# Webarchivierung für Viele
## Vorschlag für eine kooperative Webarchivierungsinfrastruktur

*Mona Ulrich, Staatliche Akademie der Bildenden Künste Stuttgart*
*Claus-Michael Schlesinger*
*Leonie Rodrian, Zentral- und Landesbibliothek Berlin*

**Zusammenfassung**
Der Artikel skizziert Notwendigkeiten und Potenziale für eine Ausweitung von Webarchivierungsaktivitäten durch Kompetenzaufbau und Infrastruktur vor dem Hintergrund andauernder Verluste des digitalen Kulturerbes im Web. Für den Bereich Kompetenzaufbau wird eine Reihe von Workshops zur Einführung in die Webarchivierung an der Amerika-Gedenkbibliothek in Berlin sowie das Rahmenkonzept für eine partizipative Webarchivierung der Zentral- und Landesbibliothek Berlin dokumentiert. Mit Bezug zur Verlustbeschreibung und zu Bedarfen aus den Workshops wird ein Vorschlag für eine minimale kooperativ nutzbare Webarchivierungsinfrastruktur auf Basis der Software Browsertrix vorgestellt.

**Summary**
Against the backdrop of the ongoing loss of digital cultural heritage, this article describes the need for and the potential of expanding web archiving initiatives through training and infrastructure development. The paper describes a series of introductory workshops on web archiving held at the Amerika-Gedenkbibliothek in Berlin, as well as the framework concept for participatory web archiving developed at the Zentral- und Landesbibliothek Berlin. It proposes a minimal solution for a cooperative web archiving infrastructure based on the Browsertrix software to address the ongoing loss and the needs identified in the workshop series.

**Schlagwörter:** Webarchivierung, Infrastruktur, Kooperation

**Zitierfähiger Link (DOI):** https://doi.org/10.5282/o-bib/6283

**Autorenidentifikation**: Mona Ulrich, ORCID: 0000-0001-9591-5614,
Claus-Michael Schlesinger, ORCID: 0000-0001-6718-5773,
Leonie Rodrian



## 1. Einleitung: Verluste

Das Web gehört mittlerweile zu den gesellschaftlichen Gegebenheiten, die so wirken können, als wären sie schon immer da gewesen. Für Menschen, die seit den 1990er Jahren geboren worden sind, ist das in Bezug auf die eigene Erinnerung auch richtig. Wer älter ist, ist Zeitzeug*in einer Vergangenheit, die diese *web natives* nicht mehr erfahren haben. In den rund 35 Jahren seit den initialen Imple-

mentierungen von Protokollen und Software, die das Web begründen, wurde viel produziert, publiziert und rezipiert, aber nur wenig archiviert. Mediengeschichtlich befindet sich das Web also noch in einer Art Inkunabelzeit, in der das, was im Medium erscheint, zirkuliert und gesellschaftlich wirksam ist. Seit Ende der 1990er Jahr wird die Notwendigkeit der Webarchivierung in der Fachdiskussion und insbesondere von Nationalbibliotheken thematisiert, was jedoch in Deutschland bisher nicht zu den breiten Archivierungsaktivitäten geführt hat, die nötig wären, um mit den seit dieser Zeit andauernden Verlusten umzugehen. Die tatsächliche Praxis erweist sich auch international als eher heterogen.[1] Das damit verbundene Verlustproblem hat Kathrin Passig in ihrem Vortrag auf einer Tagung zu den "Literaturarchiven der Zukunft" auf drastische Weise mit einem Bild der brennenden Anna-Amalia-Bibliothek in Weimar illustriert. Digitale Gegenstände gingen, so Passig, in einem ähnlichen Umfang verloren – jeden Tag.[2]

Ein Grund für den andauernden Verlust ist die spezifische Publikationsform von Webseiten. Die Metaphorik der „Seite" ist hier trügerisch, denn Webseiten existieren nur im Original, es gibt keine identischen Exemplare, mit denen Verluste kompensiert oder im Bedarfsfall neue Kopien erstellt werden könnten. Eine abgeschaltete Webseite ist schlicht nicht mehr erreichbar. Eine Studie des Pew Research Center hat in diesem Zusammenhang die durchschnittliche Dauer für die Erreichbarkeit von Webseiten im *live web* berechnet. Geprüft wurde auf Basis von zeitlich gestaffelten Stichproben aus dem Common Crawl, einem umfangreichen und frei verfügbaren Webarchiv. Die Fragestellung war, welcher Anteil der in den Stichproben enthaltenen Seiten nach bis zu zehn Jahren noch erreichbar war.[3] Aus der Stichprobe des Vorjahrs 2022, relativ zum Auswertungszeitpunkt im Jahr 2023, waren bereits neun Prozent der Seiten nicht mehr erreichbar. Aus der Stichprobe der zehn Jahre alten Crawls aus dem Jahr 2013 waren 38 Prozent der Seiten nicht mehr erreichbar.[4] Die Haltbarkeit von Webseiten im *live web* ist damit recht kurz. Verbunden mit der Publikationsform – keine Kopien, keine Exemplare – hat man es im Web also mit einem schnellen, umfangreichen und konsequenten Verlust zu tun, der nachträglich kaum kompensiert werden kann.

Damit ist noch nicht bemessen, welcher Anteil der nicht mehr im *live web* erreichbaren Seiten auch nicht archiviert wird, schließlich setzt die Methodik der Studie voraus, dass die geprüften Seiten im Common Crawl enthalten sind. In einer sehr viel kleineren und spezifischen Studie für den Bereich Netzliteratur ließ sich zeigen, dass der Verlust auf diesem Feld zwar weniger dramatisch ist, weil hier bereits seit vielen Jahren institutionalisiert archiviert wird. Von 71 geprüften literarischen Arbeiten waren 34 Prozent nicht mehr im *live web* erreichbar. 14 Prozent waren noch nicht archiviert, und

1 So etwa die Feststellung von Costa et al., dass trotz beobachtbarer Zunahme von Webarchivierungsaktivitäten der Umfang der Archive im Verhältnis zum Gegenstand klein bleibt. Costa, Miguel; Gomes, Daniel; Silva, Mário J.: The evolution of web archiving, in: International journal on digital libraries 18 (3), 2017, S. 191–205. https://doi.org/10.1007/s00799-016-0171-9.

2 Passig, Kathrin: 3,5 Kilo Daten, in: Frankfurter Rundschau, 18.03.2022, I, Stand: 04.02.2026; Der Artikel ist Stand 23.08.2026 nicht mehr online. Im Literaturverzeichnis ist für Webquellen ohne DOI jeweils zusätzlich ein Link zu einer archivierten Fassung hinterlegt, sofern verfügbar; Passig, Kathrin: Den Heuhaufen archivieren, Deutsches Literaturarchiv Marbach, März 2021, https://docs.google.com/presentation/d/1CQuVkXaDvsl0psAfkCSkzBcyitxsAAIXEjhLmQOIVaY, Stand: 08.08.2023; Passig, Kathrin: Rucksack oder Rechenzentrum, DHd 2022 Kulturen des digitalen Gedächtnisses, 11.03.2022, https://www.dhd2022.de/closing-keynote/, Stand: 19.07.2023.

