

Praxisbericht zur Langzeitarchivierung der Intranet-Website „Straßennamendatenbank“ im Stadtarchiv Bochum

von Annett Schreiber

Einführung

Das Stadtarchiv Bochum befasst sich innerhalb eines ämterübergreifenden Arbeitskreises (GeLaTi = **G**esetzliche **L**angzeitarchivierung und **t**echnische **I**nfrastruktur), der vom Amt für Personalmanagement, Informationstechnologie und Organisation geleitet wird, seit 2012 mit konkreten Maßnahmen zum Langzeiterhalt von Daten städtischer Dienststellen.¹ Eines der wichtigsten Projektziele neben der Beschaffung einer Langzeitarchivierungslösung war die Erstellung der Bochumer Fachverfahrensliste.² Die dort als vorläufig oder als abgeschlossen gelisteten Bewertungsentscheidungen ermöglichten im Jahr 2016 konkrete Gespräche mit Dienststellen über die Aussonderung und Abgabe archivreifer und archivwürdiger Daten sowie die Priorisierung dieser Aussonderungsprojekte. Zeitgleich erfragten das LWL-Archivamt und die LWL.IT den Bedarf an Projekten und Schnittstellen für alle DiPS.kommunal³-Mandanten mithilfe einer Fachverfahrens-Checkliste. Diese Checkliste ermöglichte den DiPS-Betreibereinrichtungen und den Archiven die Planungen von Aussonderungsschnittstellen und Kooperationsmöglichkeiten innerhalb der DiPS-Nutzer⁴. Es ging für das Stadtarchiv Bochum zu diesem Zeitpunkt jedoch (noch) nicht darum, bereits ein Langzeitarchivierungskonzept für das eigene Archiv oder bestimmte Datentypen vorzulegen oder umzusetzen – es ging darum, mit den zur Verfügung stehenden personellen, organisatorischen, technischen und finanziellen Ressourcen realistische Ziele zu erreichen und diese mit anderen teilnehmenden Kommunen zu diskutieren.

In Bochum wurde u. a. die HTML-Version einer Straßennamendatenbank (eine im Intranet zugängliche Website),

die von einem Mitarbeiter des Amtes für Geoinformation, Liegenschaften und Kataster (Amt 62) Mitte der 2000er-Jahre in Eigenregie und als freiwillige zusätzliche Leistung entwickelt wurde, für die Checkliste benannt. Ursprung dieser Website und der damit verknüpften Access-Datenbank ist die im Jahr 1992 vom selben Amt herausgegebene Publikation „Bochumer Straßennamen: Herkunft und Deutung“⁵, die sich u. a. der Quellen des Stadtarchivs bedient. Die Website, die jeweils nach Beschlussfassung im zuständigen Ausschuss bzw. im Rat aktualisiert wird, umfasst Erläuterungen zur Be- und Umbenennung aller Stra-

1 Der AK wurde durch einen Beschluss des Verwaltungsvorstandes mit dem Ziel gegründet, alle Maßnahmen zu erarbeiten und einzuleiten, die die Langzeitarchivierung von Daten im Stadtarchiv sicherstellen. Das Erreichen des Projektziels wurde nicht nur durch die Beschaffung einer Langzeitarchivierungslösung definiert, sondern vor allem durch die prototypische Aussonderung aus zwei bis drei Fachverfahren. Die dabei etablierten Routinen sollten das Thema Langzeitarchivierung innerhalb der gesamten Stadtverwaltung wahrnehmbar machen. Vgl. dazu Annett Schreiber, DiPS in der Praxis. Einstieg in die Elektronische Langzeitarchivierung im Stadtarchiv Bochum, vgl. <https://archivamt.hypothesen.org/11062>, [Stand: 23.09.2020, gilt ebenfalls für alle nachfolgenden Hinweise auf Internetseiten].

