

# Neue Ansätze für die Archivierung von Twitter-Accounts<sup>1</sup>

von Peter Worm

Die am 17. Mai 1994 neu gefasste Gemeindeordnung sorgte für eine einschneidende Veränderung in den nordrhein-westfälischen Gemeinde-, Stadt- und Kreisverwaltungen. Durch die Gesetzesänderung wurde die seit 1946 von der Besatzungsmacht festgelegte Norddeutsche Ratsverfassung mit einer Doppelspitze aus (Ober-)Bürgermeister und (Ober-)Stadtdirektor in die Süddeutsche Ratsverfassung geändert. Dadurch fiel das Amt des Stadt- oder Gemeindedirektors weg und die Verwaltung wird seitdem von einem gewählten hauptamtlichen (Ober-)Bürgermeister geleitet, der zugleich Vorsitzender des Rates ist. In der Gemeindeordnung NRW (GO NRW) heißt es:

„Der Bürgermeister ist kommunaler Wahlbeamter. Der Bürgermeister ist verantwortlich für die Leitung und Beaufsichtigung des Geschäftsgangs der gesamten Verwaltung.“<sup>2</sup>

„Unbeschadet der dem Rat und seinen Ausschüssen zustehenden Entscheidungsbefugnisse ist der Bürgermeister der gesetzliche Vertreter der Gemeinde in Rechts- und Verwaltungsgeschäften.“<sup>3</sup>

Durch die doppelte Spitzenfunktion für die Verwaltung und für den Rat besitzt die Amtsinhaberin/der Amtsinhaber eine besondere Stellung: Sie oder er ist quasi „das Gesicht“ der Stadt. Aus diesem Grund vergeht kaum ein Tag, an dem die Lokalpresse nicht eine Stellungnahme, ein Interview oder einen Bericht unter Beteiligung des (Ober-)Bürgermeisters/der (Ober-)Bürgermeisterin bringt. Dem gegenüber beklagt schon Konrad Schneider 2001 die aus der amtlichen Tätigkeit erwachsende Schriftlichkeit: „Leider lassen sich politische Willensbildungen und unmittelbare Handlungen von Oberbürgermeistern und hauptamtlichen Magistratsmitgliedern nur sehr begrenzt aus den nach Ausscheiden aus dem Amt abgegebenen Vorzimmerregistraturen erkennen. Allem Anschein nach ist es die Furcht, vertrauliche Dinge könnten Nachfolgern aus einem anderen politischen Lager in die Hände fallen, die zu übereilten Kassationen führt.“<sup>4</sup> Katharina Tiemann und Hans-Jürgen Höötman kommen in ihrem fast zur gleichen Zeit erschienenen Beitrag über Archivische Bewertung zu einem ähnlichen Urteil: „Die aufgrund der herausgehobenen Stellung des Registraturbildners geknüpften Erwartungen auf eine umfassende und aussagekräftige Überlieferung erfüllen sich jedoch oftmals nicht.“ Sie empfehlen stattdessen, sich für die überlieferungsmäßige Abbildung der Leitungsfunktion auf die Gegenüberlieferung bei den Ämtern zu verlassen,<sup>5</sup> nehmen von dieser Faustregel bezeichnender Weise aber die „öffentlichkeitsrelevanten Unterlagen (z. B. Reden, Repräsentationsangelegenheiten)“ aus.

## Twitter – ein Fall fürs Archiv?

In diese Kategorie der „öffentlichkeitsrelevanten Unterlagen“ fällt ein Unterlagentyp, der sich erst vor zwölf Jahren entwickelt hat: die Twitter Nachricht (kurz „Tweet“). Bei Twitter handelt sich um einen internetbasierten Microblogging-Dienst, mit dessen Hilfe jeder eine Art Online-Tagebuch führen kann. Zu Anfang (bis 2017) durfte ein Eintrag nicht mehr als 140 Zeichen umfassen, inzwischen sind auch doppelt so lange Tweets erlaubt. Diese sind standardmäßig einsehbar, also auch für unangemeldete Leser sichtbar. Sie können durch sogenannte Hashtags (das englische Wort für das Rautezeichen „#“, mit dem ein solcher Eintrag eingeleitet wird) verschlagwortet werden, Links können als URL eingefügt werden, man kann Verweise auf andere Nutzerprofile mit einem „@“ setzen sowie Bilder und Videos über eine Verlinkung oder direkt einfügen. Erste Zielgruppe dieser Tweets sind die sogenannten Follower eines Benutzers, im Gegensatz zu Facebook steht aber der Kontakt mit bekannten Freunden nicht unbedingt im Vordergrund. Über Hashtags, Verlinkungen und das Weiterleiten von fremden Tweets über das eigene Benutzerkonto (sogenannte Retweets) kann ein breites Publikum erreicht werden.<sup>6</sup>