3 Chapekis, Athena; Bestvater, Samuel; Remy, Emma u. a.: When online content disappears, Pew Research Center, 2024, https://www.pewresearch.org/?p=167501, Stand: 26.05.2026.

4 Ebd.; vgl. Weigle, Michele: Some URLs are immortal, most are ephemeral, Web science and digital libraries, 20.09.2024, https://ws-dl.blogspot.com/2024/09/2024-09-20-some-urls-are-immortal-most.html, Stand: 18.04.2026.

5,6 Prozent waren weder im *live web* noch in einem der für die Erhebung genutzten Archive erreichbar und müssen damit vorerst als möglicherweise verloren gelten. Einbezogen wurden die Netzliteratur-Archive des Deutschen Literaturarchivs Marbach, des Innsbrucker Zeitungsarchivs, der Electronic Literature Organization und das Internet Archive.[5]

Vor diesem Hintergrund wurde im Rahmen des Projekts Science Data Center for Literature (Deutsches Literaturarchiv Marbach und Universität Stuttgart) im Jahr 2022 ein erster Workshop zur Einführung in die Webarchivierung konzipiert und mit dem geschichtswissenschaftlichen Projekt #SocialMediaHistory durchgeführt. Ziel war es, den Forschenden Methoden der Webarchivierung zu vermitteln, damit sie ihre Forschungsgegenstände – in diesem Fall Social-Media-Inhalte – erhalten können.[6] Im Jahr 2023 wurde der Workshop „Einstieg in die Webarchivierung – mit Low Hanging Fruits und vielseitigen Herausforderungen" im Bereich Kunst- und Kulturarchive am Valie Export Center Linz angeboten.[7] Seit 2024 wird der Workshop in Kooperation mit der Zentral- und Landesbibliothek Berlin weiter entwickelt und zweimal jährlich angeboten.[8] Das Angebot richtet sich an Personen in den Bereichen Bibliotheken, Archive, Forschung und soziale Communities mit Archivierungsinteressen. Die bundesweite Resonanz auf das zunächst für den Berliner Raum gedachte Angebot – die Workshops waren jeweils vollständig ausgebucht – hat gezeigt, dass es in den adressierten Bereichen einen starken Bedarf und damit auch ein gutes Potenzial für einen Ausbau von Webarchivierungsaktivitäten gibt. Die Gruppe der inzwischen mehr als achtzig Workshopteilnehmenden kam dabei aus ganz unterschiedlichen Bereichen: Stadtarchive, Landesarchive, Spezialarchive, Forschungsbibliotheken, Forschungsprojekte, einzelne Wissenschaftler*innen, und nicht zuletzt historisch orientierte Arbeitsgruppen zur reflektierten Dokumentation politischer Bewegungen oder sozialer Communities. Dabei haben sich drei Desiderate gezeigt, die für eine Verstetigung und Ausweitung der Webarchivierungsaktivitäten gebraucht werden: Webarchivierungsinfrastruktur, Vernetzung und Expertise (Beratung). Die Evaluation zeigt, dass ein Großteil der Teilnehmenden weiter mit dem in den Workshops für die automatisierte Webarchivierung bereitgestellten Tool Browsertrix arbeiten möchte. Aufbau und Betrieb einer eigenen Webarchivierungsinfrastruktur mit Browsertrix oder einem vergleichbaren Tool wird jedoch in der Regel aus Kapazitätsgründen nicht verfolgt, auch die Buchung von Webarchivierungskapazitäten bei Anbietern wie Webrecorder mit Browsertrix oder Archive-IT (Internet Archive)[9] ist in vielen Fällen finanziell nicht darstellbar. Im Folgenden möchten wir vor diesem Hintergrund deshalb einen programmatischen Vorschlag für eine minimale arbeitsfähige Lösung machen, mit der diese Bedarfe adressiert werden können.

5 Schlesinger, Claus-Michael; Ulrich, Mona; Jorczick, Adam: Counting the Loss of E-Lit works. How big are our time capsules?, Zenodo, 31.05.2022, https://doi.org/10.5281/zenodo.6587607.

6 Science Data Center for Literature: Webarchivierung für Forscher*innen, 18.07.2022, https://www.sdc4lit.de/veranstaltungen/default-title/, Stand: 24.07.2026; SocialMediaHistory. Geschichte auf Instagram und TikTok, https://smh.blogs.uni-hamburg.de, Stand: 24.07.2026.

7 Ulrich, Mona: Einstieg in die Webarchivierung – mit Low Hanging Fruits und vielseitigen Herausforderungen, Valie Export Center Linz, https://www.valieexportcenter.at/einstieg-in-die-webarchivierung-mit-low-hanging-fruits-und-vielseitigen-herausforderungen, Stand: 23.07.2026.

8 Foliensatz zum Workshop: Ulrich, Mona; Schlesinger, Claus-Michael: Workshop Webarchivierung. Clientseitige Webarchivierung, Zenodo, 14.10.2024, https://doi.org/10.5281/zenodo.13918959.

9 Webrecorder: Browsertrix, https://webrecorder.net/browsertrix/, Stand: 29.07.2026; Internet Archive: Archive-IT, https://archive-it.org/, Stand: 29.07.2026.

## 2. Institutioneller Ausbau von Webarchivierungsaktivitäten und partizipative Ansätze

Auch wenn im Verhältnis zur Größe der Aufgabe die Aktivitäten im Bereich Webarchivierung für den deutschsprachigen Raum weiterhin als unzureichend einzuschätzen sind,[10] bewerten Bibliotheken, Archive und Forschungseinrichtungen im deutschsprachigen Raum Webarchivierung zunehmend als relevanten Aufgabenbereich. Davon zeugen themenspezifische Konferenzen wie „Social Media Archivieren" (2024 und 2026),[11] die Keynote von Kathrin Passig bei der Konferenz „Digital Humanities im deutschsprachigen Raum 2022",[12] die reflektierende Beschäftigung mit Webarchivierungsthemen in bibliothekarischen Fachzeitschriften, so zum Beispiel die Themenhefte der Zeitschrift für Bibliothekswesen und Bibliografie aus den Jahren 2015 und 2025[13], und nicht zuletzt die Beschäftigung mit dem Thema im Handbuch und in Veranstaltungen des nestor-Netzwerks.[14]

Hinzu kommen der Ausbau eigener Webarchivierungsinfrastrukturen und die Implementierung von langfristig orientierten Webarchivierungsstrategien an größeren Institutionen wie der Deutschen Nationalbibliothek, der Landesbibliotheken Baden-Württemberg, der Bayerischen Staatsbibliothek oder der Zentral- und Landesbibliothek Berlin (ZLB). So arbeitet die Deutsche Nationalbibliothek aktiv an der Implementierung einer erneuerten Webarchivierungsstrategie und verstärkt die Zusammenarbeit mit Landes- und Regionalbibliotheken.[15] Auch auf Ebene der Landesbibliotheken und Landesarchive werden kooperative Strategien, Teams und Angebote weiter ausgebaut, um Webarchivierung langfristig umzusetzen. So beispielsweise im Rahmen des DIMAG-Verbunds, für den das Landesarchiv Baden-Württemberg ein Modul für die Webarchivierung (DIWI) entwickelt hat.[16] Eine weitere länderübergreifende Vernetzung bildet der Kooperationsverbund Digitale Archivierung Nord (DAN), der eine gemeinsame Infrastruktur mit der DIMAG-Software über einen IT-Dienstleister betreibt. Das Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) bietet seit 2004 einen Dienst für die Webarchivierung an, der bis 2017 auf einer eigenständig entwickelten und betriebenen Plattform basierte und seitdem von einem externen Dienstanbieter, Archive-IT des Internet Archive, eingekauft wird.[17] Mit SWBregio betreibt das BSZ auf dieser Basis einen Webarchivierungs-Dienst, der sich an kommunale und regionale Archive richtet, die Webarchivierung betreiben wollen.[18]

10 Beinert, Tobias; Schmid, Katharina; Weimer, Konstanze: Infrastrukturen und Services für die wissenschaftliche Nutzung von Webarchiven, in: o-bib. Das offene Bibliotheksjournal 9 (3), 2022, S. 1–15. https://doi.org/10.5282/O-BIB/5821.

