der Ebene der Top Level Domain eines ganzen Landes (z. B. *.de) archiviert oder, wie im Fall des Webarchivs Thüringen, auf der Grundlage definierter Archivierungsprofile selektiv gesammelt. Websites werden in Deutschland nicht allein durch die Deutsche Nationalbibliothek eingesammelt und langzeitarchiviert, sondern darüber hinaus in einer

- 1 In selbstständiger Form publizierte Webveröffentlichungen (z. B. E-Books) werden von der Thüringer Universitäts- und Landesbibliothek im Rahmen der Digitalen Bibliothek Thüringen erfasst, gespeichert und präsentiert. Die Digitale Bibliothek Thüringen ist unter der folgenden URL online zugänglich: <www.db-thueringen.de>, Stand: 10.10.2020. Für die Erfassung, Speicherung und Präsentation sowohl digitalisierter als auch ausschließlich in digitaler Form publizierter Zeitschriften und Zeitungen steht an der Thüringer Universitäts- und Landesbibliothek das Zeitschriftenportal „Journals@UrMEL“ zur Verfügung. Das Zeitschriftenportal „Journals@UrMEL“ ist online unter der folgenden URL zugänglich: <<https://zs.thulb.uni-jena.de/content/below/index.xml>>, Stand: 10.10.2020. Diese beiden Repositorien enthalten auch Pflichtexemplare aus dem Zuständigkeitsbereich der Thüringer Universitäts- und Landesbibliothek. Für die Erschließung, Speicherung, Aufbereitung und Präsentation digitalisierter Kulturüberlieferung stehen an der Thüringer Universitäts- und Landesbibliothek Thüringen eine Reihe weiterer Repositorien bereit; vgl. dazu auch Mutschler, Thomas: Neue Wege der Kulturgutdigitalisierung in Thüringen, in: Bibliotheksdienst 51, 2017, H. 3–4, S. 310–321. Die an der Thüringer Universitäts- und Landesbibliothek betriebenen Repositorien basieren auf der Open Source Software MyCoRe.
- 2 Der Autor bedankt sich bei Susanne Puls und Britta Woldering, beide von der Deutschen Nationalbibliothek, für die Unterstützung und die konstruktiven Hinweise bei der Abfassung des Beitrags.
- 3 Niggemann, Elisabeth: Im weiten endlosen Meer des World Wide Web. Vom Sammelauftrag der Gedächtnisinstitutionen, in: Zeitschrift für Bibliothekswesen und Bibliographie 62, 2015, H. 3–4, S. 157. Online: <<http://dx.doi.org/10.3196/1864295015623439>>.
- 4 Einen Überblick über den aktuellen Stand der deutschsprachigen Literatur zum Thema Webarchivierung bietet das Themenheft zur Webarchivierung in Bibliotheken der Zeitschrift für Bibliothekswesen und Bibliographie 62, 2015, H. 3–4, sowie die dortigen Einzelbeiträge, die unterschiedliche auch internationale Aspekte des Themas beleuchten. Beiträge zu Aktivitäten einzelner deutscher Bibliotheken im Bereich der Webarchivierung bietet außerdem Bibliotheksdienst 51, 2017, H. 6. Zum Themenkomplex insgesamt vgl. auch Niu, Jinfang: An Overview of Web Archiving, in: D-Lib Magazin 18, 2012, H. 3–4. Online: <<https://doi.org/10.1045/march2012-niu1>>.

Reihe von Bundesländern durch die jeweils zuständigen Landesbibliotheken in eigenen Webarchiven verfügbar gemacht.⁵

Der Terminus „Website“ bezieht im Folgenden publizierte Webauftritte bzw. Webpräsenzen im Rahmen des Web ein. Websites sind dadurch gekennzeichnet, dass sie hypertextuell gestaltet sowie multimedial und interaktiv aufbereitet sind und in der Regel aus mehreren Unterseiten bestehen, die einen inhaltlichen Kontext bilden; sie werden über einen Uniform Resource Locator (URL) mit Hilfe eines Webbrowsers abgerufen. Archiviert wird die Website durch das Kopieren einer Domain zu einem bestimmten Zeitpunkt. In diesem Zusammenhang werden Snapshots bzw. Zeitschnitte von Websites angefertigt, welche den jeweiligen Stand einer Website zu einem definierten Zeitpunkt festhalten.

3. Der mühsame Weg zum Webarchiv Thüringen

Die Thüringer Universitäts- und Landesbibliothek hat sich erstmalig im Jahr 2013 im Rahmen einer umfassenden Testphase intensiver mit der Problematik der Websitearchivierung beschäftigt und eine entsprechende Anwendung für eine Testinstanz aufgesetzt.⁶ Für die Durchführung des Experiments bot sich die zentrale Webpräsenz des Freistaats Thüringen (Domain *.thüringen.de) an - schon zum damaligen Zeitpunkt ein Webangebot mit einem Volumen von rund 21 Gigabyte und nicht weniger als 133.000 einzelnen URLs, die es bei der Spiegelung zu berücksichtigen galt. Infolge ihres großen Umfangs nahm die Spiegelung mehrere Tage in Anspruch und ließ das Projektteam so erahnen, welche Ressourcen bei einer umfassenderen Sammeltätigkeit vorzusehen sind.

