

Themenschwerpunkt

Sammlungen von Websites an deutschen Regionalbibliotheken

Lars Jendral

Einleitung

DOI 10.1515/bd-2017-0050

Auf ihrer Sitzung 2002 in Eutin beschloss die AG Regionalbibliotheken, die Ausweitung des traditionell bestehenden Sammelauftrags für gedruckte Pflichtexemplare auf Netzpublikationen zu forcieren und auf die politische Agenda zu setzen. In der Folge wurde bei den zuständigen Landesministerien gesetzgeberischer Handlungsbedarf angemahnt, um die drohenden Erinnerungslücken im kulturellen Gedächtnis nicht anwachsen zu lassen. Zeitgleich entstanden die ersten Archiv- und Depotsysteme, um die Sammlung von Netzpublikationen technisch zu ermöglichen. Schnell stellte sich die Frage, welche Publikationsformen aus der unüberschaubaren Vielfalt der Online-Welt gesammelt werden sollen. Dazu waren vorrangig definitorische und typologische Klarstellungen vonnöten,¹ bei denen sich im Laufe der folgenden Jahre ein Entscheidungskonsens deutlich herauslesen ließ: je ähnlicher die Netzpublikationen bereits bekannten Publikationstypen aus der Druckwelt waren, desto größer die Bereitschaft der Regionalbibliotheken, diese als sammelwürdig einzuordnen. Dementsprechend kontrovers verliefen die Positionen zur Archivierung von Websites, die so wenig in tradierte Schemata passten. Während die Skeptiker auf das Massenproblem sowie die technischen Probleme einer Speicherung der Webangebote hinwiesen, stellten für die Befürworter inhaltlich hochwertige landeskundliche Websites klassisches Sammelgut einer Regionalbibliothek dar, das aufgrund seiner Schnelllebigkeit und eines fehlenden Pendantes in der Druckwelt besonderer Aufmerksamkeit bedürfte.

So begannen ab 2003 mit den neuen Archivservern für elektronische Pflicht in Rheinland-Pfalz, Baden-Württemberg und im Saarland auch die ersten Aktivitäten auf dem Gebiet der Webarchivierung. 2010 folgte dann die Bayerische Staatsbibliothek (BSB), die neben den landesbezogenen Websites auch eine Lösung für die im Rahmen der Sondersammelgebiete relevanten Webangebote

¹ Heidrun Wiesenmüller et al.: Auswahlkriterien für das Sammeln von Netzpublikationen. In: Bibliotheksdienst 38 (2004), S. 1423–1444.

im Focus hatte. Im Rahmen einer Kooperation nutzt seit 2013 auch die Staats- und Universitätsbibliothek Hamburg die an der BSB aufgebaute Infrastruktur für die Sammlung von Websites.

Der Entwicklung wie dem aktuellen Stand dieser fünf Webarchivierungsvorhaben an deutschen Regionalbibliotheken im Frühjahr 2017 widmet sich dieses Themenschwerpunktheft des Bibliotheksdienstes.² Welche Erfahrungen wurden in den letzten Jahren gesammelt? Lassen sich Problembereiche bei der Sammlung von Websites identifizieren? Und nicht zuletzt: welchen Beitrag können die bestehenden regionalen Webarchive zur „Sammlung des deutschen Netzes“ bereits jetzt leisten? Neue Aktualität ist dem Thema seit dem August 2016 erwachsen: auf einem gemeinsamen Workshop von Deutscher Nationalbibliothek (DNB) und AG Regionalbibliotheken zum „digitalen Sammelauftrag“ in Frankfurt bestand nahezu einhellig Bereitschaft, auf dem Gebiet verstärkt zu kooperieren. Wenn-gleich das Thema mit digitaler Sammlung denkbar weit gefasst war, stand auf Seiten der DNB der Kooperationswunsch auf dem Felde der Websitesammlung deutlich im Vordergrund. Während bereits aktive Landesbibliotheken ihre Projekte kurz vorstellten, zeigten sich bislang zurückhaltende Bibliotheken bereit, das Thema für sich erneut zu prüfen – eine Haltung, die auf der Herbstsitzung der AG Regionalbibliotheken bestärkt wurde.

Rechtliche Grundlagen und Sammlungsschwerpunkte

Die bereits existierenden Websitesammlungen sind durch die Pflichtabgaberegungen in den entsprechenden Pflichtexemplar-, Presse-, Medien- oder Bibliotheksgesetzen und/oder durch die Erlasse zur Abgabe amtlicher Veröffentlichungen legitimiert, in einigen Fällen ist jedoch auch eine Genehmigungseinholung notwendig. Dementsprechend stellt die Sammlung von Websites der Ministerien, Behörden oder kulturellen Institutionen der Bundesländer einen wichtigen Schwerpunkt bei allen vorgestellten Projekten dar. Ebenso einhellig werden thematische Webangebote zum Land oder einzelnen Regionen als zentrales Sammelgut benannt.

² Vgl. hierzu auch ZfBB 62 (2016), H.3/4: Webarchivierung in Bibliotheken, hrsg. von Reinhard Altenhöner und Achim Oßwald, in diesem Kontext besonders: Tobias Beinert/Astrid Schoger: Vernachlässigte Pflicht oder Sammlung aus Leidenschaft? Zum Stand der Webarchivierung in deutschen Bibliotheken, S. 172–183.