11 Deutsche Nationalbibliothek: Nachhaltige Archivierung sozialer Medien. Twitter und danach, 19.12.2023, https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/Twittertagung/twittertagung2024.html, Stand: 18.04.2026.

12 Passig: Rucksack oder Rechenzentrum.

13 Zeitschrift für Bibliothekswesen und Bibliographie Jahrgang 72 (2), 2025. https://doi.org/10.3196/1864-2950-2025-2; Zeitschrift für Bibliothekswesen und Bibliographie 62 (3–4), 2015. https://doi.org/10.3196/1864-2950-2015-3-4.

14 Liegmann, Hans; Rauber, Andreas: Kap. 17.9 Webarchivierung zur Langzeiterhaltung von Internet-Dokumenten, in: Neuroth, Heike; Oßwald, Achim; Scheffel, Regine u. a. (Hg.): nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Göttingen 2010. https://nbn-resolving.org/urn:nbn:de:0008-20100305365. Netzwerkaktivitäten sind verzeichnet auf nestor: Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland e. V., https://langzeitarchivierung.de; Stand: 29.07.2026.

15 Deutsche Nationalbibliothek: Webarchivierung, 02.04.2025, https://www.dnb.de/webarchiv, Stand: 11.02.2026; Mutschler, Thomas: Zum Stand der kooperativen Webarchivierung in Thüringen, in: o-bib. Das offene Bibliotheksjournal 7 (4), 2020, S. 1–12. https://doi.org/10.5282/O-BIB/5632.

16 Eberlein, Miriam; Jehn, Mathias; Heiden, Detlev u. a.: Entwicklung und Betrieb von DIMAG, in: Archivar 74 (2), 2021, S. 76–83.

17 Hannemann, Renate: Webarchivierung im BSZ mit Archive-It, Konferenzfolien zum 107. Deutschen Bibliothekartag, Berlin 2018, https://nbn-resolving.org/urn:nbn:de:0290-opus4-36116.

Die Tendenz geht also in Richtung mehr Archivierung in Gedächtnisinstitutionen durch kooperative, z.T. auch kollaborativ betriebene Software-Lösungen und Infrastrukturen. Trotz des Ausbaus existiert ein unadressierter Bedarf, da durch die Ausrichtung der Sammlungsprofile nur bestimmte Teile der webbasierten Öffentlichkeit abgedeckt sind. Soziale Bewegungen, wichtige öffentliche Debatten und andere Kommunikationsereignisse im Web werden beispielsweise nur nachrangig oder gar nicht erhalten. Denn kleinere spezialisierte Archive, Bewegungsarchive und Communities haben oftmals keinen Zugang zu entsprechender Webarchivierungsinfrastruktur. Anforderungen der historischen Forschung an einen möglichst breiten Sammlungsaufbau[19] und gesellschaftliche Bedarfe nach Möglichkeiten einer Reflexion sozialer Bewegungen und politischer Diskurse sind damit durch existierende Sammlungsaktivitäten oft nicht abgedeckt. Hinzu kommt, dass eigenständig arbeitende Spezialarchive und historisch orientierte Arbeitsgruppen in sozialen Communities meist selbstverantwortlich archivieren möchten, um bei Auswahl, Objektkonstituierung, Verwaltung und Bereitstellung unabhängig zu bleiben von Sammlungs- und Infrastrukturentscheidungen möglicher Partnereinrichtungen aus dem Bibliotheks- und Archivwesen.

Um den bestehenden Lücken in Sammlungspraxis, Wissensvermittlung und Infrastruktur zu begegnen, setzen neue Lösungsansätze in diesen Bereichen auf Empowerment und Partizipation. Stakeholder werden dann nicht als Empfänger*innen oder Nutzer*innen von Dienstleistungen verstanden, sondern wirken gestaltend und verantwortlich am Aufbau und der Nutzung von Strukturen mit. Zwei wichtige Elemente solcher partizipativen Konzepte im Bereich Webarchivierung sind Angebote für Kompetenzaufbau und Infrastrukturentwicklung, wobei beides in einer wechselseitigen Abhängigkeit steht. Kompetenzaufbau bleibt ohne verfügbare Mittel eingeschränkt; umgekehrt ist eine Bereitstellung von Infrastruktur abhängig von Kompetenzentwicklung, weil ein nachhaltiger Aufbau von Webarchiven und die Konstituierung von Archivobjekten technische, mediengeschichtliche und arbeitspraktische Kenntnisse voraussetzt.

Neben infrastrukturellen Kooperationen zwischen Gedächtnisinstitutionen entstehen deshalb auch partizipative Ansätze, die zivilgesellschaftliche Akteur*innen und Communities einbeziehen. Das nachfolgend ausführlicher vorgestellte, an der ZLB entwickelte Webarchivierungskonzept ist ein Beispiel dafür, wie bibliotheks- und wissenschaftsexterne Communities bei der Entwicklung des Sammlungskonzepts und beim Sammlungsaufbau kooperativ mit eingebunden werden, etwa durch gezielte Austauschformate und Kompetenzvermittlung für ein Empowerment der Community-Partner.

Ein weiterer Anwendungsbereich partizipativer Webarchivierungsinfrastrukturen ergibt sich aus der Forschung. Webinhalte sind zunehmend Gegenstand wissenschaftlicher Untersuchungen, wodurch Webarchivierung für Forschende zur Konstituierung und Sicherung ihrer Forschungsgegenstände relevant wird. So weist bereits die Einleitung zum ZfBB-Themenheft zur Webarchivierung 2015 auf die Bedeutung von Webinhalten für die Forschung und die daraus abgeleiteten Bedarfe und Anforderungen an archivierende Institutionen hin.[20] Im Anschluss daran zeigt sich inzwischen, dass Forschende

18 Beinert, Tobias; Hagenah, Ulrich; Kugler, Anna: Es war einmal eine Website. Kooperative Webarchivierung in der Praxis, in: o-bib. Das offene Bibliotheksjournal 1 (1), 2014, S. 291–304. https://doi.org/10.5282/O-BIB/2014H1S291-304.

19 Müller, Sophie: Webarchivierung aus kulturwissenschaftlicher Perspektive, in: Dialog mit Bibliotheken 30 (2), 2018, S. 61–62.