Welche Ressourcen wurden im Rahmen des an der Thüringer Universitäts- und Landesbibliothek durchgeführten Testlaufs eingesammelt? Der Abzug der Webpräsenz des Landes Thüringen enthielt rund 41 verschiedene Dateitypen; dabei machten die in die Website eingebetteten, druckbildähnlichen Online-Publikationen, insbesondere PDF-Dateien, welche von den Ministerien und Landesbehörden über deren Websites veröffentlicht wurden, den größten Teil des eingesammelten Webcontents aus. Derartige Quellen umfassten etwas mehr als die Hälfte des Gesamtvolumens der Spiegelung - gefolgt von Bildmaterialien und Audio/Video-Dateien. Zur Durchführung des Testlaufs wurden die Software *Heritrix* zum Einsammeln der Websites sowie für die Präsentation der archivierten Zeitschnitte die *Wayback Machine* genutzt, welche an der Thüringer Universitäts- und Landesbibliothek für die Zeit der Testphase zuvor auf einem Testserver installiert worden waren.

Im Ergebnis vermittelte der Testlauf einen sehr guten Eindruck davon, welche Herausforderungen die Webarchivierung mit sich bringt, sowohl in technischer als auch in organisatorischer Hinsicht. Der Testlauf zeigte außerdem auf, dass bestimmte Bereiche bzw. Unterseiten des Portals nicht

5 Für einen Überblick über die Aktivitäten der Bundesländer im Bereich der Webarchivierung und zum Ansatz in Nordrhein-Westfalen vgl. Ammendola, Andrea Pietro: Webarchivierung in NRW aus Sicht der Universitäts- und Landesbibliothek Münster, Köln 2020, S. 82. Online: <<https://nbn-resolving.org/urn:nbn:de:hbz:79pbc-opus-15775>>.

6 Vgl. Mutschler, Thomas: Webarchivierung in Thüringen. Sachstand, Planungen, Perspektiven, Präsentation gehalten anlässlich der UAG „Pflichtexemplar“ am 04.04.2019 an der Saarländischen Universitäts- und Landesbibliothek Saarbrücken, Jena 2019. Online: <<https://doi.org/10.22032/dbt.40535>>.

eingesammelt werden konnten, entweder weil sie zugangsbeschränkt waren, eigenständige Datenbankapplikationen darstellten, für welche separate technische Lösungen zu schaffen sind, oder Inhalte konnten aus anderen technischen Gründen nicht eingesammelt werden. Dieser Befund war insofern wertvoll, als das Verständnis für die Komplexität der technischen und organisatorischen Prozesse auch für das fünf Jahre später gestartete Kooperationsprojekt mit der Deutschen Nationalbibliothek erkenntnisleitend waren.

Nach dem Testlauf von 2013 und der Auswertung der Testergebnisse wurde es an der Thüringer Universitäts- und Landesbibliothek in Sachen Webarchivierung keineswegs ruhiger, auch wenn der Aufbau des kooperativen Webarchivs noch fünf Jahre auf sich warten lassen sollte. So war in der Zwischenzeit auch das Thüringische Hauptstaatsarchiv Weimar auf das Thema Webarchivierung aufmerksam geworden und hat im Zuge des Aufbaus eines digitalen Magazins eine Wirtschaftlichkeitsbetrachtung in Auftrag gegeben, in welche auch die bis dato an der Thüringer Universitäts- und Landesbibliothek gesammelten Erfahrungen eingeflossen sind.

Ein weiterer positiver Nebeneffekt, welcher mit der Wirtschaftlichkeitsbetrachtung einherging, war die Erarbeitung eines Sammelprofils für Thüringer Websites durch die Thüringer Universitäts- und Landesbibliothek. Ein Sammelprofil zu definieren, war allein schon aus dem Grund notwendig, um die Menge bzw. den Umfang eines künftigen Webarchivs abschätzen zu können. Das für den Testlauf erstellte Sammelprofil bot der Bibliothek eine gute Basis für die spätere Kooperation mit der Deutschen Nationalbibliothek. Bevor die Zusammenarbeit im Jahr 2018 konkrete Züge annahm, wurden an der Thüringer Universitäts- und Landesbibliothek zuvor noch weitere Optionen geprüft, so z.B. die Nutzung der vom *Internet Archive* bereitgestellten Archivierungs- und Präsentationssoftware *Archive-It*; nach negativer Evaluierung wurde diese Option jedoch nicht weiterverfolgt.⁷

4. Prinzip der „exemplarischen Vielfalt“ als Sammlungskriterium

Reinhard Rinn hat als Vertreter der Deutschen Bibliothek bereits 1998 im Rahmen des Bibliothekartags darauf hingewiesen, dass die Vermischung der diversen Publikationsarten im Zusammenhang mit den damals noch jungen Netzpublikationen für Bibliotheken zunehmend ein Problem darstelle und Sammelprofile entsprechend deutlicher zu formulieren seien: „Die bisher angewendeten Selektions- und Sammlungskriterien, die sich hauptsächlich auf gedruckte Publikationen beziehen, müssen überprüft und den neuen Gepflogenheiten angepasst werden“.⁸ Angesichts des rasanten Wachstums des Web in den vergangenen 20 Jahren hat diese Aussage auch und gerade in Bezug auf das Webarchiv Thüringen nichts an ihrer Aktualität verloren – das Gegenteil trifft zu.

7 Das negative Ergebnis der hausinternen Evaluierung bezüglich Archive-It ist vor allem darauf zurückzuführen, dass Archive-It für die Thüringer Universitäts- und Landesbibliothek nicht im Rahmen eines Konsortiums oder über einen Verbund lizenzierbar ist, sondern bilateral bezogen werden müsste.