Im politischen Kontext hat dieses Medium spätestens seit den amerikanischen Präsidenten Barack Obama und Donald Trump einen solch hohen Stellenwert erreicht, dass damit historische Vorgänge beeinflusst werden. Die Library of Congress sah es deshalb in der Zeit von 2010 bis 2017 als eine Aufgabe an, alle weltweit erstellten Twitter-Nachrichten zu sichern. Bis 2013 waren schon 170 Millionen Tweets zusammengekommen, eine gewaltige Datenmenge, selbst wenn man sich auf die Sicherung der Textanteile beschränkte. An die Nachricht angehängte Bilder, Videos oder Verlinkungen zu anderen multimedialen Inhalten wurden nicht gespeichert. 2017 kapitulierte man vor der Masse an Tweets

1 Es handelt sich um die Druckfassung eines Vortrags auf dem Deutsch-Niederländischen Archivsymposium in Vreden vom 5.11.2021 und basiert auf zwei Blogbeiträgen: Peter Worm, Twitter Accounts von kommunalen Wahlbeamten. Bewertung und Überlieferungsstrategien, in: Archive 2.0 am 03.02.2020 (<https://archive20.hypotheses.org/8710>), zuerst veröffentlicht 03.02.2020, zuletzt aktualisiert am 22.10.2021) und Michael Jerusalem/Marc Malwitz/Peter Worm, Ein neuer Ansatz für die Langzeitarchivierung von Twitter-Accounts, in: Archive 2.0 am 09.03.2021 (<https://archive20.hypotheses.org/10031>), zuerst veröffentlicht 09.03.2021, zuletzt aktualisiert 10.03.2021.

2 § 62, Abs. 1 Gemeindeordnung für das Land Nordrhein-Westfalen (GO NRW), Bekanntmachung der Neufassung vom 14.07.1994 mit Stand vom 05.07.2019.

3 § 63, Abs. 1 GO NRW.

4 Konrad Schneider, Das Ende der Aktenzeit? Eine Herausforderung für die Archive, in: Archivar 54 (2001), S. 206.

5 Hans-Jürgen Höötman/Katharina Tiemann, Archivische Bewertung – Versuch eines praktischen Leitfadens, in: Archivpflege in Westfalen-Lippe 52 (2000), S. 1–11, hier S. 6.

6 Zur Funktionsweise, Geschichte und Verbreitungsgrad von Twitter vgl. den Wikipedia-Eintrag unter <https://de.wikipedia.org/wiki/Twitter> [Stand: 09.11.2021, gilt ebenfalls für alle nachfolgende Hinweise auf Internetquellen].

und beschränkt sich seitdem auf die Archivierung von bestimmten Nachrichten: „thematic and event-based, including events such as elections, or themes of ongoing national interest, e. g. public policy“.<sup>7</sup> In Deutschland ist die politische Wirksamkeit der sozialen Netzwerke zwar nach wie vor umstritten, aber in einer großangelegten Studie zu Twitter von 2016 stellen die zwei Autoren, Mathias und Wolfgang König, das grundsätzlich Neue dieses Mediums dar: Es entsteht eine „neue Form von Öffentlichkeit – so eine weit verbreitete Ansicht – [...], weil Web-2.0-Angebote alternative Publikationsstrukturen ermöglichen, herkömmliche Rollen aufweichen (hier Produzierende, dort Konsumierende) und weil mit vergleichsweise kleinem Aufwand große Aufmerksamkeit erzielt werden kann“.<sup>8</sup> Deutsche Spitzenpolitikerinnen und Spitzenpolitiker setzen diese direkte Kommunikation nicht flächendeckend ein, und auch im kommunalen Umfeld ist Twitter sehr unterschiedlich verbreitet. Schaut man auf Münster, so erreicht der Oberbürgermeister (OB) Markus Lewe mit seinem Twitter-Account rund 4.000 Follower (ca. 2 % der Einwohner seiner Stadt), auf eine vergleichbare Quote kommen z. B. die Kölner Oberbürgermeisterin Reker und der Bochumer OB Eiskirch, die ihre Twitter-Accounts in ähnlicher Weise einsetzen. Der Dortmunder OB Sierau wurde 2015 Opfer eines Fake-Accounts, den ein Unbekannter in seinem Namen mit „Neuigkeiten“ und Meinungen bestückte.<sup>9</sup>