20 Altenhöner, Reinhard; Oßwald, Achim: Im Fokus: Webarchivierung in Bibliotheken. in: Zeitschrift für Bibliothekswesen und Bibliographie 62 (3–4), 2015, S. 139–143. https://doi.org/10.3196/1864-2950-2015-3-4

im Zuge von Forschungsprozessen und Forschungsdatenmanagement auch selbst archivieren oder an Archivierungsprozessen aktiv beteiligt sind. Daraus lässt sich ergänzend ein Bedarf für nutzbare Webarchivierungsinfrastruktur und Kompetenzvermittlung ableiten. Das Angebot „Webarchivierung Aix Marseille" (WAAM) des Research Data Training and Support Center (CEDRE) der Universität Marseille ist ein Beispiel für den Betrieb einer Webarchivierungsplattform auf Basis von Browsertrix, die der Forschungscommunity in einem geografisch lokal begrenzten Gebiet — hier Aix-en-Provence und Marseille — zur Verfügung steht. Diese ist gleichzeitig an die nationale französische Pflichtabgabestruktur der Bibliothèque nationale de France angeschlossen.[21] Eine weitere kooperative Initiative zum Erhalt forschungsbezogener Inhalte ist *Rogue Scholar*, ein von der Firma Front Matter entwickeltes Wissenschaftsblog-Repository, das in Zukunft über eine Genossenschaft von Organisationen aus den Bereichen Bibliothek und Wissenschaft finanziert und gesteuert werden soll.[22] Ein Beispiel für projektbezogene kooperative Archivierung ist schließlich das Twitter-Archiv an der Deutschen Nationalbibliothek, das in Zusammenarbeit mit dem Forschungs- und Infrastrukturprojekt Science Data Center for Literature unter Beteiligung von weiteren Forschenden erstellt wurde.[23]

## 3. Partizipative Webarchivierung an der Zentral- und Landesbibliothek Berlin

Die Zentral- und Landesbibliothek Berlin hat den gesetzlichen Auftrag, in Berlin veröffentlichte Publikationen zu archivieren und der Öffentlichkeit zugänglich zu machen. Mit der Novellierung des Berliner Pflichtexemplar-Gesetzes im Jahr 2021 wurde dieser Auftrag um digitale Publikationen erweitert, zu denen auch Webseiten zählen. Angesichts der großen Menge an Online-Inhalten und der begrenzten Ressourcen, die der ZLB zu deren Erhalt zur Verfügung stehen, wird deutlich, dass Webseiten nur in Auswahl archiviert werden können. Einen wichtigen Aspekt bei der Erfüllung des gesetzlichen Auftrags stellt damit die kontinuierliche Auseinandersetzung mit der Frage dar, wie Auswahl getroffen wird, was Eingang in das Archiv findet und welche Zugangsbarrieren es gibt. Das Ziel war und bleibt, Prozesse für die Praxis zu entwickeln, die möglichst transparent und inklusiv gestaltet sind.

2023 stand für die ZLB die konzeptuelle Auseinandersetzung mit der Frage der Beteiligung der Stadtgesellschaft bei der Webarchivierung im Mittelpunkt. In verschiedenen öffentlichen Veranstaltungen wurde grundlegend diskutiert, was unter digitalem kulturellem Erbe verstanden wird, welche Verantwortung größere Kulturinstitutionen für dessen Erhalt tragen und welche Ausschlussmechanismen in bestehenden Archivierungspraktiken wirken. Dabei zeigte sich auch, dass es Gruppen gibt, die öffentlichen Institutionen wie der ZLB mit Skepsis begegnen und die Frage nach Zusammenarbeit auf Augenhöhe wurde betont.

21 Gebeil, Sophie: WAAM. Web Archiving Aix Marseille, Pôle Bibliothèques et Archives de la MMSH, 03.10.2025, https://pba.mmsh.fr/?p=35306, Stand: 27.05.2026.

22 Fenner, Martin: Rogue Scholar is becoming a German non-profit organization, Front Matter, 03.11.2025, https://doi.org/10.53731/rftfk-qv692.

23 Schlesinger, Claus-Michael; Woldering, Britta: „Den Heuhaufen archivieren". Archivierungsinitiative für deutschsprachige Tweets, in: Zeitschrift für Bibliothekswesen und Bibliographie 71 (4), 2024, S. 236–242. https://doi.org/10.3196/186429502471445.

Ein zentraler Impuls aus den öffentlichen Diskussionen war die Forderung, dass Institutionen, die durch die Unterstützung der öffentlichen Hand strukturell privilegiert sind, dieses Privileg reflektieren und ihre Ressourcen mit weniger Privilegierten teilen. Diese Forderung bildete eine wesentliche Motivation für das Angebot des Workshops Webarchivierung.

Wichtig für eine langfristig auf Beteiligung ausgelegte Sammlungspraxis ist es außerdem, Austauschorte zu schaffen, an denen Netzwerke und Beziehungen entstehen können. Neben dem Workshop Webarchivierung sollen auch die Online-Veranstaltungsreihe „collect and connect" und der in Kooperation mit der Technologiestiftung Berlin organisierte Stammtisch Webarchivierung Gelegenheit dazu geben.[24]

Die erste partizipativ gestaltete Webseiten-Sammlung der ZLB entstand 2024 in Kooperation mit den Bezirksbibliotheken und deren Kooperationspartner*innen. Diese wurden gebeten, Webseiten von Personen, Projekten, Vereinen und Initiativen vorzuschlagen, die das Leben in Berlin und seinen Kiezen prägen. Nach Zustimmung der Betreiber*innen wurden die Webseiten archiviert und sind je nach Zugriffsbeschränkung entweder online oder in den Räumlichkeiten der ZLB einsehbar und über den VÖBB-Katalog recherchierbar.[25]

## 4. Workshops und ihre Zielsetzung

Der oben genannte eineinhalbtägige Workshop wurde in den Jahren 2024-2026 zweimal jährlich an der Amerika-Gedenkbibliothek in Berlin durchgeführt.[26] Er richtet sich an alle Interessierten; Vorkenntnisse sind nicht vorausgesetzt. Bis dato nimmt vor allem Fachpublikum aus den Bereichen Bibliothek, Archiv, Museum und Wissenschaft das Angebot wahr, aber auch Vertreter*innen kleinerer Archive und aktivistischer Projekte finden sich unter den Teilnehmenden, die beispielsweise Webseiten von ausgelaufenen Forschungsprojekten, politischen Bewegungen oder themenspezifischen Communities bewahren möchten. Mit einem praxisorientierten Einstieg werden Grundlagen zu Webseiten als Archivierungsgegenstand, zur clientseitigen Webarchivierung und einem minimalen Workflow vermittelt, der in den Hands-on-Phasen praktisch eingeübt wird.

Einerseits soll der Workshop Menschen befähigen, selbst Webseiten zu archivieren und so wichtige Netzkultur ihrer Gruppen oder Interessengebiete vor dem Verschwinden zu bewahren, ohne dabei die Kontrolle über diese Kulturgüter an Institutionen abgeben zu müssen, denen sie möglicherweise nicht ausreichend vertrauen. Andererseits sind die begrenzten Ressourcen, die den Umfang der Webarchivierungsbemühungen der ZLB generell stark einschränken, ebenfalls Motivation für den Workshop. Die ZLB ist aktuell nicht in der Lage, großflächig Kultur aus dem Netz zu archivieren. Angesichts der Dringlichkeit dieser Aufgabe erscheint es sinnvoll, die Verantwortung auf mehrere Schultern zu verteilen und nicht auf eine hypothetische zukünftige Ausweitung der eigenen Ressourcen und Vorhaben zu warten.