8 Rinn, Reinhard: Metadaten und nationalbibliographische Verzeichnung von Netzpublikationen, in: Wefers, Sabine (Hg.): „Nur was sich ändert, bleibt“. 88. Deutscher Bibliothekartag in Frankfurt am Main 1998, Frankfurt am Main 1999 (Zeitschrift für Bibliothekswesen und Bibliographie, Sonderheft 75), S. 223.

Dass die Orientierung an den landesbibliothekarischen Sammelrichtlinien für gedruckte Publikationen im Bereich der Netzpublikationen und Websites im Sinne der Einschätzung von Reinhard Rinn nur sehr bedingt möglich ist, wurde auch an der Thüringer Universitäts- und Landesbibliothek bereits zu einem frühen Zeitpunkt deutlich. Insbesondere der in der Printwelt angestrebte Anspruch einer größtmöglichen Vollständigkeit der Sammlung ist für Websites völlig neu zu definieren, zumal das Webarchiv Thüringen aufgrund des selektiven Ansatzes notwendigerweise eine Auswahl von Quellen aus dem Web darstellt. Gefragt war demnach die Definition eines mehr oder minder großen und qualitativ bewerteten Ausschnittes aus der Vielfalt der Angebote, wofür sich der von Christoph Classen aus kulturwissenschaftlicher Sicht geprägte Begriff der „exemplarischen Vielfalt“ anbot.⁹ Im Rahmen eines 2018 in der Deutschen Nationalbibliothek stattgefundenen Workshops zum Thema „Webarchivierung aus kulturwissenschaftlicher Perspektive“ empfahl der Zeithistoriker Classen, Sammlungen von Websites pluraler und flexibler sowie stärker fallbezogen anzulegen, etwa zu bestimmten zeittypischen Phänomenen oder sozialen Gruppen.¹⁰ Darüber hinaus ist im Zusammenhang mit der Ausarbeitung eines Sammelprofils auf die Arbeiten der AG Regionalbibliotheken zu verweisen, in welcher auch die Thüringer Universitäts- und Landesbibliothek als Mitglied vertreten ist und welche sich seit längerer Zeit mit dem Thema „Sammelrichtlinien“ für Websites konstruktiv auseinandersetzt.¹¹

An diese Überlegungen anknüpfend, orientiert sich das digitale Archiv Thüringer Websites grundsätzlich am Sammelauftrag der Thüringer Universitäts- und Landesbibliothek als regionaler Pflicht-exemplarbibliothek für den Freistaat Thüringen. Gesammelt werden demzufolge Websites, die über den öffentlich zugänglichen Teil des Web publiziert und ohne Kosten frei zugänglich gemacht werden, die sich auf das Bundesland beziehen und Themen von historischer, sozialer, politischer, kultureller, religiöser, wissenschaftlicher und ökonomischer Bedeutung enthalten. Gesammelt werden Websites von wissenschaftlichen, bildungsmäßigen und kulturellen Einrichtungen, von Parteien, Religionsgemeinschaften, Stiftungen, politischen Institutionen, Vereinen, Verbänden, Akademien, Gesellschaften, Kommunen, Gemeinden und Landkreisen sowie von Unternehmen. Websites von Ministerien, Behörden und Organisationen des Freistaats stehen ebenfalls im Fokus der Sammlung. Das Sammelprofil ist grundsätzlich dynamisch angelegt und wird durch das Hinzufügen neuer Websites und vor allem durch das Anlegen ereignisspezifischer Sammlungen kontinuierlich weiterentwickelt bzw. aktualisiert.

Neben einer regulären Kollektion können im Rahmen des Webarchivs Thüringen themenspezifische oder ereignisbezogene Sammlungen angelegt werden. Jedes Ereignis bildet eine eigene, in sich abgeschlossene Sammlung. Spezialsammlungen wurden bislang zu den Themen „Europa- und Landtagswahl 2019“ und den Auswirkungen der „Corona-Pandemie“ in Thüringen angelegt. Diese Spezialsammlungen ergänzen die Sammlung von landeskundlich relevanten Websites. Die Auswahl der einzusammelnden Websites wird von Fachreferentinnen und Fachreferenten der Thüringer Universitäts- und Landesbibliothek und der Redaktion der Landesbibliographie anhand sachlicher, regionaler und zeitlicher Kriterien unterstützt.

9 Zitat bei Müller, Sophie: Webarchivierung aus kulturwissenschaftlicher Perspektive, in: Dialog mit Bibliotheken 30, 2018, H. 2, S. 61. Online: <<https://nbn-resolving.org/urn:nbn:de:101-2018091409>>.

10 Vgl. ebenda.

11 Vgl. Pfau, Pia und Wiesenmüller, Heidrun: Die Sammlung von amtlichen Veröffentlichungen in elektronischer Form an regionalen Pflichtexemplarbibliotheken, in: o-bib 7, 2020, H. 1, S. 7ff. Online: <<https://doi.org/10.5282/o-bib/5583>>.