Sehr viel höher ist der Verbreitungsgrad von Tweets, der durch das Teilen innerhalb von Twitter (Re-Tweet) oder in anderen Netzwerken erreicht wird. Diese sogenannten „Impressions“ können nur von den angemeldeten Benutzern eingesehen werden. Die so erreichte Reichweite liegt bei Nachrichten von Münsters OB zwischen 1.500 und 10.000, meist zwischen 5.000 und 7.000 Aufrufen. Eine aufwändige Studie von ARD und ZDF ermittelte für 2019, dass nur rd. 2 % der Bevölkerung täglich Twitter nutzen, interessanter Weise deutlich mehr Männer als Frauen. Der Verbreitungsgrad alleine ist jedoch kein Kriterium für eine positive Bewertungsentscheidung – ginge es danach, müsste man sich vielmehr um vorhandene Facebook- oder Instagram-Auftritte kümmern.<sup>10</sup>

Für die archivische Bewertung ist ein wesentliches Kriterium, „ob Twitter in der politischen Kommunikation als verlängerter Arm der Pressestelle von Politikern oder tatsächlich als Kommunikationsplattform zwischen Politiker und Bürger fungiert“.<sup>11</sup> Nur wenn die Tweets (oder zumindest ein erkennbarer Anteil der Tweets) vom Amtsträger als direkter Kommunikationskanal zur Bürgerschaft genutzt wird, sind daraus Rückschlüsse auf Beweggründe und persönliche Schwerpunktsetzungen möglich, die sich nicht an anderer Stelle ausführlicher und detaillierter nachvollziehen lassen. Bezogen auf die Stadt Münster und den Account des Oberbürgermeisters heißt es im Impressum für den Facebook- und den Twitter-Account des OB: „Es posten er selbst (ML) und sein Team (TL bzw. /tl)“. In der Praxis fehlt eine solche Unterscheidung manchmal; in diesen Fällen stammt der Tweet in aller Regel vom OB selbst. Der archivische Wert besteht dabei weniger in den Inhalten, son-

dern in den Themen, die vom Verwaltungsvorstand durch die Posts gesetzt und akzentuiert werden. Darüber hinaus stellen die Tweets oft die unmittelbarste und schnellste Reaktion der Verwaltung auf tagesaktuelle Ereignisse dar und der Gesamtauftritt bei Twitter kann als Quelle für die Selbstdarstellung des Amtsinhabers/der Amtsinhaberin genutzt werden. Bisher hat sich in der deutschsprachigen archivischen Welt noch kein Best-Practice-Verfahren herausgebildet, wie eine adäquate Langzeitarchivierung aussehen könnte. Twitter wurde von Archiven bisher zumeist in seiner Bedeutung als Mittel der Öffentlichkeitsarbeit untersucht, es gibt aber erste Erfahrung mit der Archivierung.<sup>12</sup>

7 Laurel Wamsley, Library Of Congress Will No Longer Archive Every Tweet, in: The Two-Way am 26.12.2017, online abrufbar unter: <https://www.npr.org/sections/thetwo-way/2017/12/26/573609499/library-of-congress-will-no-longer-archive-every-tweet?t=1563203305427>.

8 Mathias König/Wolfgang König, #MythosTwitter. Chancen und Grenzen eines sozialen Mediums (OBS-Arbeitspapier 24), Frankfurt 2016, S. 3.