24 Zentral- und Landesbibliothek Berlin: collect and connect. Die Talkreihe der Landesbibliothek digital, https://digital.zlb.de/viewer/collect-and-connect/, Stand: 23.07.2026.

25 Zentral- und Landesbibliothek Berlin: Webarchiv der Digitalen Landesbibliothek Berlin, https://digital.zlb.de/viewer/webarchiv/, Stand: 26.05.2026.

26 Ulrich; Schlesinger: Workshop Webarchivierung, 2024.

Ziel des Workshops ist daher, die Teilnehmer*innen in die Lage zu versetzen, einen Webarchivierungsprozess strukturiert umzusetzen und dafür die vorgestellten Tools selbständig anzuwenden. Im Workshop werden ausschließlich Tools mit grafischer Benutzeroberfläche von Webrecorder vorgestellt, da diese einen niedrigschwelligen Zugang ermöglichen und über umfangreiche Funktionen verfügen, mit denen sich qualitativ hochwertige Webarchive erstellen und wiedergeben lassen.[27] Der Prozess ist in die Schritte Objektanalyse, Capturing, Wiedergabe und Qualitätskontrolle unterteilt. Der Ablauf des Workshops ist an diesem Workflow orientiert. Die Objektanalyse liefert Hinweise für die Auswahl der Methode, des Capturingtools und dessen Handhabung. Zu den analysierten Kriterien zählen insbesondere Umfang, Objektgrenzen, Inhalt, Funktionalitäten und das Konzept der Seite. Beim anschließenden Capturing werden Webarchivierungstools eingesetzt, um die Inhalte im *live web* archivgerecht im WARC-Format zu sichern.[28] Die daraus resultierenden Webarchive können anschließend mit Wiedergabe-Tools als (archivierte) Webseite dargestellt werden. In der Qualitätskontrolle werden die archivierten Webseiten auf Vollständigkeit und Funktionsfähigkeit, basierend auf den Eigenschaften, die bei der Objektanalyse erkannt wurden, überprüft.

Für das Capturing werden zwei Webarchivierungstools vorgestellt und angewendet: Erstens ArchiveWeb.Page für die manuelle Archivierung. Hierbei werden während des Archivierungsprozesses alle zu archivierenden Inhalte manuell aufgerufen. ArchiveWeb.Page ist ein Browser-Add-on und steht somit niedrigschwellig zur Verfügung. Zweitens Browsertrix für die automatisierte Archivierung. Hier werden Webseiten anhand einer festgelegten Konfiguration mithilfe eines Crawlers automatisiert archiviert. Browsertrix ist eine Software mit grafischer Nutzeroberfläche, die für den Workshop temporär als Webservice bereitgestellt wird und eine serverbasierte Installation voraussetzt. Die Anwendung wird von Webrecorder als Open-Source-Software entwickelt und seit 2023 als Webservice angeboten.

Der Ablauf des Workshops folgt einer didaktischen Struktur, bei der sich der Anteil an selbständiger Arbeit der Teilnehmer*innen schrittweise erhöht. Zu Beginn werden technische Grundlagen und der Webarchivierungsworkflow vorgestellt. Anschließend folgen angeleitete Praxisphasen zu den einzelnen Workflowschritten. Den Abschluss bildet ein Archiveathon, in dem die Teilnehmer*innen weitgehend selbständig arbeiten und die gelernten Schritte und Tools selbständig anwenden. Anders als der klare Ablauf vermuten lässt, verläuft die Praxis der Webarchivierung selten linear. Oft kann beim ersten Versuch nicht alles umfassend erkannt, adressiert und archiviert werden. Es ist ein iterativer Prozess, bei dem die Workflowschritte Objektanalyse, Capturing, Wiedergabe und Qualitätskontrolle wiederholt ausgeführt werden müssen. Daher muss ausreichend Zeit für die selbständige Arbeit während des Archiveathons geplant werden, damit die Teilnehmer*innen entlang ihrer Objekte entsprechend arbeiten können.

27 Webrecorder: Web Archiving for All, https://webrecorder.net/, Stand: 18.08.2026. Der Titel dieses Beitrags "Webarchivierung für Viele" ist an den Slogan "Web Archiving for All" von Webrecorder angelehnt und greift das Anliegen auf, Webarchivierung einem breiten Kreis von Interessierten zugänglich zu machen.

28 WARC steht für Web ARChive, es ist aktuell das Standardformat für die Erstellung, Ablage und Weiterverarbeitung von archivierten Webseiten; International Organization for Standardization (ISO): ISO 28500:2017. Information and documentation, WARC file format, 2017, https://www.iso.org/standard/68004.html, Stand: 23.8.2026; siehe ergänzend: International Internet Preservation Consortium (IIPC): The WARC Format 1.1, https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/, Stand: 23.08.2026.

# 5. Vorschlag für eine minimale kooperative Webarchivierungsinfrastruktur

Ein zentrales Desiderat, das sich in den Workshops gezeigt hat, ist der dauerhafte Zugriff für die Teilnehmenden auf die Webanwendung Browsertrix, die für den Workshop nur temporär bereitgestellt wird. Teilnehmende lernen den Umgang mit dem Tool, das die Hauptschritte eines Webarchivierungs-Workflows unterstützt, erarbeiten erste Ansätze zur Arbeit mit den eigenen Gegenständen, können aber später nicht ohne weiteres auf die Anwendung zugreifen. Ähnlich verhält es sich mit dem Mentoring, das nach Ende des Workshops nicht mehr im gleichen Maß verfügbar ist. Daraus folgen drei konkrete Bedarfe: Aufbau und Bereitstellung von Webarchivierungsinfrastruktur, Vernetzung von Akteur*innen und kontinuierliche Bereitstellung von Expertise für die Unterstützung einer fortgesetzten Integration webarchivarischer Workflows im Anschluss an einführende und weiterführende Veranstaltungsformate.

Zielgruppen für ein solches Angebot sind insbesondere kleinere Akteure, also Spezialarchive und Forschungsprojekte, die zu Gegenständen im Bereich des digitalen Kulturerbes arbeiten, die nicht archiviert sind. Hinzu kommen Arbeitsgruppen im Bereich sozialer Bewegungen und Communities, die eine eigenständige und unabhängige geschichtliche Selbstreflexion anstreben.

