

Digitales kulturelles Erbe bewahren: Archivierung kommunaler Websites durch das Bibliothekservice-Zentrum Baden-Württemberg (BSZ)

Webarchivierung im BSZ

Das BSZ bearbeitet das Thema der Webarchivierung seit Beginn der 2000er-Jahre aktiv. In den Jahren 2004 bis 2016 bot das BSZ einen Dienst zur Archivierung von Webauftritten an, für Kommunalarchive unter dem Label SWBregio.

Dabei handelte es sich um eine eigenentwickelte Webanwendung, die eine Erfassung hierarchischer Metadatenstrukturen bereitstellte, die Festlegung differenzierter Crawlingparameter erlaubte und Elemente zur Jobsteuerung enthielt. Die Anwendung wurde von den teilnehmenden Archiven bedient, welche die Spiegelungen selbständig durchführten. Die Verantwortung für Entwicklung und Pflege der Anwendung, Datenhaltung und Support lag beim BSZ.

Als Crawler kam zunächst HTTrack zur Anwendung, 2012 wurde parallel dazu auch der vom Internet-Archiv entwickelte Crawler Heritrix (technisches Archivformat WARC) eingeführt.

Diese Eigenentwicklung wurde 2017 aufgegeben, da sie sich als sehr zunehmend aufwendig in Betrieb und Weiterentwicklung erwies. Damit wurde auch das Modell der eigenverantwortlich von den Archiven durchgeführten Spiegelungen abgelöst. Seither betreibt das BSZ den Dienst SWBregio unter Nutzung des Angebots von Archive-It¹, einer Einrichtung des Internet-Archivs², als einen zentralen Archivierungs-Service, welcher sich an kommunale, regionale und Kreisarchive richtet. Die in der Vorgänger-Applikation enthaltenen Archivkopien wurden vollständig nach Archive-It übernommen, die HTTrack-Daten dazu zuvor ins WARC-Format migriert.

Aktuell nutzen 19 Archive³ dieses Angebot – 17 davon aus Baden-Württemberg, 2 aus anderen Bundesländern.

Daneben führt das BSZ ebenfalls auf der Basis von Archive-It für die Saarländische Universitäts- und Landesbibliothek Saarbrücken (SULB) die Archivierung aller im Rahmen ihres Pflichtauftrags gesammelten saarländischen Websites durch.

Insgesamt werden aktuell rd. 670 Domains laufend durch das BSZ gespiegelt, zumeist mit ein- bis zweimal jährlich erzeugten Zeitschnitten, in der Regel fortlaufend über Jahre hinweg. Bei Bedarf findet jedoch auch ein zeitlich befristetes Event-Crawling mit kürzeren Spiegelungsintervallen oder ein Rettungs-Crawling (bei anstehender Einstellung oder Relaunch einer Website) statt.

Die Archivierung durch das BSZ erfolgt nach Beauftragung durch die teilnehmenden Archive, welche die Auswahl der Websites durchführen, die Spiegelungserlaubnisse bei der betreibenden Einrichtung einholen, sowie die Spiegelungsintervalle und die Modalitäten für die spätere Zugänglichkeit zu den Archivkopien festlegen. In der Regel sind diese öffentlich zugänglich, auf Wunsch können aber auch Einschränkungen eingeführt werden (z.B. Nutzung nur innerhalb des Archivs). Eine Bereitstellung zur ausschließlich internen Benutzung durch das Archivpersonal ist ebenfalls möglich.

Die Erschließung erfolgt in einrichtungsbezogenen Sammlungen für die jeweiligen Archive durch beschreibende Metadaten sowie eine Volltextindexierung der archivierten Ressourcen.

Die Spiegelungen werden physisch beim Internet Archiv gespeichert und von dort über gängige Webbrowser präsentiert. Parallel dazu werden sie vom BSZ regelmäßig als Sicherungskopie in Baden-Württembergischer Landesspeicherinfrastruktur gesichert.

Das BSZ übernimmt bei diesem Modell den gesamten technischen Workflow von der Erfassung und Pflege der Metadaten über die Analyse der zu spiegelnden Websites, die Parametrisierung des Crawlers, das laufende Jobmanagement bis hin zur inhaltlichen wie technischen Qualitätskontrolle für alle erzeugten Zeitschnitte.