9 Im Februar 2020 hatte @LeweMarkus noch rund 2.300 Follower und erreichte somit ca. 0,73 % der Bevölkerung. Inzwischen sind es wie oben geschrieben knapp 4.000 Follower (Stand: 09.11.2021) und 2 % der Bevölkerung. Vgl. zum Fake-Account des Dortmunder OBs den Artikel in BILD vom 01.11.2015, online unter <https://www.bild.de/regional/ruhrgebiet/twitter/twitter-posse-um-dortmunds-ob-43182450.bild.html>. nach der Löschung durch Twitter in der Folge unter dem username @nichtUlliSierau angelegte Nachfolge-Account des Unbekannten ist heute noch erreichbar.

10 Für die Zahlen und einen Erfahrungsbericht danke ich Michael Vogt vom Büro des OB in Münster; zum Konsum von Social Media vgl. Natalie Beisch/Wolfgang Koch/Carmen Schäfer, ARD/ZDF-Onlinestudie 2019: Mediale Internetnutzung und Video-on-Demand gewinnen weiter an Bedeutung, in: Media Perspektiven 9 (2019), S. 374–388, bes. Tab. 9, S. 383, online abrufbar unter: [http://www.ard-zdf-onlinestudie.de/files/2019/0919\\_Beisch\\_Koch\\_Schaefer.pdf](http://www.ard-zdf-onlinestudie.de/files/2019/0919_Beisch_Koch_Schaefer.pdf), während es bei anderen Social-Media-Plattformen quantitative Veränderungen gibt, bleibt die Twitternutzung in Deutschland recht stabil bei 2 %, vgl. dies., 25 Jahre ARD/ZDF-Onlinestudie: Unterwegsnutzung steigt wieder und Streaming/Mediatheken sind weiterhin Treiber des medialen Internets abrufbar unter: [https://www.ard-zdf-onlinestudie.de/files/2021/Beisch\\_Koch.pdf](https://www.ard-zdf-onlinestudie.de/files/2021/Beisch_Koch.pdf).

11 Zur Twitternutzung vgl. Mathias König/Wolfgang König, Digitale Öffentlichkeit – Facebook und Twitter im Bundestagswahlkampf 2017. Beitrag auf den Seiten der Bundeszentrale für politische Bildung am 08.05.2018, online unter: <https://www.bpb.de/dialog/podcast-zur-bundestagswahl/264748/digitale-oeffentlichkeit-facebook-und-twitter-im-bundestagswahlkampf-2017> bei Anmerkung 21; eine sicher nicht vollständige Liste prominenter Politikerinnen und Politiker und das Zitat bietet: So twittert die Politik, veröffentlicht am 04.09.2012 [https://www.basecamp.digital/twitter\\_politik/](https://www.basecamp.digital/twitter_politik/). Lewe nutzte schon früh (2010–2012) internetbasierte Möglichkeiten des Bürgerkontakts, so auch die Plattform direktzu@ (<https://direktzu.com/politischer-buergerdialog/>) – sein Account ist zwar inzwischen eingestellt, die Ergebnisse aber noch aufrufbar unter <https://direktzu.de/lewe>. Die Seite des Bonner OB Jürgen Nimptsch wurde bis 2014 fortgeführt, vgl. <https://direktzu.bonn.de/nimptsch>, ebenso die Seite des Münchener OB Christian Ude, vgl. <https://direktzu.muenchen.de/ude>.

12 Kate Theimer, What is the meaning of Archives2.0?, in: The American Archivist 74/1 (2011), S. 58–68; Handreichung zum Umgang mit Sozialen Netzwerken (Social Media) im Bereich der Historischen Bildungs- und Öffentlichkeitsarbeit im Archiv. Empfehlung der Bundeskonferenz der Kommunalarchive beim Deutschen Städtetag vom 30.09.2015, online unter [http://www.bundeskonferenz-kommunalarchive.de/empfehlungen/Handreichung\\_social\\_media\\_Endfassung.pdf](http://www.bundeskonferenz-kommunalarchive.de/empfehlungen/Handreichung_social_media_Endfassung.pdf); Bastian Gillner, Likens, sharen, crowdsourcing. Neue Formen des archivischen Nutzerkontaktes im Netz“ (Vortrag Landesarchivtag Sachsen-Anhalt 2015), <https://archive.20.hypotheses.org/2416>; Antje Diener-Staeckling, Alles kann, nichts muss. Warum Kommunalarchive die Möglichkeiten der Welt des Web 2.0 kennen und nutzen sollten, in: Archivpflege in Westfalen-Lippe 84 (2016), S. 36–40 sowie die Beiträge zum Schwerpunktthema des ARCHIVAR 71 (2018), Heft 1, S. 6–46. In den Niederlanden arbeitet man im „Netzwerk Digitaal Erfgoed“ an einer kooperativen Lösung vgl. Sophie Ham, Monument van onze generatie. Archivering van sociale media, in: archievenblad 10 (2019), S. 6–7. Auf dem Deutsch-Niederländischen Archivsymposium berichtete Mirjam Schaap vom Stadsarchief Amsterdam über ihr Vorgehen, siehe in diesem Heft Markus Stumpf, Tagungsbericht, S. 4.