2 Annett Fercho, Erfassung und Bewertung der elektronischen Fachverfahren durch das Stadtarchiv Bochum im Rahmen des Projekts „Erarbeitung einer Konzeption zur Sicherstellung einer reversionssicheren und vertrauenswürdigen digitalen Langzeitarchivierung bei der Stadt Bochum“, 2015 (interner Bericht); Annett Fercho/Stefan Pätzold, Die Erfassung und Bewertung der elektronischen Fachverfahren der Stadt Bochum. Ein Werkstattbericht, in: *Archivpflege in Westfalen-Lippe* 81 (2014), S. 40 ff.

3 Vgl. <https://www.stadt-koeln.de/artikel/62898/index.html>.

4 Der DiPS-Anwenderkreis besteht aus den kommunalen DiPS-Mandaten und den DiPS-Betreibern, der LWL.IT und dem Amt für Informationsverarbeitung der Stadt Köln. Der AK trifft sich in der Regel zwei Mal jährlich, um über alle im Zusammenhang mit der Langzeitarchivierung auftretenden Fragen zu beraten und zu informieren.

5 Stadt Bochum, Vermessungs- und Katasteramt (Hrsg.), *Bochumer Straßennamen: Herkunft und Deutung*, Bochum 1992.

Stadt Bochum

Straßenlisten

- Sortierung nach:
 - Name
 - Gemarkung (siehe Übersicht)
 - Gebietskörperschaft
- Motivgruppe
- Namensverbund
- Ehem. Straßennamen
- Zeitliche Sortierung
- Sonstige inoffizielle Bezeichnungen

• Statistik

Allgemeines

- Startseite

- Vorwort
- Einleitung
- Anleitung
- Verwendete Abkürzungen
- Literaturverzeichnis
- Weiterführende Literatur
- Bildnachweis
- Impressum

Bochumer Straßennamen - Herkunft und Deutung

Wittener Straße

Straßenname: Wittener Straße

Benannt: 1874

ehemalige Namen: Weststraße (1979 tlw. einbez.), Chaussee von Bochum nach Witten

Gemarkung: Bochum, Altenbochum, Laer, Querenburg, Langendreer, Wiemelhausen

Verlauf: Kurt-Schumacher-Platz bis Universitätsstraße

Planquadrat (ETRS): J 7, K 7, 8, L, M 8, N 8, 9, O 8, 9

Postleitzahl: 44789 (von 1 bis 87), 44803 (von 89 bis 291), 44803 (von 293 bis 361), 44892 (von 529 bis Ende), 44789 (von 2 bis 134), 44803 (von 136 bis 260), 44803 (von 262 bis 270), 44892 (von 512 bis Ende)

Straßenschlüssel: 5905

Gebietskörperschaft: Bundesstraße B 226 (Kurt-Schumacher-Platz - Auf-/Abfahrt A44), Kreisstraße K 8 (Auf-/Abfahrt A44 - Universitätsstraße)

Motivgruppe: Nachbarstadt

Erläuterung: Diese Straße war schon in sehr frühen Zeiten der Verbindungsweg zwischen Bochum und der ehemaligen kreisfreien Stadt Witten (seit 1975 zum Ennepe-Ruhr-Kreis gehörend). Vor 1790 waren die Hauptausfallstraßen oft schlammige, mit Pferdefuhrwerken kaum befahrbare Wege. Um 1790 begann man mit dem Ausbau der sogenannten Kunststraßen oder Chausseen. "Die beiden Männer, denen die Mark die schönen neuen Kunststraßen verdankte, waren der Vorsitzende der Regierung der Grafschaft Mark Kammerdirektor Frh. vom Stein und der Minister v. Heinitz; letzterer bereiste 1787 selbst die Wegestrecke und bestimmte den Bau der Straße, welche von Essen über Steele, Bochum, Kringeldanz nach Hörde, Unna, Soest sowie über Hagen nach dem Sauerlande führen sollte. Am 13. Okt. 1789 konnte Bergassessor Grolmann dem Kriegsrat v. Bernuth in Hamm melden, daß die Wittener Chaussee von Bochum bis an Hiltrops Haus fertig gestellt sei; von da ab konnte alles Fuhrwerk gegen Erlegung eines Wegegeldes auf dieser Strecke verkehren. 1791 erst wurde das Stück jener Chaussee vor dem Butenbergsthore bis zum Schlagbaum an der Lohbergstraße hergerichtet." Quelle: Darpe, F., Geschichte der Stadt Bochum, Bochum, 1894



Bild zum Vergrößern anklicken

Wittener Straße

Beispielseite mit Navigationsleiste, Version 10/2013. In der Wittener Straße befindet sich das Stadtarchiv.