Archiviert werden in SWBregio ausschließlich institutionelle Websites. Neben Webauftritten kommunaler Einrichtungen können dies beispielsweise auch Seiten von Vereinen, Initiativen und Kulturträgern sein. In welchem Umfang der Dienst genutzt wird, liegt vollständig im Ermessen der beauftragenden Archive - hier liegt die Bandbreite zwischen 1 und rd. 80 beauftragten Sites pro Archiv. Vom BSZ nicht für die Archivierung angenommen werden Social Media-Accounts (Facebook, Twitter, Instagram etc.), private und gewerbliche Webauftritte.

Bei der Durchführung der Spiegelungen liegt der Fokus auf einem möglichst hohen Grad der Authentizität der Archivkopien. Dies betrifft auf der einen Seite die optischen Gestaltungsmerkmale des Originals, das Look-and-feel der gespiegelten Site, ihre Funktionalität und Navigierbarkeit. Große Bedeutung wird zum anderen der Tiefe der Spiegelung und der Vollständigkeit und Verfügbarkeit ihrer Inhalte (HTML-Seiten, pdf-Dokumente, Grafiken, Videos u.a.) beigemessen. Die sehr dynamische technische Entwicklung im Bereich des Webdesigns stellt bei der Erreichung dieses Ziels eine stete Herausforderung dar.

[Beweggründe für die Archivierung kommunaler Websites](#)

Das World Wide Web ist heute über 30 Jahre alt und inzwischen aus dem täglichen Leben und der Welt der Information nicht mehr fortzudenken. Die meisten Städte und Gemeinden in Deutschland präsentieren sich, ihre Einrichtungen, Organe und Dienstleistungen im Netz, die Tendenz ist steigend. Das Angebot wandelt sich inhaltlich wie technisch laufend, neue, auch interaktive Formen der Kommunikation zwischen kommunalem Anbieter und Öffentlichkeit etablieren sich.

Das digitale Informationsangebot ersetzt und verdrängt dabei das konventionelle mehr und mehr: Analoge Informationsmittel befinden sich in vielen Bereichen der kommunalen Öffentlichkeitsarbeit auf dem Rückzug und verschwinden damit zunehmend als Sammelgut.

Stellten Webauftritte anfänglich noch ein rein ergänzendes Angebot dar, welches parallel zum klassischen Publizieren existierte und dieses lediglich digital abbildete, so hat sich dieses Verhältnis im Laufe der Jahre gravierend geändert. Websites haben sich zu eigenständigen Publikationsmedien weiterentwickelt – mit technischen Möglichkeiten der Gestaltung und Interaktion, für welche sich im analogen Bereich keine Entsprechung mehr findet. Die Sammlung und Erschließung dieser digitalen Quellen geschieht trotz der Entwicklungen der letzten Jahre auf diesem Gebiet in Deutschland noch nicht im selben Maße, wie das für ihre analogen Pendanten der Fall war und ist.

Das hat gravierende Folgen für die Kontinuität in der Dokumentation kommunalen Wirkens und erzeugt Brüche, welche sich nachträglich nicht mehr schließen lassen. Ein Großteil des an die Öffentlichkeit gerichteten Spektrums kommunaler Veröffentlichungen findet sich heute an dieser zentralen Stelle: gebündelt, in der Regel gut recherchier- und herunterladbar, rund um die Uhr erreichbar und weltweit für Jedermann verfügbar. Umso gravierender sind die Folgen ihres Verschwindens.

Websites und ihre Inhalte verschwinden

Websites sind in sich sehr dynamische Informationsplattformen, deren Content kontinuierlich erweitert, aktualisiert und ausgetauscht wird. Die Pflege wird in der Regel unter dem Aspekt der Aktualität betrieben, so dass als veraltet bewertete Inhalte gelöscht werden. Selbst bei einer kontinuierlich über Jahre hinweg verfügbaren Website findet also schleichend ein steter, aber in jedem Fall endgültiger Verlust von Inhalten statt. Nicht mehr verfügbare Inhalte sind dadurch zumindest für die Öffentlichkeit – und in vielen Fällen auch darüber hinaus – dauerhaft nicht mehr verfügbar.

Gravierender jedoch ist der Verlust ganzer Webauftritte: Tagtäglich werden aus den unterschiedlichsten Gründen Websites vollständig eingestellt. Einige Aspekte hierzu möchte ich im Folgenden aufgreifen.