## Der rechtliche Rahmen

Die erste zu klärende Frage ist, ob Twitter-Daten der Anbietungspflicht unterliegen. Das Archivgesetz NRW fasst in § 2, Abs. 1 unter dem Begriff „Unterlagen“ alle „auch elektronischen Aufzeichnungen, unabhängig von ihrer Speicherungsform, sowie alle Hilfsmittel und ergänzenden Daten, die für die Erhaltung, das Verständnis dieser Informationen und deren Nutzung notwendig sind“, und unterwirft sie der Anbietungspflicht.<sup>13</sup> Dass elektronische Kommunikation keinen Sonderstatus genießt, sondern die Anbietung dem Löschpflichten des Datenschutzrechts vorgeht, wurde in Baden-Württemberg höchstrichterlich im Zuge der sogenannten Mappus-Affäre beantwortet.<sup>14</sup> Für private Benutzerkonten bei Web 2.0-Dienstleistern gilt dies auch, wenn Mitglieder der Verwaltung diese Dienste zu amtlichen Zwecken einsetzen. Für die Daten, die im Messenger-Bereich von Twitter enthalten und die nicht für die Öffentlichkeit sichtbar und bestimmt sind, bedarf es der genauen Abwägung. Überwiegt der private Charakter, kommen sie für eine Archivierung nicht in Frage, Gleiches gilt für die gespeicherten Suchanfragen und die Werbe-Verknüpfungen, die mit diesem Account verbunden sind.<sup>15</sup>

Neben der archivrechtlichen macht eine urheberrechtliche Prüfung Sinn. Tweets sind sehr kurze Äußerungen, sodass nicht davon auszugehen ist, dass sie die notwendige Schöpfungshöhe erreichen, um als Werk zu gelten. Die Rechtsprechung ist jedoch uneinheitlich und vom Einzelfall abhängig.<sup>16</sup> Bei Re-Tweets z. B. von Nachrichtenbeiträgen oder Zeitungsartikeln ist ebenfalls nicht davon auszugehen, dass die Kurznachricht urheberrechtlichen Schutz genießt – anders als der Artikel oder der Video-Ausschnitt selbst, auf den verwiesen wird, der selbstverständlich die nötige Schöpfungshöhe aufweist. Etwas anders sieht es für an die Tweets angehängte Bild- oder Videodateien aus, für die in aller Regel ein urheberrechtlicher Schutz besteht und bei denen sich zumindest für Nutzungskontexte die Frage stellt: Besitzt der twitternde Wahlbeamte (oder die Stadt) die Urheberrechte bzw. zumindest die Nutzungsrechte an dem Material? Auch auf die Erhaltungsstrategie hatte zumindest die alte Rechtslage beim Urheberrecht Auswirkungen: So war es den Archiven bisher nicht erlaubt, mehr als eine Kopie von Objekten zu machen. Diese Problematik wurde jedoch auf europäischer Ebene im Sinne der Archive (und „Einrichtungen des kulturellen Erbes“) geklärt.<sup>17</sup> Die Problematik stellt sich je nach Archivierungsmethode und Bewertungsentscheidung nicht, z. B. wenn sich das Archiv nur für die Sicherung der textlichen Nachricht entscheidet und Bilder und Videos nicht übernimmt oder diese Inhalte beim sogenannten Harvesting der Internetseite (s. u.) ausschließt.