## 5.1 Use Cases

Nur ein Teil des digitalen Kulturerbes wird von institutionellen Sammlungsprofilen abgedeckt und archiviert. Eine Archivierung ist oft zeitkritisch, da Inhalte im Web sich ständig verändern und auch wieder verschwinden können. Zudem können Archive und Bibliotheken archivierte Webseiten oft nicht online zugänglich machen, sondern nur *on-premise*[29] und auch dann zum Teil mit weitreichenden Einschränkungen, was Zugang und Analysemöglichkeiten einschränkt. Und schließlich kann der Wunsch nach autonomer Archivierung und Kontrolle über das eigene Archiv ausschlaggebend sein, um Inhalte selbständig zu erhalten. Davon abgeleitet lassen sich zwei spezifische Use Cases definieren:

1. Erhalt von Forschungsgegenständen: Für viele geistes- und sozialwissenschaftliche Fächer sind Webinhalte mittlerweile zentrale Forschungsgegenstände. Für die Wahrung einer fortlaufenden Nachvollziehbarkeit von Forschungsergebnissen ist eine Erhaltung der Gegenstände notwendig.
2. Communityorientierte Archivierung: Zivilgesellschaftliche Akteur*innen, Initiativen oder kleinere Organisationen möchten ihre Inhalte langfristig sichern und zugänglich halten.

29 Mit der Bereitstellung *on-premise* ist gemeint, dass Bibliotheken und Archive den Zugang zu Webarchiven auf das eigene Netzwerk oder sogar einen dezidierten Computer vor Ort beschränken. So sind beispielsweise große Teile des Webarchivs der DNB nur an den beiden Standorten in Leipzig und Frankfurt am Main zugänglich. Grund sind meist rechtliche Schranken, die eine offene Bereitstellung im Internet nicht erlauben.

## 5.2 Browsertrix

Die zentrale Software-Komponente der hier vorgeschlagenen kooperativen Webarchivierungsinfrastruktur ist die Open-Source-Software Browsertrix, die 2022 von Webrecorder veröffentlicht wurde und stetig weiterentwickelt wird.[30] Browsertrix ist eine State-of-the-Art-Lösung für den Betrieb einer cloudbasierten Webarchivierungsanwendung mit Nutzungs- und Ressourcenverwaltung, grafischer Benutzungsoberfläche und Programmierschnittstellen zur Anbindung an bestehende Infrastrukturen, insbesondere Speicher.

Alle zentralen Anforderungen an eine Webarchivierungsinfrastruktur werden aus Sicht von Webarchivar*innen von ihr nutzungsfreundlich erfüllt:

- Hochwertige Crawls mit präziser Erfassung von Webseiten
- Grafische Benutzungsoberfläche
- Integriertes Workflow-Management
- Komponente zur Wiedergabe und Bereitstellung
- Integrierte Unterstützungen zur Qualitätskontrolle

Zudem erfüllt die Software aus Betreiber*innenperspektive relevante Anforderungen:[31]

- Open-Source-Lizenz für eigenständigen Betrieb, Anpassung und Weiterentwicklung
- Aktive Nutzer*innen- und Entwickler*innencommunity
- Schnittstelle (API) zur technischen Einbindung in bestehende Infrastrukturen
- Ausführliche Dokumentation

## 5.3 Infrastruktur

Webarchivierung ist compute-intensiv.[32] Daraus ergibt sich ein relevanter Kostenfaktor bei der Inanspruchnahme von Webarchivierungskapazitäten bei externen Dienstanbietern wie Browsertrix oder Archive-IT. Aufbau und Betrieb einer eigenen Webarchivierungsinfrastruktur lohnt sich wegen der Kosten für Compute, Systemadministration und Nutzungsverwaltung erst ab einer bestimmten Größe und Nutzungsintensität der Installation. Kleinere Institutionen und Arbeitsgruppen können daher meist keine eigene Webarchivierungsinfrastruktur betreiben oder externe Ressourcen einkaufen.

Ein Lösungsansatz für dieses Problem muss also erstens Bedarfe bündeln, um mit einem ausreichend großen System Skalierungseffekte zu nutzen und zweitens die Compute-Kosten zu reduzieren. Leisten könnte dies der kooperative Aufbau einer Webarchivierungsinfrastruktur mit zugehörigem Bera-

30 Webrecorder: Browsertrix, Github, https://github.com/webrecorder/browsertrix, Stand: 11.06.2026.

31 Alternative Lösungen mit einem ähnlichen Funktionsumfang, die mit kalkulierbarem Aufwand auf eigener Infrastruktur betrieben werden können, sind derzeit nicht verfügbar.

32 Mit dem Begriff compute-intensiv sind CPU-Ressourcen und Arbeitsspeicher gemeint, die für Webarchivierung benötigt werden. Der Begriff wird als Lehnwort im Bereich Cloud Computing verwendet, um Rechenkapazitäten und -bedarf zu beschreiben.

tungsangebot in den Bereichen Archivierung und Digitale Objekte, Bereitstellung und Analytics. Die Kosten für Compute und Speicherplatz können in einer initialen Phase durch Beteiligung von Universitäts- oder Bibliotheksrechenzentren abgedeckt werden. Webarchivierung ist zwar im Vergleich mit anderen datenorientierten Verfahren im Bereich digitales Kulturerbe aufwändig, verglichen mit den sehr umfangreichen Ressourcen, die etwa für naturwissenschaftliche Simulationsverfahren oder inferenzbasierte Systeme (Künstliche Intelligenz) benötigt werden, ist dieser Aufwand aus Sicht eines großen Rechenzentrums aber gleichzeitig auch überschaubar.

Eine iterative Entwicklung eines solchen Dienstes erscheint vor dem Hintergrund offener Fragen hinsichtlich der Bedarfe für Nutzung und technischen Anschüssen als beste und nachhaltigste Möglichkeit. Iterative Entwicklung bedeutet in diesem Fall, dass ein minimal funktionierender Prototyp in einen Testbetrieb gebracht wird. Ein Testbetrieb kann mit den folgenden Komponenten gewährleistet werden:

1. Serverkapazitäten
2. Systemadministration
3. Nutzungsverwaltung

Für die Systemadministration ist erwähnenswert, dass das Webrecorder-Projekt großen Wert auf gute Dokumentation legt und explizit für Self-Hosting-Support ansprechbar ist.[33] Die Nutzungsverwaltung ist technisch bereits in Browsertrix integriert. Ein Cluster kann von mehreren Arbeitsgruppen unabhängig voneinander genutzt werden. Für die Verwaltung von Arbeitsgruppen und einzelnen Nutzer*innen wird eine verantwortliche Person benötigt. Denkbar ist außerdem, dass Arbeitsgruppen Nutzer*innen und ihre Berechtigungen selbst verwalten (Administration eines abgegrenzten Bereichs mit eingeschränkten Verwaltungsrechten).