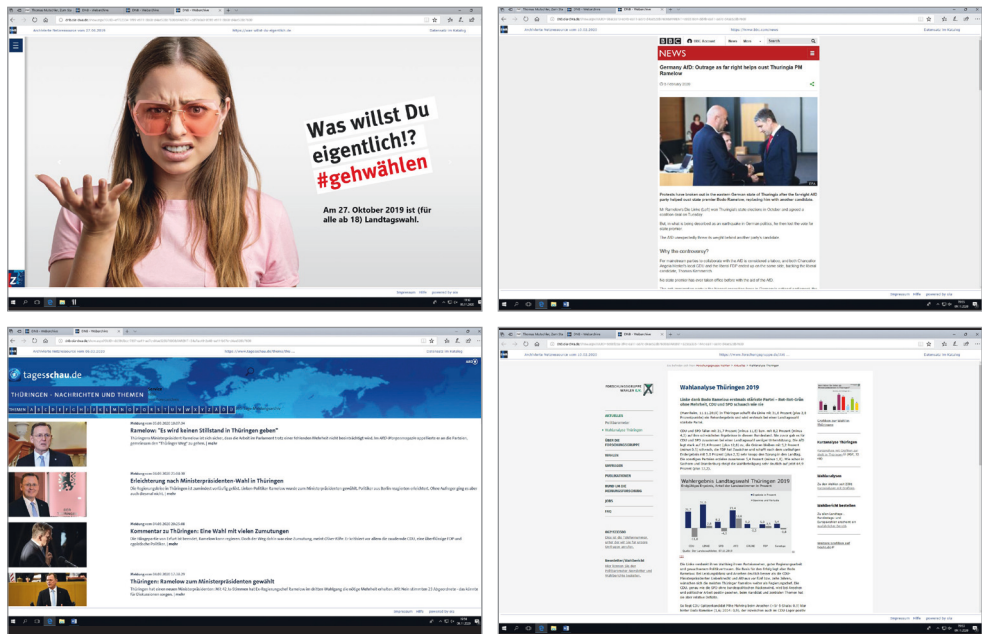


Abb.1: Beispiele von archivierten Websites aus der Kollektion „Wahlen 2019“

5. Rechtliche Situation in Thüringen

Das Thema „Webarchivierung“ unterliegt jedoch nicht allein technischen und sammlungsspezifischen Prämissen, sondern hängt ganz wesentlich auch von rechtlichen Parametern ab. Der überwiegende Teil der digitalen Inhalte im Web genießt nämlich urheberrechtlichen Schutz. Die Anfertigung einer Kopie durch das Herunterladen der Dateien, ihre dauerhafte Speicherung und öffentliche Bereitstellung bedürfen demzufolge einer gesetzlichen Ermächtigungsgrundlage. Den rechtlichen Rahmen dafür bilden im Freistaat Thüringen derzeit drei Bestimmungen. Für den Bereich der digitalen Publikationen (Elektronisches Pflichtexemplar) ist zunächst das Thüringer Pressegesetz einschlägig. Dieses Gesetz sieht neben der Abgabe von gedruckten Publikationen auch die Ablieferung von Medien, die in unkörperlicher Form verbreitet werden, an die Thüringer Universitäts- und Landesbibliothek vor.¹² Den Modus der Ablieferung der digitalen Publikationen regelt eine Verordnung, welche auch die Möglichkeit des automatisierten Abholens (Harvesting) ausdrücklich einbezieht.¹³

12 Vgl. §§ 6 und 12 Thüringer Pressegesetz (TPG) vom 31.07.1991 in der Fassung von 2008. Online: <<http://landesrecht.thueringen.de/jportal/?quelle=jlink&query=PresseG+TH+%C2%A7+12&psml=bsthueprod.psm1&max=true>>, Stand: 10.10.2020.

13 Vgl. Verordnung über die Ablieferung digitaler Publikationen an die Thüringer Universitäts- und Landesbibliothek vom 08.02.2011 in der Fassung vom 20.11.2015. Online: <<http://www.landesrecht.thueringen.de/jportal/?quelle=jlink&query=DigPublAbIV+TH&psml=bsthueprod.psm1&max=true>>, Stand: 10.10.2020.

Darüber hinaus ist im Kontext der Webarchivierung für Thüringen ein Erlass über die Abgabe amtlicher Veröffentlichungen einschlägig.¹⁴ Dieser verpflichtet die Behörden, Dienststellen und Einrichtungen des Landes alle durch sie herausgegebenen oder in ihrem Auftrag erscheinenden Veröffentlichungen u. a. an die Thüringer Universitäts- und Landesbibliothek abzugeben und bezieht auch elektronische Publikationen ein. Der Bibliothek wird mit der Abgabe der digitalen Veröffentlichung das Recht eingeräumt wird, die Daten zu speichern, zu vervielfältigen, zu verändern und sie öffentlich zugänglich zu machen, ohne dass es der vorherigen Einholung einer zusätzlichen Einverständniserklärung bedarf. Für in Thüringen publizierte Amtsdruckschriften gilt es noch festzuhalten, dass diese bereits seit vielen Jahren durch die Thüringer Universitäts- und Landesbibliothek Thüringen auf der Ebene der einzelnen Titel und Ausgaben im Rahmen des elektronischen Pflichtexemplars gesammelt, erfasst und über Repositorien zugänglich gemacht werden.¹⁵