Ben Bochums und bietet zusätzlich diverse Sortiermechanismen.

Die HTML-Version wurde aus mehreren Gründen als Pilotprojekt für den GeLaTi-Arbeitskreis sowie für das Stadtarchiv Bochum als DiPS-Mandanten angegeben:

Das Stadtarchiv Bochum hatte sich schon vor längerer Zeit dazu entschieden, als Pilotkunde bei der Umsetzung von Projekten, die auch andere DiPS-Mandaten und Kommunalarchive betreffen, mitzuwirken. Da alle anderen Projekte innerhalb des DiPS-Anwenderkreises um „klassische“ Behördenaussonderungen aus weit verbreiteten kommunalen Anwendungen⁶ und den Entwurf kompatibler Schnittstellen für diese kreisten, sei die Beschäftigung mit anderen Datentypen wie Websites sehr erwünscht, teilte das LWL-Archivamt für Westfalen damals mit.

Da bereits zahlreiche Kooperationsprojekte im Bereich der Nutzung und Bereitstellung sowie Digitalisierung von Karten und Plänen des Stadtarchivs mit dem Amt 62 bestanden, sicherte der Ersteller der Datenbank für dieses Projekt formlos unbürokratische, aktive und zeitnahe Unterstützung zu. Letztendlich musste für diese Quelle auch die Datenschutzbeauftragte nicht bemüht werden, da diese Website bereits veröffentlichtes Material sowie keinerlei schützenswerte personenbezogene Daten enthält.

Beschreibung der Quelle

Straßennamendatenbanken werden im kommunalen Sektor entweder vom Stadtarchiv oder aber von anderen dafür prädestinierten Fachdienststellen erstellt oder betrieben.⁷ Sie sind ein niedrigschwelliges Angebot für Bürgerinnen und Bürger, die Informationen zur Namensherkunft und -vergabe einer bestimmten Straße benötigen. Zudem sind sie ein probates Medium, um z. B. die Straßenumbenennungen während und nach der NS-Zeit nachzuvollziehen.

(Bochumer) Bürgerinnen und Bürgern steht die HTML-Straßennamendatenbank als mindestens einmal jährlich aktualisierte CD-ROM-Version zum Kauf zur Verfügung; für den Dienstgebrauch ist sie ausschließlich im Intranet zugänglich. Eine über den Browser frei abrufbare Online-Version existiert nicht. Zur Beantwortung von Einzelfragen der Lesesaalbenutzer besaß das Stadtarchiv unabhängig vom Intranetzugriff bereits eine lokal gespeicherte und über einen Browser nutzbare Version. Als Besitzer der Daten beantwortet(e) jedoch überwiegend das Amt 62 regelmäßig Bürgeranfragen mit dem aktuellen Datenbestand.

Die Website stellt zahlreiche Sortiermöglichkeiten sowie statistische Informationen zur Verfügung und wird laufend inhaltlich und strukturell angepasst.⁸ Ihren Hauptzweck erfüllt sie durch die alphabetische Anzeige aller Straßennamen (ca. 1.900) mit entsprechenden Erläuterungen (wie oben); ca. 500 Datensätze enthalten ein Foto. Die Straßen können aber auch als zur Gemarkung (Stadtbezirk) zugehörig oder nach der Gebietskörperschaft (Unterhaltsbehörde der Straße) angezeigt werden. Die Kategorien Person, Gebäude, Geografischer Begriff oder Flora und Fauna, Mythen- und Sagenwelt bezeichnen die sog. Motivgruppen mit diversen Unterbegriffen. Die Wittener Straße gehört demnach zum Oberbegriff „Geografischer Begriff“, der weiter in „Nachbarstadt“ aufgegliedert wurde. Sachthematisch sortiert sind die Straßen außerdem über den Link „Namensverbund“, der die Benennung von komplet-

6 Hier sind vorrangig die Aussonderungen aus dem Personenstandsverfahren Autista, aus MIGEWA und sonstigen Gewereregisterverfahren sowie aus Ratsinformationssystemen zu nennen.