33 Webrecorder: Browsertrix, https://webrecorder.net/browsertrix/, Stand: 23.08.2026.

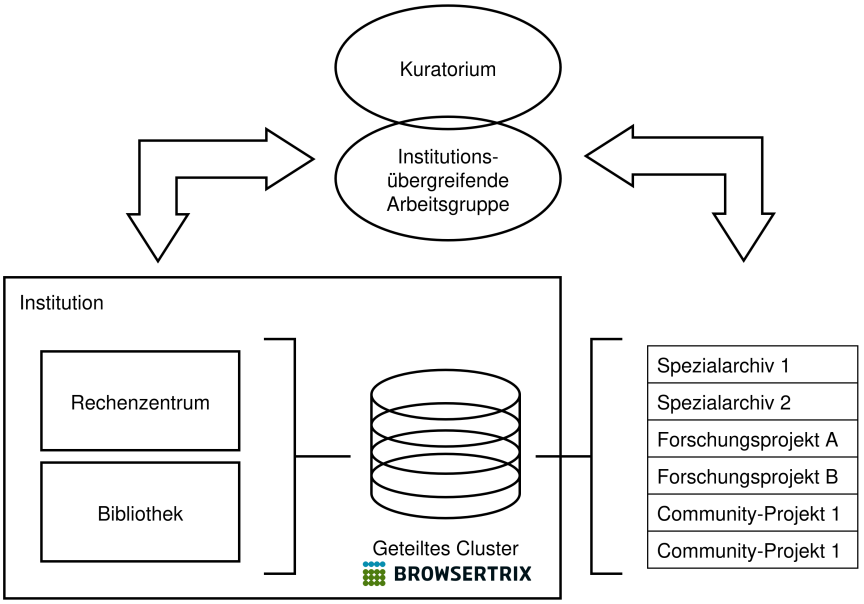


***Abb. 1:*** *Strukturdiagramm eines minimalen Systems mit zugehöriger Governance durch eine übergreifende Arbeitsgruppe und ein Kuratorium. Grafik: Mona Ulrich und Claus-Michael Schlesinger.*

Das Diagramm (Abb. 1) zeigt ein minimales Modell für einen solchen Betrieb einschließlich möglicher Governance-Strukturen, um Entscheidungsprozesse so zu gestalten, dass alle Beteiligten verantwortlich in die Entwicklung des Gesamtsystems einbezogen werden. Eine übergreifende Arbeitsgruppe fördert die Vernetzung der Teilnehmer*innen und ermöglicht eine fortlaufende Entwicklung von Richtlinien für die Nutzung und weitere Entwicklung der Infrastruktur. Ein Kuratorium, bestehend aus Repräsentant*innen der teilnehmenden Stakeholder, also Nutzende und Träger*innen der Infrastruktur, übernimmt die Vorbereitung, Konkretisierung und Umsetzung von entsprechenden Entscheidungen. Spezialarchive, Forschungsprojekte und Communities bringen dabei ihre Expertise in Bezug auf die eigenen Gegenstände mit ein. Dies schafft im Austausch untereinander und mit den institutionellen Infrastruktur-Providern Möglichkeiten für gegenstandsbezogene Kooperationen.

Im minimalen Modell liegt der Fokus auf Archivierung. Teilnehmende erhalten im Ergebnis ein Webarchiv in Form von einer oder mehreren Dateien im WACZ-Format[34], das anschließend eigenständig weiterverarbeitet werden kann, dann jeweils in Verantwortung der Teilnehmenden. Es handelt sich bei diesem Modell also nicht um einen vollumfassenden Dienst, sondern um die Unterstützung der Phase aktiver Archivierung. Fragen der Bereitstellung, Analyse oder Langzeitarchivierung sind hier

34 Das WACZ-Format basiert auf dem WARC-Format und wird vom Webrecorder-Projekt entwickelt und eingesetzt. Es handelt sich um eine Zip-Datei mit den WARC-Dateien und erweiterten Metadaten zum Crawl. WACZ-Dateien vereinfachen den Transport und die Weiterverarbeitung von Webarchiven mit den Webrecorder-Tools und mit Webrecorder-konformen Tools; Web Archive Collection Zipped (WACZ), Webrecorder, 03.06.2021, https://specs.webrecorder.net/wacz/1.1.1/, Stand: 23.08.2026.

nicht berücksichtigt, können aber in weiteren Iterationen perspektivisch ergänzend hinzugenommen werden.

Als Basis für einen initialen Testbetrieb und eine Ausarbeitung von Richtlinien braucht es für dieses Modell eine kleine Gruppe von Partnereinrichtungen und Arbeitsgruppen, um im Sinne einer iterativen Entwicklung handlungsfähig zu werden. Am Horizont sind als Implementierungsziel verschiedene Modelle denkbar, zum Beispiel die Einrichtung einer kooperativen Arbeitsstelle mit Drittmittelstrategie im Kontext aktueller Forschungsdateninitiativen für einen mittelfristigen Betrieb oder langfristig eine genossenschaftliche Struktur unter Beteiligung aller Kooperationspartner. Mit Rückbezug auf den Ausgangspunkt unserer Überlegungen — der aus den Workshops heraus entstandene Bedarf nach fortgesetzter Nutzung von Browsertrix für die eigene Arbeit im Kontext von Archiv, Bibliothek, Forschung und Community — sehen wir die Etablierung der skizzierten Minimalstruktur als wichtigsten nächsten Schritt, der die bereits erhobenen Bedarfe adressiert und auf dieser Basis eine Adressierung auch weiterführender Fragen bezüglich Skalierung und langfristiger Sicherung des Betriebs ermöglicht.

## Fazit

Ausgehend von den umfangreichen Verlusten im Bereich des digitalen Kulturerbes und mit Blick auf aktuelle archivarische und bibliothekarische Aktivitäten im Bereich Webarchivierung wurden Lücken beschrieben, die als Potenzial für einen Ausbau der Webarchivierung gewertet werden können. Die Einbeziehung von Stakeholdern mit direktem Verhältnis zum Gegenstand — Forschung, soziale Communities, Diskursbeteiligte — erschließt zusammen mit Bibliotheken und Archiven Bereiche mit Spezialwissen, das von institutionellen Sammlungsaufträgen und -gegenständen bislang nicht abgedeckt ist. Dieser Vorschlag für den Aufbau einer minimal wirksamen kooperativen Infrastruktur setzt auf bereits verfügbare Strukturelemente mit einer möglichst niedrigen Implementierungsschwelle. Der Fokus auf partizipative Nutzung, Entwicklung und Governance fördert Aktivierung und Vernetzung und ermöglicht im besten Fall einen eigenständigen und kooperativen Aufbau von Sammlungen mit einer jeweils spezialisierten und übergreifend breiten Abdeckung der Medien-, Kultur- und Diskursgeschichte im Web.[35]

## Literaturverzeichnis

Altenhöner, Reinhard; Oßwald, Achim: Im Fokus: Webarchivierung in Bibliotheken. in: Zeitschrift für Bibliothekswesen und Bibliographie 62 (3–4), 2015, S. 139–143. https://doi.org/10.3196/1864-2950-2015-3-4.

Beinert, Tobias; Hagenah, Ulrich; Kugler, Anna: Es war einmal eine Website. Kooperative Webarchivierung in der Praxis, in: o-bib. Das offene Bibliotheksjournal 1 (1), 2014, S. 291–304. https://doi.org/10.5282/O-BIB/2014H1S291-304.

Beinert, Tobias; Schmid, Katharina; Weimer, Konstanze: Infrastrukturen und Services für die wissenschaftliche Nutzung von Webarchiven, in: o-bib. Das offene Bibliotheksjournal 9 (3), 2022, S. 1–15. https://doi.org/10.5282/O-BIB/5821.

35 Alle genannten Autor*innen haben gleichwertig zu diesem Artikel und zur zugrundeliegenden Forschungs- und Vermittlungsarbeit beigetragen.

Chapekis, Athena; Bestvater, Samuel; Remy, Emma u. a.: When online content disappears, Pew Research Center, 2024, https://www.pewresearch.org/?p=167501, Stand: 26.05.2026, archiviert: https://web.archive.org/web/20260802054009/https://www.pewresearch.org/wp-content/uploads/sites/20/2024/05/pl_2024.05.17_link-rot_report.pdf.