Die Einholung einer vorherigen Genehmigung durch den Rechteinhaber der Website für die Speicherung und für die Publikation der gespiegelten Version würde allerdings in solchen Fällen notwendig, wenn es sich um Websites handelt, die nicht von Einrichtungen des Freistaats herausgegeben wurden. Ein solches Verfahren zur Einholung von Genehmigungen durch den Rechteinhaber für das Vervielfältigen, Archivieren und Präsentieren der eingesammelten Websites wäre im Workflow „Webarchivierung“ in jedem Fall dann zu berücksichtigen, sobald eine Einrichtung das Webarchiv selber hostet. Da die Thüringer Universitäts- und Landesbibliothek die archivierten Objekte, zumindest momentan, nicht selbst speichert, sondern die Speicherung der Websites durch die Deutsche Nationalbibliothek auf der Grundlage ihres gesetzlichen Sammelauftrags erfolgt, entfällt dieser Schritt zumindest aus Thüringer Sicht. Das Urheberrechtsgesetz ermöglicht den Zugriff auf das Webarchiv auch in den Räumlichkeiten von regionalen Pflichtexemplarbibliotheken derjenigen Bundesländer, die über ein entsprechendes Gesetz verfügen.

6. Geschäftsgang für das Webarchiv Thüringen

Der Geschäftsgang für das Webarchiv Thüringen umfasst diverse Arbeitsschritte und beginnt mit der Selektion der einzusammelnden Websites. Deren Auswahl und inhaltliche Evaluierung wird von Expertinnen und Experten der Thüringer Universitäts- und Landesbibliothek vorgenommen. Websites mit landeskundlicher Relevanz werden in Orientierung am landesbibliothekarischen Sammelprofil durch gezieltes Recherchieren im Web identifiziert. Die so getroffene Auswahl an URLs wird anschließend der Deutschen Nationalbibliothek übergeben und von den dortigen Bibliothekarinnen und Bibliothekaren in die Client-Software des Anbieters „oia GmbH“ mit Sitz in Düsseldorf eingepflegt. Dabei wird die Sammlungsfrequenz so parametrisiert, dass von den Websites standardmäßig zweimal jährlich Snapshots angefertigt werden. Im Rahmen ereignisbezogener Crawls erfolgen die Spiegelungen nach Absprache situationsbedingt.

14 Vgl. Abgabe amtlicher Veröffentlichungen an Bibliotheken und das Hauptstaatsarchiv vom 19.11.2008. Online: <<https://staatsbibliothek-berlin.de/de/die-staatsbibliothek/abteilungen/bestandsaufbau/amtsdruckschriften/abgabeerlasse/thueringen/>>, Stand: 10.10.2020.

15 Zum Pflichtexemplar in Thüringen vgl. Mutschler, Thomas: Thüringen, in: Bibliotheksdienst 47, 2013, H. 8–9, S. 659–663. Online: <<https://doi.org/10.1515/bd-2013-0078>>, Stand: 10.10.2020.

Die zum Einsatz kommende Crawler-Software des Anbieters verfolgt die Links innerhalb einer Website; dahinterliegende Dateien werden dabei, soweit technisch möglich, ebenfalls bis zu einer bestimmten Tiefe heruntergeladen. Per Default werden bei einem Crawl alle Links innerhalb der gleichen Domain verfolgt. Darüber hinaus kann konfiguriert werden, nach welcher Zeit der Crawl vom System spätestens beendet werden soll. Es kann jedoch auch vorkommen, dass Inhalte aus technischen Gründen teilweise nicht kopiert werden können. Für solche Fälle muss eine Qualitätskontrolle vorgesehen werden. Eine grundlegende Qualitätssicherung aller durchgeführten Crawls ist im Leistungsumfang des Anbieters der Deutschen Nationalbibliothek enthalten. Allerdings umfasst diese keine vollständige Überprüfung aller heruntergeladenen Inhalte. Aus diesem Grund erfolgt durch die Thüringer Universitäts- und Landesbibliothek derzeit eine stichprobenartige Qualitätskontrolle auf intellektueller Basis. Derzeit wird darauf geachtet, dass die zentralen Inhalte einschließlich visueller Elemente möglichst vollständig heruntergeladen werden und das *look and feel* der Website erhalten bleiben. Erfüllen die Archivkopien die Mindestanforderungen nicht, bittet die Deutsche Nationalbibliothek den Anbieter um Prüfung. Lassen sich technische Probleme nicht klären, kann in einzelnen Fällen von der Archivierung abgesehen werden. Da die Qualitätskontrolle erhebliche Ressourcen bindet, ist zu hoffen, dass die automatisierten Verfahren der Qualitätsprüfung im Bereich der Webarchivierung weiterentwickelt werden. Der Wunsch nach mehr Automatisierung betrifft jedoch nicht nur das Qualitätsmanagement, sondern darf auch auf die Arbeitsschritte der inhaltlichen Selektion der einzusammelnden Websites sowie deren Erschließung übertragen werden. Hier wird die Entwicklung von KI-Verfahren mit Spannung verfolgt.