7 Beispiele hierfür sind: <https://www.iserlohn.de/aktuelles/strassenverzeichnis/> sowie <https://www.marchivum.de/de/strassennamen>.

8 In der Version von 2020 steht auch die Sortierung nach Geschlecht (Anzahl der nach Frauen, Männern und Intersexuellen vergebenen Straßennamen) zur Verfügung.

ten (oder komplett neu gebauten) Wohnvierteln z. B. nach bestimmten Vogelarten oder Naturwissenschaftlern dokumentiert. Zudem lassen sich die Straßen nach Ihrem Benennungsjahr abrufen, was insbesondere für Zwecke zur Aufarbeitung der NS-Zeit sowie bei Eingemeindungen und für den Städtezusammenschluss Bochum und Wattenscheid im Jahr 1975 von Bedeutung ist. Für Archivrecherchen hat die Gegenüberstellung (Konkordanz) alter und neuer Straßenbezeichnungen sowie eine Liste inoffizieller Bezeichnungen, die aus dem lokalen Sprachgebrauch stammen, besondere Relevanz.

Der Vollständigkeit halber sei hier erwähnt, dass ergänzend zu dieser Quelle im Amt 62 eine Beschlussliste des zuständigen Ausschusses zu allen Straßenbe- und -umbenennungen existiert. Zudem wird für Bochum seit ca. 1950 je Straße eine analoge Akte geführt (für Wattenscheid ab dem Städtezusammenschluss 1975), die zumeist einen Lageplan, die Vorschlagsliste zur jeweiligen Benennungsentcheidung, Zeitungsartikel, Widmungen, Leserbriefe und Material des Stadtarchivs enthält. Die Unterlagen sind langfristig nicht abgabereif, weil sie dauerhaft zur Aufgaben erledigung in der Dienststelle benötigt werden, können jedoch bei berechtigtem Interesse zumindest durch Dienstkräfte der Stadtverwaltung genutzt werden.

Das abgebende Amt hat ein gesteigertes Interesse daran, alle Vorversionen von den dortigen Servern zu löschen und jeweils nur die aktuelle Version für interne Zwecke und die Bürgerauskunft vorzuhalten.

Suche nach dem geeigneten Tool

Das Archivierungsziel umfasste den vollständigen Erhalt der Website mit allen Unterseiten sowie Funktionalitäten (Erhalt des Look & Feel) als ein Archivobjekt. Vorkenntnisse auf dem Gebiet der Langzeitarchivierung von Websites bestanden weder im Stadtarchiv noch in der IT-Abteilung oder dem Amt 62. Als hilfreich für dieses Projekt erwiesen sich jedoch die im Stadtarchiv vorhandenen HTML-Grundkenntnisse.

Seit 2016 erfolgten mehrere Besprechungen mit dem Ersteller der Website sowie dem LWL-Archivamt und der LWL-IT, die Überlegungen zum möglichst zweckmäßigen und ressourcenorientierten Einsatz von gängigen Internetarchivierungsmethoden⁹ zum Gegenstand hatten.

Nach Übermittlung der Version 10/2013 durch das abgebende Amt über ein sicheres Datenaustauschprogramm im Herbst 2017 wurden die Daten lokal im Stadtarchiv gesichert. Zunächst wurde die HTML-Version als Website inhaltlich vollständig analysiert; die Ergebnisse wurden im Kontext zu signifikanten Eigenschaften dokumentiert. Letztendlich wurde die Website in Gänze als archivwürdig eingestuft, auch wenn es von Anfang an attraktiv erschien, nur die Webseiten zu den Straßennamen-Erläuterungen zu erhalten, bedenkt man das hier offensichtliche Nutzungsziel „Feststellung der Namensherkunft einer Straße“ im Lesesaal. Zusätzlich erschien hier der klassische Text mit Inhalt (die intellektuell verstehbare Information) als die vor-

rangig zu erhaltende signifikante Eigenschaft des Objekts, nicht die Verlinkung als solche oder die Art der Darstellung. Die Access-Datenbank wurde aus diversen Gründen nicht zur Langzeitarchivierung vorgesehen.