Lokale Anlässe mit zeitlich begrenztem Charakter können zur Erstellung eigenständiger Websites für die Öffentlichkeit führen. Dies betrifft zum Beispiel kulturelle Events, Stadtjubiläen, Kommunalwahlen oder städtische und städtebauliche Projekte⁴ mit ihren Planungsprozessen und Maßnahmen zur Bürgerbeteiligung. Diese Web-Angebote werden oftmals nach Beendigung des Events zeitnah ersatzlos eingestellt, womit sämtliche veröffentlichte Inhalte verloren gehen.

Die Corona-Pandemie⁵ beispielsweise, welche auf kommunaler Ebene vielfach ein enormes digitales Informationsaufkommen generierte, ist heute im Bereich der Webarchivierung nur rudimentär abgebildet. Kommunen hatten ab 2020 Webangebote zur Pandemie, den wechselnden Bestimmungen und Maßnahmen, zum bürgerschaftlichen Engagement geschaltet, welche bereits heute nicht mehr existieren. Ihre Archivierung hätte es ermöglicht, den Umgang mit der Pandemie und das wechselvolle Geschehen zu dokumentieren und eines Tages nachzeichnen und analysieren zu können.

Ein weiterer Faktor, welcher zur Einstellung bestehender Websites führt, ist aktuell der Trend, bisher eigenständige Sites kommunaler Einrichtungen in das Hauptangebot der Kommunen einzugliedern⁶. Dies lässt sich beispielsweise im Zusammenhang mit Musikschulen, Museen, Bibliotheken, Theater, Feuerwehren, städtischen Betrieben etc. beobachten. Selbständige Websites mit ihrer bisher individuellen Form der Präsentation werden geschlossen. Damit einher gehen in der Regel eine starke inhaltliche Verschlankung des Angebots sowie der Verlust aller zuvor publizierten Inhalte.

Daneben führen eine Vielzahl weiterer Gründe kontinuierlich zur Aufgabe existierender Webangebote⁷. Durch welche Umstände auch immer die Einstellung einer Website begründet ist: In allen Fällen führt sie zum vollständigen Verlust der zuvor vorhandenen Inhalte, welche damit als mögliche historische Quellen verschwinden. In Archivkopien bleiben sie – erschlossen durch eine Volltextindexierung – dauerhaft auffindbar und verfügbar.

Fazit

Kommunale Websites sind Teil des Gedächtnisses und des Erbes einer Stadt und ihrer Bürgerinnen und Bürger. Sie bilden politische, gesellschaftliche, kulturelle, wirtschaftliche und soziale Prozesse ab. Als Quellen für die historische Betrachtung sind sie daher erhaltens- und bewahrenswerte Zeugnisse kommunaler Geschichte. Die Webarchivierung ermöglicht es, diese kommunalen Online-Inhalte systematisch und dauerhaft für die Zukunft zu bewahren.

¹ <https://archive-it.org/>

² <https://archive.org/>

³ <https://wiki.bsz-bw.de/display/WEBARCHIV/Referenzen>

⁴ https://wayback.archive-it.org/8530/*/http://www.buga2019.de/ Städtebaulicher Planungsprozess, Bürgerbeteiligung zur Bundesgartenschau 2019 in Heilbronn. Heute: <https://www.heilbronn.de/umwelt-mobilitaet/gruenes-heilbronn/buga-gelaende.html>

https://wayback.archive-it.org/8430/*/https://asylnetz-zollernalbkreis.de/ Zentrales Webangebot des Landkreises Zollernalb zum Thema Asyl. Website wurde eingestellt.

⁵ https://wayback.archive-it.org/8861/*/https://corona-engagiert.stuttgart.de/ Corona, bürgerschaftliches Engagement Stadt Stuttgart. Heute: <https://www.stuttgart.de/buergerengagement/>

⁶ https://wayback.archive-it.org/8507/*/https://naturkundemuseum.kassel.de/ Naturkundemuseum Kassel. Heute im städtischen Web-Angebot enthalten.

https://wayback.archive-it.org/8475/*/http://www.rego.gd/ E-Mobilität Schwäbisch Gmünd. Heute im städtischen Web-Angebot enthalten.

⁷ https://wayback.archive-it.org/9100/*/http://www.fvzmannheim.de/ Website wurde eingestellt.

https://wayback.archive-it.org/8507/*/http://www.friedrich-ebert-strasse.net/ Website wurde eingestellt.

https://wayback.archive-it.org/9172/*/http://www.feuerwehr.neu-ulm.de/ Website wurde eingestellt