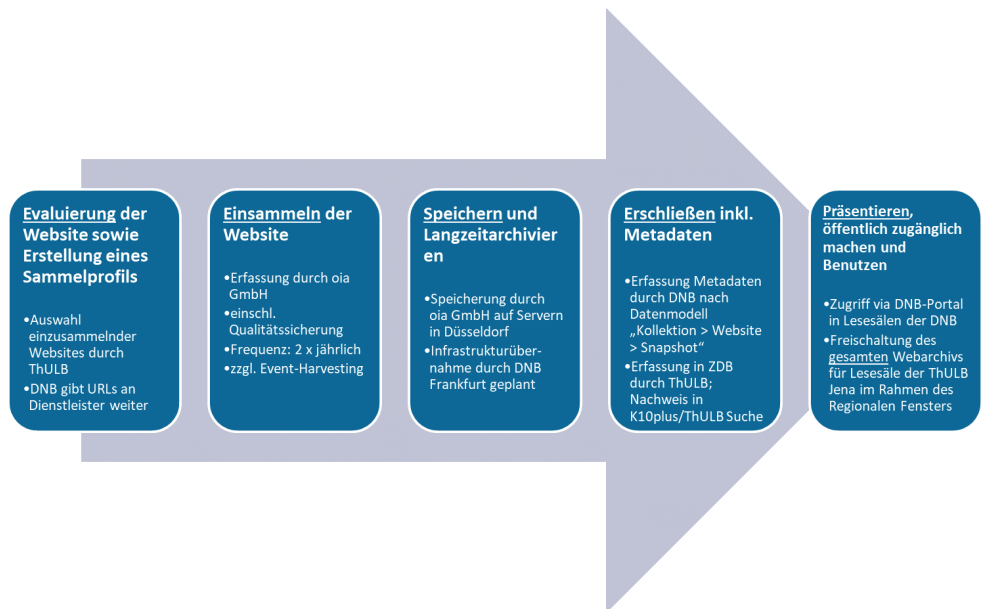


Abb. 2: Geschäftsgang für das Webarchiv Thüringen

Archiviert werden die Website-Kopien durch die Deutsche Nationalbibliothek im WARC-Format (= Web ARChive file format; ISO 28500) als standardisiertem Format für die Archivierung von Websites. WARC ist in der Lage, die verschiedenen heterogenen Teile, aus denen Websites zumeist zusammengesetzt sind, in einem Container zusammenzufassen, so dass die archivierte Website auch später noch als solche erfahrbar ist.

Im Ergebnis der nunmehr zweijährigen Pilotphase des Projekts werden im Rahmen des Webarchivs Thüringen ca. 380 Websites regelmäßig zwei Mal im Jahr eingesammelt und archiviert. Insgesamt wurden bisher rund 1.350 Zeitschnitte eingesammelt (Stand Juli 2020). Ergänzt wird dieser Bestand durch derzeit 80 URLs, die im Rahmen von 2 Ereigniscrawls gesammelt wurden. Die Sammlung soll kontinuierlich ausgebaut werden.

7. Verzeichnung, Präsentation und Zugriff auf das Webarchiv Thüringen

Abschließend geht es um die Frage, wie die archivierten Websites bibliographisch verzeichnet und wie sie zugänglich gemacht bzw. präsentiert werden. Im Unterschied zur Deutschen Nationalbibliothek, welche die Titelinformationen für die im nationalen Ausschnitt des Webarchivs enthaltenen Websites in ihrem Katalog erfasst, werden die Thüringer Websites von der Thüringer Universitäts- und Landesbibliothek zusätzlich in der Zeitschriftendatenbank (ZDB) erfasst. Die Titeldaten und Bestandsinformationen zu den Thüringer Websites fließen über den Datenlieferdienst der Zeitschriftendatenbank anschließend in den Katalog der Thüringer Universitäts- und Landesbibliothek bzw. in deren Discovery-System ein, von wo aus Nutzerinnen und Nutzer Zugang zu der Sammlung erhalten. Die landesbibliographische Verzeichnung der archivierten Websites innerhalb der Thüringen-Bibliographie ist zukünftig ebenfalls vorgesehen.

Im Einzelnen sind nur wenige Schritte notwendig, um zur gewünschten Website-Kopie zu gelangen: Nach dem Aufrufen des entsprechenden Datensatzes im Suchportal der Thüringer Universitäts- und Landesbibliothek erscheint der dazugehörige Link zum Katalog der Deutschen Nationalbibliothek, in welchem die entsprechenden Zeitschnitte präsentiert werden. Vom Katalogisat der Thüringer Universitäts- und Landesbibliothek aus gelangen Nutzerinnen und Nutzer zum Webarchiv der Deutschen Nationalbibliothek und erhalten hier Zugriff auf die archivierte Netzressource. Um Verwechslungen mit der jeweils aktuellen Website auszuschließen, sind die einzelnen Zeitschnitte der eingesammelten Website als Archivversion markiert und entsprechend datiert.

Wer es bevorzugt, direkt im Webarchiv Thüringen zu browsen, dem bietet sich die Möglichkeit, über das Portal der Deutschen Nationalbibliothek auf den digitalen Bestand zuzugreifen.¹⁶ Hier lässt sich das Webarchiv Thüringen über eine alphabetische Liste identifizieren und per Mausklick öffnen. Die archivierten Websites sind alphabetisch aufgelistet und dort ebenfalls in einzelnen Zeitschnitten

16 Das Webarchiv Thüringen ist im Rahmen des Webarchivs der Deutschen Nationalbibliothek über deren zentrale Webpräsenz zugänglich (Einstieg über Digitale Sammlungen > Websites). Online: <https://www.dnb.de/DE/Home/home_node.html>, Stand: 10.10.2020.

zugänglich. Unter „Ereignisse Thüringen“ findet sich sodann der Zugang zu den ereignisbasierten Sammlungen. Für Letztere muss die Frage der bibliographischen Verzeichnung allerdings noch abschließend geklärt werden, da hier nicht nur komplette Webauftritte, sondern auch (unselbstständige) Teile von Websites oder Unterseiten enthalten sind, welche als solche nur schwer zu erfassen sind.