Da es sich trotz aller Anpassungen nach wie vor um eine statische Website im Stil der 1990er-Jahre ohne dynamische Inhalte handelt, erfolgte der erste Versuch zur Archivierung der Straßennamen-Einzelseiten in Form von PDF-Dateien mithilfe der Funktion „Microsoft Print to PDF“ des Firefox-Browsers. Diese Methode war trotz der Beschränkung auf die Seiten zu reinen Straßennamenerläuterungen zwar aufwändig, führte jedoch zur gewünschten Extrahierung der Text-Informationen. Probleme mit einer durchweg konsistenten Dateibenennung sowie der Nachbildung einer Seitennavigation ließen die Vorgehensweise letztendlich als nicht durchführbar erscheinen.

Das Programm Pablo¹⁰, das eine Bilddatei der Website sowie eine METS-XML-Datei zur Darstellung der Verknüpfungen erzeugt, eignete sich – obwohl aus Perspektive der Langzeiterhaltung favorisiert – nicht, da es konkrete URLs (www.beispiel.de) für das Crawlen benötigt. In Bochum werden die Daten jedoch in Form eines Pfadnamens über ein lokales internes Laufwerk auf einem Webserver abgefragt, mit dem die Software nicht umgehen kann. Pablo kann zwar Websites, die über einen Login im Intra- oder Internet erreichbar sind, crawlen, technische Schwierigkeiten in der Soft- und Hardwareumgebung sowie in der Firewall der Stadt Bochum standen diesem Versuch, der unterstützt durch die Firma startext und das LWL-Archivamt begleitet wurde, jedoch entgegen.¹¹

Der Prozess erfuhr eine entscheidende Wende durch einen Termin mit den für die Langzeitarchivierung zuständigen Kollegen im Landesarchiv NRW in Münster (LAV) im Januar 2019. Sie hatten sich freundlicherweise bereit erklärt, sowohl dem LWL-Archivamt als auch dem Stadtarchiv Bochum die bis dato gemachten Erfahrungen mit der Webarchivierung vorzustellen, darunter Heritrix¹², HTTrack¹³,

9 Bundeskonferenz der Kommunalarchive beim Deutschen Städtetag (Hrsg.), Empfehlung: Speicherung von kommunalen Webseiten – Teil 2: Technik, Beschluss der BKK vom 14.04.2011 in München, vgl. <http://www.bundeskonferenz-kommunalarchive.de/empfehlungen.html>.

10 Vgl. <https://www.startext.de/produkte/pablo>.

11 Vgl. dazu Michael Korn, Vortrag „Werkstattbericht. Erfahrungen des Stadtarchiv Sankt Augustin bei der Archivierung von stadtrelevanten Websites mit PABLO“ bei der Sitzung der AG Kommunalarchive im Rhein-Sieg-Kreis am 07.11.2017 in Rheinbach (unveröffentlicht).

12 Heritrix ist ein weit verbreiteter und kostenfrei nutzbarer Webcrawler, der vom Internet Archive und den Nordic National Libraries entwickelt wurde und einen sog. WARC-Container erzeugt. Im Wesentlichen funktioniert er über die Kommandozeile. Vgl. <https://github.com/internetarchive/heritrix3>.

13 HTTrack ist ein kostenfrei verwendbares offline Browser Tool zum Download einer Website auf ein lokales Verzeichnis unter Beibehaltung der Linkstruktur. Vgl. <https://www.httrack.com/>; Angela Ullmann, ARNE – Archivierung von Netzressourcen des Deutschen Bundestags, in: H. Neuroth/A. Oßwald u. a. (Hrsg.), nestor Handbuch. Eine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3, 2010, Kapitel 18.4, vgl. https://www.langzeitarchivierung.de/Webs/nestor/DE/Publikationen/nestor_Handbuecher/nestor_handbuecher_node.html?sessionid=38FEC0E5535EF3853133EBDE74F52616.internet562.