Costa, Miguel; Gomes, Daniel; Silva, Mário J.: The evolution of web archiving, in: International journal on digital libraries 18 (3), 2017, S. 191–205. https://doi.org/10.1007/s00799-016-0171-9.

Eberlein, Miriam; Jehn, Mathias; Heiden, Detlev u. a.: Entwicklung und Betrieb von DIMAG, in: Archivar 74 (2), 2021, S. 76–83.

Deutsche Nationalbibliothek: Nachhaltige Archivierung sozialer Medien. Twitter und danach, 19.12.2023, https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/Twittertagung/twittertagung2024.html, Stand: 18.04.2026, archiviert: https://web.archive.org/web/20250904081348/https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/Twittertagung/twittertagung2024.html.

Deutsche Nationalbibliothek: Webarchivierung, 02.04.2025, https://www.dnb.de/webarchiv, Stand: 11.02.2026, archiviert: https://web.archive.org/web/20260413150541/https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/sammlung_websites_node.html.

Fenner, Martin: Rogue Scholar is becoming a German non-profit organization, Front Matter, 03.11.2025, https://doi.org/10.53731/rftfk-qv692.

Gebeil, Sophie: WAAM. Web Archiving Aix Marseille, Pôle Bibliothèques et Archives de la MMSH, 03.10.2025, https://pba.mmsh.fr/?p=35306, Stand: 27.05.2026, archiviert: https://web.archive.org/web/20260625063303/https://pba.mmsh.fr/?p=35306.

Hannemann, Renate: Webarchivierung im BSZ mit Archive-It, Konferenzfolien zum 107. Deutschen Bibliothekartag, Berlin 2018, https://nbn-resolving.org/urn:nbn:de:0290-opus4-36116.

International Internet Preservation Consortium (IIPC): The WARC Format 1.1, https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/, Stand: 23.08.2026, archiviert: https://web.archive.org/web/20211218112357/https://iipc.github.io/warc-specifications/specifications/warc-format/warc-1.1/.

International Organization for Standardization (ISO): ISO 28500:2017. Information and documentation, WARC file format, 2017, https://www.iso.org/standard/68004.html, Stand: 23.8.2026.

Internet Archive: Archive-IT, https://archive-it.org/, Stand: 29.07.2026.

Liegmann, Hans; Rauber, Andreas: Kap. 17.9 Webarchivierung zur Langzeiterhaltung von Internet-Dokumenten, in: Neuroth, Heike; Oßwald, Achim; Scheffel, Regine u. a. (Hg.): nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Göttingen 2010. https://nbn-resolving.org/urn:nbn:de:0008-20100305365.

Müller, Sophie: Webarchivierung aus kulturwissenschaftlicher Perspektive, in: Dialog mit Bibliotheken 30 (2), 2018, S. 61–62.

Mutschler, Thomas: Zum Stand der kooperativen Webarchivierung in Thüringen, in: o-bib. Das offene Bibliotheksjournal 7 (4), 2020, S. 1–12. https://doi.org/10.5282/O-BIB/5632.

nestor: Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland e. V., https://langzeitarchivierung.de; Stand: 29.07.2026.

Passig, Kathrin: 3,5 Kilo Daten, in: Frankfurter Rundschau, 18.03.2022, I, Stand: 04.02.2026, archiviert: https://web.archive.org/web/20230404101116/http://www.fr.de/kultur/gesellschaft/kilo-daten-91419725.html.

Passig, Kathrin: Den Heuhaufen archivieren, Deutsches Literaturarchiv Marbach, März 2021, https://docs.google.com/presentation/d/1CQuVkXaDvsl0psAfkCSkzBcyitxsAAIXEjhLmQOIVaY, Stand: 08.08.2023.

Passig, Kathrin: Rucksack oder Rechenzentrum, DHd 2022 Kulturen des digitalen Gedächtnisses, 11.03.2022, https://www.dhd2022.de/closing-keynote/, Stand: 19.07.2023.

Schlesinger, Claus-Michael; Ulrich, Mona; Jorczick, Adam: Counting The loss of E-Lit works. How big are our time capsules?, Zenodo, 31.05.2022, https://doi.org/10.5281/zenodo.6587607.

Schlesinger, Claus-Michael; Woldering, Britta: „Den Heuhaufen archivieren": Archivierungsinitiative für deutschsprachige Tweets, in: Zeitschrift für Bibliothekswesen und Bibliographie 71 (4), 15.08.2024, S. 236–242, https://doi.org/10.3196/186429502471445.

Science Data Center for Literature: Webarchivierung für Forscher*innen, 18.07.2022, https://www.sdc4lit.de/veranstaltungen/default-title/, Stand: 24.07.2026, archiviert: https://perma.cc/SKC4-WYSP.

SocialMediaHistory. Geschichte auf Instagram und TikTok, https://smh.blogs.uni-hamburg.de, Stand: 24.07.2026, archiviert: https://perma.cc/G9VW-UGDT.

Ulrich, Mona: Einstieg in die Webarchivierung – mit Low Hanging Fruits und vielseitigen Herausforderungen, Valie Export Center Linz, https://www.valieexportcenter.at/einstieg-in-die-webarchivierung-mit-low-hanging-fruits-und-vielseitigen-herausforderungen, Stand: 23.07.2026.

Ulrich, Mona; Schlesinger, Claus-Michael: Workshop Webarchivierung. Clientseitige Webarchivierung, Zenodo, 14.10.2024, https://doi.org/10.5281/zenodo.13918959.

Web Archive Collection Zipped (WACZ), Webrecorder, 03.06.2021, https://specs.webrecorder.net/wacz/1.1.1/, Stand: 23.08.2026.

Webrecorder: Browsertrix, https://webrecorder.net/browsertrix, Stand: 24.08.2026.

Webrecorder: Browsertrix, Github, https://github.com/webrecorder/browsertrix, Stand: 11.06.2026.

Webrecorder: Web Archiving for All, https://webrecorder.net/, Stand: 18.08.2026.

Weigle, Michele: Some URLs are immortal, most are ephemeral, Web science and digital libraries, 20.09.2024, https://ws-dl.blogspot.com/2024/09/2024-09-20-some-urls-are-immortal-most.html, Stand: 18.04.2026, archiviert: https://web.archive.org/web/20260525183721/https://ws-dl.blogspot.com/2024/09/2024-09-20-some-urls-are-immortal-most.html.

Zeitschrift für Bibliothekswesen und Bibliographie 72 (2), 2025. https://doi.org/10.3196/1864-2950-2025-2.

Zeitschrift für Bibliothekswesen und Bibliographie 62 (3–4), 2015. https://doi.org/10.3196/1864-2950-2015-3-4.

Zentral- und Landesbibliothek Berlin: Webarchiv der Digitalen Landesbibliothek Berlin, https://digital.zlb.de/viewer/webarchiv/, Stand: 26.05.2026.

Zentral- und Landesbibliothek Berlin: collect and connect. Die Talkreihe der Landesbibliothek digital, https://digital.zlb.de/viewer/collect-and-connect/, Stand: 23.07.2026.