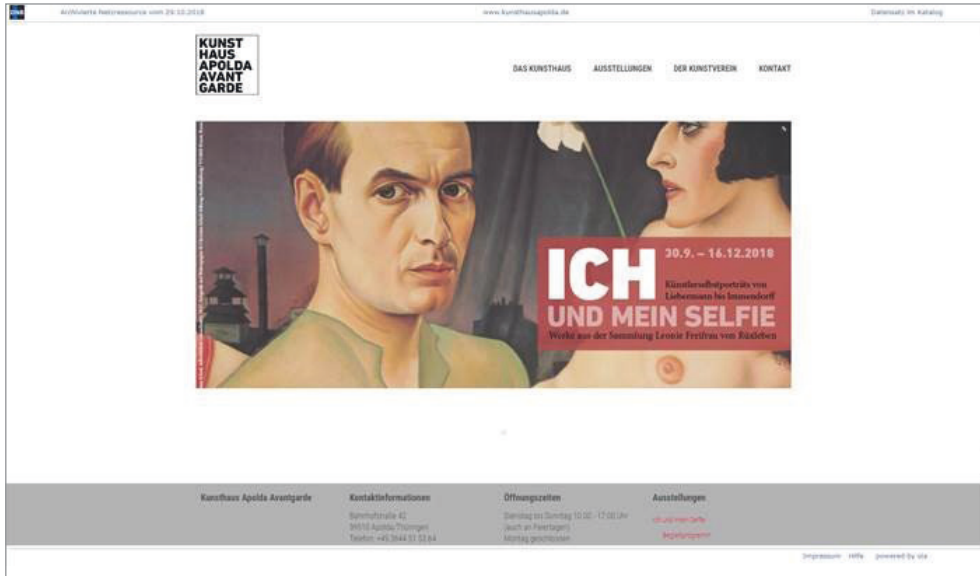


Abb. 3: Zeitschnitt einer archivierten Website aus dem Webarchiv Thüringen

Der Zugang zum Webarchiv Thüringen ist für Nutzerinnen und Nutzer der Thüringer Universitäts- und Landesbibliothek auf die Räume der Thüringer Universitäts- und Landesbibliothek beschränkt. Diese Einschränkung basiert auf der sog. Bibliotheksschranke im Urheberrechtsgesetz, die eine solche Beschränkung vorschreibt. Allerdings haben Nutzerinnen und Nutzer im Rahmen der Kooperation mit der Deutschen Nationalbibliothek die Möglichkeit, nicht nur die Inhalte des Webarchivs Thüringen zu nutzen, sondern auch auf die übrigen Website-Sammlungen der Deutschen Nationalbibliothek zuzugreifen.

8. Fazit und Ausblick

Webarchivierung ist wichtig und notwendig und reichert das landesbibliothekarische Profil mit neuen Aufgaben an. Wie aufgezeigt wurde, verfolgt die Thüringer Universitäts- und Landesbibliothek im Bereich der Webarchivierung keine isolierten Lösungen, sondern arbeitet als bundesweit erste Partnereinrichtung seit dem Jahr 2018 mit der Deutschen Nationalbibliothek zusammen. Das so ins Leben gerufene Webarchiv Thüringen steht der Forschung und der interessierten Öffentlichkeit als Teil der landesbibliothekarischen Dienstleistungen der Thüringer Universitäts- und Landesbibliothek

fortan zur Verfügung. Damit werden ausgewählte, landesspezifische Webangebote gesichert, und die Thüringer Universitäts- und Landesbibliothek erweitert ihr digitales Serviceangebot um archivierte Websites. Dass durch die Zusammenarbeit zwischen der Deutschen Nationalbibliothek und der Thüringer Universitäts- und Landesbibliothek eine Win-win-Situation entsteht, liegt ebenfalls auf der Hand: So erfährt die Thüringer Universitäts- und Landesbibliothek hinsichtlich der technischen Aspekte des Einsammelns, der Speicherung und Langzeitarchivierung der Websites eine deutliche Entlastung, und umgekehrt kommt die Deutsche Nationalbibliothek auf diesem Weg an Websites, die zu ihrem Sammelauftrag gehören, ohne dafür jedoch den Aufwand des Auswählens zu haben.

Was sind die noch offenen Fragen und welche Perspektiven eröffnen sich für die Weiterentwicklung des Webarchivs Thüringen? Ähnlich wie es Tobias Beinert in einem Beitrag zur Webarchivierung an der Bayerischen Staatsbibliothek für das Segment der digitalen Amtsschriften aufgezeigt hat,¹⁷ wäre auch für Thüringen zu prüfen, ob sich insbesondere mit Blick auf die amtlichen Websites durch die Webarchivierung neue Perspektiven ergeben. Amtliche Publikationen werden auch in Thüringen, wie bereits aufgezeigt, als eigenständige Veröffentlichungen gesammelt, erschlossen und archiviert, aber in vielen Fällen auch in das Webarchiv übernommen. Das zweigleisige Vorgehen ist perspektivisch zu hinterfragen, wobei zu berücksichtigen ist, dass die Thüringer Universitäts- und Landesbibliothek (zumindest noch) nicht über die digitalen Kopien der eingesammelten Websites verfügt.