Webrecorder¹⁴ und Wget¹⁵. Zudem erläuterten sie Bewertungsmaßnahmen, Vorgehensweisen für eigene Projekte innerhalb der Landesverwaltung NRW¹⁶, Formatfragen, notwendige Konfigurationen der von ihnen geprüften und sonstigen verfügbaren Tools, die Metadatenvergabe sowie Nutzungsmöglichkeiten archivierter WARC-Dateien¹⁷. Es wurde außerdem darüber beraten, die zugrundeliegende Access-Datenbank im SIARD¹⁸- oder im CSV-Format¹⁹ zu archivieren. „Doch das ist in diesem Fall, als würde man mit Kanonen auf Spatzen schießen“, kommentierte der Kollege des LAV das Vorhaben zu den beiden letztgenannten Optionen. Der Installations- und Umsetzungsaufwand sowie die Rekrutierung notwendiger technischer Ressourcen in der Stadt Bochum standen in keinem Verhältnis zum Projekt als Ganzem. Der Sachbearbeiter im Amt 62 hatte bereits die Archivierung der Access-Datenbank im CSV-Format erläutert – am Ende sahen wir wegen der Schwierigkeiten im Nutzungsprozess der Dateien und des hohen Erstellungsaufwandes davon ab.

Nach diesem Termin wurde dem Stadtarchiv klar, dass keines der zur Verfügung stehenden Programme ad hoc oder testweise mit den vorhandenen technischen Mitteln und Kenntnissen eingesetzt werden konnte, weder im Stadtarchiv selbst noch in der IT-Abteilung. Es bedurfte für jedwede Maßnahme umfangreicher Fachkenntnisse und einer dauerhaften Investition der Ressource Zeit in dieses Unterfangen. Es schien, als würden bei aller Anstrengung, diese Aufgabe anzugehen, die anderen Aussonderungsprojekte aus Fachverfahren (dauerhaft) hinstehen müssen.

Technische Umsetzung – die Lösung

Die LWL.IT hat sich im Jahr 2019 dafür entschieden, im Rahmen ihrer Serviceleistungen für Kommunalarchive (hier Bochum) die Archivierung der HTML-Straßennamendatenbank mit Wget zu testen. Das Tool war dort bereits standardmäßig auf einem Linux-Server installiert und war direkt einsatzbereit.

Um die browserbasierte Straßennamendatenbank mit einem Crawler archivieren zu können, wurde sie in der LWL.IT auf einem Suse Linux Enterprise Server über die Python-Komponente SimpleHTTPServer als Website zugänglich gemacht. Die Website wurde vor der Archivierung auf Vollständigkeit geprüft und danach mit Wget in einen WARC-Container archiviert. Dieser erzeugte Container wurde anschließend mit JHOVE²⁰ auf Validität überprüft und zeigte dabei keine Auffälligkeiten.

WARC ist ein Containerformat, das alle Dateiformate und Ressourcen, die zur Darstellung der Ursprungswebseite notwendig sind, zusammen mit Links und Metadaten abspeichert. Der ISO-Standard wird trotz seiner weiten Verbreitung unter langzeitarchivierenden Einrichtungen²¹ und der Nutzung in der Internet Archive Way Back Machine²² kontrovers diskutiert. Die DiPS.kommunal-Betreiber Stadt Köln und LWL haben sich trotz der Herausforderungen in Bezug auf dessen Langzeiterhaltung entschieden, das

WARC-Format als Archivformat für DiPS.kommunal zuzulassen, da es offen dokumentiert, lizenzfrei und weit verbreitet ist. Damit entspricht es zumindest formal einigen Kriterien für die Verwendung von Langzeitarchivformaten.