Technische Lösungen

Übereinstimmend haben sich alle auf dem Gebiet aktiven Regionalbibliotheken zur technischen Ausgestaltung ihrer Archivierungsvorhaben kompetente Partner gesucht, in diesen Fällen die jeweiligen Verbundzentralen. Einzig die BSB betreibt ihr Webarchiv in Eigenregie und tritt ihrerseits als Partner für eine andere Bibliothek auf.

Das Bibliotheksservicezentrum (BSZ) betreut für die Landesbibliotheken und das Landesarchiv das Baden-Württembergische Online-Archiv (BOA) und Saardok, den Archivserver der Saarländischen Universitäts- und Landesbibliothek Saarbrücken. Zum Einsatz kommt dabei SWBcontent, eine JAVA-basierte browsergestützte Workflow-Lösung auf Open-Source-Basis.

Das Hochschulbibliothekszentrum Nordrhein-Westfalen (hbz) entwickelt für das Landesbibliothekszentrum Rheinland-Pfalz den Archivserver edoweb mit REGAL, das auf Basis der Open-Source-Software Fedora Commons und des CMS Drupal funktioniert.

Die BSB und die SUB Hamburg nutzen für ihre Sammlung das Web Curator Tool (WCT), die eigentliche Archivierung erfolgt zumindest in der BSB über das Exlibris-Produkt Rosetta.

Allen gemeinsam ist die integrierte Nutzung des Webcrawlers Heritrix, die gesammelten Inhalte werden in WARC-Dateien abgespeichert und mittels der Viewer Wayback und OpenWayback zur Anzeige gebracht. Damit bedienen sich die beschriebenen Projekte alle mittlerweile der ursprünglich vom in San Francisco ansässigen Internet Archive³ entwickelten Software und Standards. Nicht zuletzt deshalb ist das BSZ zu dem Entschluss gekommen, die Website-Archivierung von BOA an das Internet Archive auszulagern und dort ein eigenes Depot betreiben zu lassen.

Perspektiven

In fünf der sechzehn Bundesländer betreiben Regionalbibliotheken selektive Webarchive mit vergleichbaren, wenngleich überschaubaren Sammelschwerpunkten, die indes formal gut erschlossen sind. Inhaltlich verfolgen sie dabei auf Länderebene einen ähnlichen Ansatz, den auch die DNB auf Bundesebene beschreitet. Sollte der Plan, eine kooperative Sammlung des „deutschen Netzes“

³ <https://archive.org/> [Zugriff: 21.03.2017].

auf Bibliotheksebene Gestalt annehmen, sind grundlegende Entscheidungen zu treffen.

Bislang abwartende Regionalbibliotheken müssten die entsprechende technische Infrastruktur beschaffen bzw. ihre bislang für die E-Pflicht genutzten Systeme entsprechend erweitern. Da es existierende technische Plattformen gibt und entsprechende Dienstleister vorhanden sind, besteht an dieser Stelle nicht das Haupthindernis. Auch die in einigen Bundesländern noch zu schaffenden oder anzupassenden rechtlichen Grundlagen sind vermutlich mit der nötigen Beharrlichkeit und Geduld zu erwirken. Von entscheidender Bedeutung wird jedoch die Bereitschaft der zuständigen Ministerien und Landesparlamente sein, die entsprechenden Ressourcen zur Verfügung zu stellen. Denn selbst bei weitgehend automatisierten Geschäftsgängen ist der personelle Aufwand für Identifikation, Sammlung und Qualitätskontrolle nicht zu gering zu veranschlagen. Außerdem ist der Speicherbedarf gerade beim automatisierten Wiedereinsammeln der bereits einmal archivierten Websites in regelmäßigem Turnus rasch wachsend und stellt damit ebenfalls einen Kostenfaktor dar. Daher muss ein Schwerpunkt der Bemühungen darin bestehen, der Politik den Sinn und die Bedeutung dieses Archivierungsvorhabens einleuchtend zu vermitteln.

Die DNB muss ihrerseits über ihre selektive Sammlung von Websites hinaus verstärkt auf dem Feld des domain-crawls aktiv werden, d. h. dem Versuch, in einem angesichts des Umfangs wahrscheinlich Monate dauernden Vorgang, sämtliche Websites der domain .de einzusammeln. Allein die Nationalbibliothek hat vom Aufgabenzuschnitt die Möglichkeit, ein nach Bundesländern nun einmal nicht näher differenziertes deutsches Webangebot in seiner Gesamtheit zu harvesten. Bislang hat die DNB erst einmal einen domain-crawl mit Hilfe eines Dienstleisters durchgeführt. Dieses in anderen Staaten schon regulär zum Einsatz kommende Verfahren lieferte eine wichtige, allerdings nur über die URL erschlossene Basisdokumentation des Web, die durch besser erschlossene, qualitativ bessere Teilsammlungen der Regionalbibliotheken angereichert werden könnte. Zukünftige Kooperationen werden hoffentlich durch den geplanten Entwurf für eine Anpassung der Urheberrechts und der damit einhergehenden Änderungen des Gesetzes über die DNB erleichtert.⁴

Unabhängig davon, wie genau eine Aufgabenverteilung, Bereitstellung und Finanzierung einer solchen Zusammenarbeit ausgestaltet wird, müsste über die Bibliothekswelt hinaus ein Netz denkbar weit zu fassender Gedächtnisinstitutio-

⁴ http://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/DE/UrhWissG.html;jsessionid=A020CA352CC68D6270CE3E026FC45AF71_cid324 [Zugriff: 21.03.2017].

nen aktiv mitsammeln, um bei der Größe der Aufgabe ein den Erwartungen entsprechendes, gemeinsames Webarchiv für Deutschland zu schaffen.