Zu klären ist neben der Profilierung des Geschäftsgangs und der Bereitstellung entsprechender personeller Ressourcen an der Thüringer Universitäts- und Landesbibliothek die nicht minder wichtige Frage, ob und wie es perspektivisch gelingen kann, die Archivkopien der eingesammelten Websites auf Speicherungssysteme der Thüringer Universitäts- und Landesbibliothek umziehen zu lassen, oder ob im Sinne der bestehenden Kooperation darauf gänzlich verzichtet werden soll. Dies ist eine zentrale Frage, die in Zukunft beantwortet werden muss. Als weitere Aufgaben kommen die Weiterentwicklung sowie der Ausbau der Sammlung, die vertragliche Ausgestaltung der Kooperation mit der Deutschen Nationalbibliothek und *last but not least* deren nachhaltige Ausfinanzierung hinzu. Ein wichtiges Monitum stellt auch das Fehlen der persistenten Identifizierung der Website-Kopien durch einen Digital Object Identifier (DOI) oder Uniform Resource Name (URN) dar.

Darüber hinaus wäre es sehr zu begrüßen, wenn sich auch in Deutschland Initiativen finden könnten, die der gestiegenen Relevanz von sozialen Netzwerken wie Twitter, Facebook und Co Rechnung tragen. Während das Konzept der HTML-basierten Website nun schon seit mehr als 30 Jahren existiert, hat sich das Web in der vergangenen Dekade vor allem hinsichtlich der sozialen Medien rasant weiterentwickelt. Gemessen an ihrer gestiegenen Relevanz sind Inhalte aus sozialen Medien in den digitalen Sammlungen der Bibliotheken jedoch noch nicht berücksichtigt. Dass deren Archivierung ebenfalls gelingen kann, zeigen internationale Beispiele von Webarchiven, welche z. B. auch Tweets von Regierungen in ihr Sammel- und Archivierungsprofil erfolgreich integriert haben.¹⁸

17 Vgl. Beinert, Tobias: Webarchivierung an der Bayerischen Staatsbibliothek, in: Bibliotheksdienst 51, 2017, H. 6, S. 490–499. Online: <<https://doi.org/10.1515/bd-2017-0052>>.

18 Vgl. stellvertretend für andere Einrichtungen das UK Government Archive der britischen National Archives. Online: <<https://webarchive.nationalarchives.gov.uk/twitter/>>, Stand: 10.10.2020.

Literaturverzeichnis:

- Ammendola, Andrea Pietro: Webarchivierung in NRW aus Sicht der Universitäts- und Landesbibliothek Münster, Köln 2020, S. 82. Online: <<https://nbn-resolving.org/urn:nbn:de:hbz:79pbc-opus-15775>>.
- Beinert, Tobias: Webarchivierung an der Bayerischen Staatsbibliothek, in: Bibliotheksdienst 51, 2017, H. 6, S. 490–499. Online: <<https://doi.org/10.1515/bd-2017-0052>>.
- Bibliotheksdienst 51, 2017, H. 6 [Themenschwerpunkt Sammlungen von Websites an deutschen Regionalbibliotheken].
- Müller, Sophie: Webarchivierung aus kulturwissenschaftlicher Perspektive, in: Dialog mit Bibliotheken 30, 2018, H. 2, S. 61. Online: <<https://nbn-resolving.org/urn:nbn:de:101-2018091409>>.
- Mutschler, Thomas: Neue Wege der Kulturgutdigitalisierung in Thüringen, in: Bibliotheksdienst 51, 2017, H. 3–4, S. 310–321.
- Mutschler, Thomas: Thüringen, in: Bibliotheksdienst 47, 2013, H. 8–9, S. 659–663. Online: <<https://doi.org/10.1515/bd-2013-0078>>, Stand: 10.10.2020.
- Mutschler, Thomas: Webarchivierung in Thüringen. Sachstand, Planungen, Perspektiven, Präsentation gehalten anlässlich der UAG „Pflichtexemplar“ am 04.04.2019 an der Saarländischen Universitäts- und Landesbibliothek Saarbrücken, Jena 2019. Online: <<https://doi.org/10.22032/dbt.40535>>.
- Niggemann, Elisabeth: Im weiten endlosen Meer des World Wide Web. Vom Sammelauftrag der Gedächtnisinstitutionen, in: Zeitschrift für Bibliothekswesen und Bibliographie 62, 2015, H. 3-4, S. 157. Online: <<http://dx.doi.org/10.3196/1864295015623439>>.
- Niu, Jinfang: An Overview of Web Archiving, in: D-Lib Magazin 18, 2012, H. 3-4. Online: <<https://doi.org/10.1045/march2012-niu1>>.
- Pfau, Pia und Wiesenmüller, Heidrun: Die Sammlung von amtlichen Veröffentlichungen in elektronischer Form an regionalen Pflichtexemplarbibliotheken, in: o-bib 7, 2020, H. 1, S. 7ff. Online: <<https://doi.org/10.5282/o-bib/5583>>.
- Rinn, Reinhard: Metadaten und nationalbibliographische Verzeichnung von Netzpublikationen, in: Wefers, Sabine (Hg.): „Nur was sich ändert, bleibt“. 88. Deutscher Bibliothekartag in Frankfurt am Main 1998, Frankfurt am Main 1999 (Zeitschrift für Bibliothekswesen und Bibliographie, Sonderheft 75), S. 223.
- Zeitschrift für Bibliothekswesen und Bibliographie 62, 2015, H. 3–4 [Themenheft zur Webarchivierung].