Für die Nutzung der WARC-Datei wurde der Webrecorder Player²³ ausgewählt. Eine Qualitätskontrolle der Daten sowie der Verlinkungen durch das Stadtarchiv in diesem lokal installierten Offline-Browser ergab eine 99%ige Übereinstimmung mit der Original-Ansicht im Intranet (keine Verluste an der Inhaltsinformation). Das Archivierungsziel wurde demnach vollständig erreicht.

Einlieferung über den PIT.plus und Anbindung an Faust

Die WARC-Datei konnte nach einer Anpassung der Konfigurationsdatei für die Formaterkennung durch den PIT.plus, dem Ingesttool für unstrukturierte Daten als Vorstufe für die Einlieferung nach DiPS.kommunal, verarbeitet und mit technischen, inhaltlichen und archivischen Metadaten versehen werden. Diese Metadaten wurden z. T. von der abgebenden Stelle, überwiegend jedoch vom Stadtarchiv vorher definiert oder waren projektimmanent. Nach der Archivierung in DiPS.kommunal wurde die VZS-XML, die das Langzeitarchiv für die Weiterverarbeitung von rudimentären Erschließungsdaten im Erschließungsprogramm erstellt,

14 Das Tool erlaubt das Mitschneiden der Anlinkreihenfolge beim Besuch von Webseiten in Echtzeit; die Daten werden ebenfalls in einem WARC-Container abgelegt. Vgl. <https://webrecorder.io/>.

15 Wget ist ein freies Programm, mit dem man direkt aus einem Terminal Dateien von FTP-, HTTP- oder HTTPS-Servern herunterladen kann. Vgl. <https://www.gnu.org/software/wget/>.

16 Valentin Kramer, Konzept zur Archivierung des Webauftritts des MFKJKS NRW, Transferarbeit Archivschule Marburg, 2017.

17 Das WARC (Web ARChive) Format speichert die komplette Website mit allen verfügbaren Ressourcen und Dateiformaten sowie dazugehörigen Verlinkungen in einen Container. Vgl. <https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml> sowie <https://kost-ceco.ch/cms/warc.html?highlight=WARC>; ISO 28500:2017, Information and documentation – WARC file format, siehe <https://www.iso.org/standard/68004.html>.

18 SIARD ist eine vom Schweizer Bundesarchiv entwickelte Software zur einfachen Archivierung von relationalen Datenbanken. SIARD Suite extrahiert Inhalte aus relationalen Datenbanken und speichert sie im für die Archivierung geeigneten SIARD-Format. Dieses Dateiformat ermöglicht es, die Daten im Archiv langfristig und unabhängig von der ursprünglichen Software aufzubewahren. Bei Bedarf lassen sich die Daten in eine neue Datenbank laden. Dadurch können sie unabhängig von der ursprünglichen Datenbank aufbewahrt und auch in Zukunft in modernen Datenbanksystemen wiederverwendet werden. Vgl. <https://www.bar.admin.ch/bar/de/home/archivierung/tools---hilfsmittel/siard-suite.html>.

19 Die Abkürzung csv steht für comma-separated values oder character-separated values. CSV-Daten sind tabellarisch strukturierte ASCII-Dateien. Die einzelnen Werte, Felder oder Spalten werden durch ein spezielles Trennzeichen, beispielsweise das Komma oder Semikolon, getrennt, Tabellenzeilen durch das Zeilenumbruchzeichen. Bei der Überführung einer relationalen Datenbank in CSV-Dateien wird jede einzelne Datenbanktabelle in eine CSV-Datei kopiert. Vgl. [https://de.wikipedia.org/wiki/CSV_\(Dateiformat\)](https://de.wikipedia.org/wiki/CSV_(Dateiformat)) sowie <https://kost-ceco.ch/cms/csv.html?highlight=csv>.

20 JHOVE ist ein in Java programmiertes auf bestimmte Formate beschränktes Tool zur Objektvalidierung (Wohlgeformtheit, Validität). Vgl. <https://de.wikipedia.org/wiki/JHOVE>; <https://jhove.openpreservation.org/>.

21 Vgl. https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives.

22 Es handelt sich um eine Sammlung archivierter Websites, vgl. <https://archive.org/web/>.

23 Der Webrecorder Player ist eine Desktop-Anwendung für die Betrachtung von archivierten Websites ohne die Notwendigkeit einer Internetverbindung. Vgl. <https://github.com/webrecorder/webrecorder-player>.

erfolgreich nach Faust importiert. Das archivierte Objekt kann nun über den Archivbestand „Amt 62“ in der Erschließungssoftware recherchiert werden. Ein Link in der Verzeichnungseinheit verweist auf den DOXIS WebCube (dem Browser-Zugriff für DiPS.kommunal) und die Möglichkeit des Downloads der digital archivierten WARC-Datei. Diese kann als Kopie an einem Sachbearbeiter-PC über den Webrecorder Player digital benutzt werden. Für Benutzer des Archivs steht dieser Service noch nicht zur Verfügung.

Perspektive

Die Straßennamendatenbank soll künftig in regelmäßigen Abständen archiviert und Bürgern im Lesesaal (z.B. mit dem entwickelten DiPS-Nutzungstool) zugänglich gemacht werden. Es wurde jedoch noch nicht darüber entschieden, ob die Stadt Bochum die Technik für Wget oder ein anderes Webarchivierung-Tool implementieren, sich als Dauerprojekt auch der Langzeitarchivierung der offiziellen städtischen Website www.bochum.de oder dem städtischen Intranet sowie den Webpräsenzen von Schulen, Tochtergesellschaften und anderer städtischer Einrichtungen mit Bochum-Bezug widmen wird.²⁴ Da die Archivwürdigkeit hier zweifelsohne zu bejahen ist, wurden diese Websites bereits in der Bochumer Fachverfahrensliste vermerkt.

Die Ergebnisse zum Einsatz von Wget bei dynamischen Websites mit Videoinhalten, externen Links usw. sind noch nicht so zufriedenstellend, als dass das Tool uneingeschränkt empfohlen werden könnte.²⁵ Die hier geschilderten und im DiPS-Anwenderkreis präsentierten Erfahrungen

lassen hoffen, dass sie auch andere Archive dazu anregen, sich dem Thema Langzeitarchivierung von Webpräsenzen zu widmen. Mit dem Abschluss dieses Projekts wurde zusätzlich ein weiterer Meilenstein für den eingangs erwähnten GeLaTi-Arbeitskreis erreicht. Zudem ebnete der Umsetzungsprozess den Weg für weitere Kooperationen und Übernahmeprojekte aus dem Bereich der Geodaten im Amt 62, z. B. der Deutschen Grundkarte 1:5000.²⁶

Schlussendlich hat sich an diesem Übernahmeprozess gezeigt, dass dieser nur in Kooperation mit anderen langzeitarchivierenden Einrichtungen gelingen konnte. ■



Annett Schreiber
Stadtarchiv – Bochumer Zentrum für
Stadtgeschichte
ASchreiber@bochum.de

²⁴ Rechtlich zuständig für die Langzeitarchivierung der Top-Level-Domain www.dnb.de ist die Deutsche Nationalbibliothek. Vgl. https://www.dnb.de/DE/Professionell/Sammeln/Sammlung_Websites/sammlung_websites.html.

²⁵ Vgl. die archivierte Website des Schauspielhauses Bochum unter <http://web.archive.org/web/20191023084245/https://www.schauspielhaus-bochum.de/de/> Die Erhaltungs- und Nutzungsziele für diese Website sind durch das Stadtarchiv bereits definiert worden. Sie können zu einem Großteil durch Wget realisiert werden. Vgl. nestor-Arbeitsgruppe Digitale Bestandserhaltung (Hrsg.), Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung, Version 2, nestor-materialien, 2012. Vgl. https://www.langzeitarchivierung.de/Webs/nestor/DE/Publikationen/nestor_Materialien/nestor_materialien_node.html.

²⁶ LWL-Archivamt für Westfalen, Die Übernahme der Deutschen Grundkarte im Maßstab 1:5000 (DGK5). Handlungsempfehlungen für Katasterämter und Archive, Version 0.6 vom 20.01.2019.