Weniger kritisch ist die Prüfung nach dem Kunsturhebergesetz §§ 22–23 („Recht am eigenen Bild“): Ist der Wahlbeamte selbst abgebildet, so muss er als „absolute Person der Zeitgeschichte“ mit der Veröffentlichung von Bildern leben (auch bei einer späteren Nutzung als Archivgut), sind weitere Personen auf einem Bild oder Video zu sehen, so legt in aller Regel schon die Pose nahe, dass es

sich um ein Pressebild handelt (also die Veröffentlichung intendiert ist und die Dargestellten zugestimmt haben, vgl. zum Beispiel Abb. 5) oder die Personen nur als Beiwerk bzw. im Hintergrund auftauchen.<sup>18</sup>

## Signifikante Eigenschaften

Um die richtige Archivierungsmethode zu wählen, muss das Archiv sich darüber klarwerden, welche Informationen und Funktionen es langfristig erhalten will. Dafür ist es hilfreich, die signifikanten Eigenschaften zu ermitteln, die im Zuge der Archivierung gewahrt werden sollen:

- Geht es dem Archiv um die Sicherung des visuellen Eindrucks / der Anmutung des Twitter-Accounts oder nur um den textlichen Inhalt?
- Sollen online gestellte Bilder oder gar die Video-Sequenzen gesichert werden?
- Wie soll mit Kommentaren und verlinkten Inhalten bzw. sogenannten Re-Tweets aus anderen Twitter-Accounts umgegangen werden?
- Geht es auch um den Erhalt des Klickverhaltens und der Suchfunktionalitäten von Twitter?
- Was soll mit den internen Account-Daten wie den Protokollen zum Klick- und Suchverhalten oder den Messages passieren?

<sup>13</sup> Die Anbietungspflicht begründet fürs staatliche Archivwesen § 4 ArchivG NRW, für den kommunalen Bereich § 10, Abs. 4 mit den Übertragungen aus § 10, Abs. 5, vgl. Martina Wiech, Die digitale Herausforderung im Spiegel der aktuellen deutschen Archivgesetzgebung, in: Archivpflege in Westfalen-Lippe 80 (2014), S. 4–8, bes. S. 4.

<sup>14</sup> Lt. Urteil des Verwaltungsgerichtshofs erläutert in der Pressemitteilung mit Datum vom 04.08.2014, online abrufbar unter: <http://vghmannheim.de/pb/Lde/2271892/?LISTPAGE=2271610>; dazu auch Robert Kretzschmar, Alles neu zu durchdenken? Archivische Bewertung im digitalen Zeitalter, in: Archivpflege in Westfalen-Lippe 80 (2014), S. 9–14, hier S. 9.

<sup>15</sup> Für einen vorsichtig abwägenden Ansatz zwischen „Recht auf Erinnern“ und „Recht auf Vergessen“ bei der Bewertungs- und Übernahmepaxis plädiert vor dem Hintergrund der DSGVO Clemens Rehm, Das „Recht auf Erinnerung“. Zur Relevanz des Archivwesens im Zeitalter der EU-Datenschutzgrundverordnung, in: RECHTSicher. Archive und ihr rechtlicher Rahmen. 89. Deutscher Archivtag in Suhl (Tagungsdokumentation zum Deutschen Archivtag 24), Fulda 2020, S. 45–72 und ders. in diesem Heft, S. 5–13.

<sup>16</sup> Das Landgericht Bielefeld hat in einem aktuellen Fall die Schutzfähigkeit für eine Twitter-Meldung („Tweet“) verneint (Beschluss v. 03.01.2017 – 4 O 141/16). Kurzen Sätzen wird nur selten die notwendige Individualität und Gestaltungshöhe zugebilligt. Gerade weil sie im Umfang begrenzt sind, müssen an ihre eigenschöpferische Prägung besondere Anforderungen gestellt werden. Diese juristische Auffassung überrascht durchaus, da in anderen Kontexten (z. B. bei sog. „Knipsbildern“ von Hobby-Fotografen) der Werkcharakter durch die Gerichte i. d. R. zugebilligt wird. Die Rechtsprechung schätzt dies im Fall kurzer Texte bisher nur in wenigen Ausnahmefällen so ein. Sie stammen alle aus literarischem Kontext, zum Beispiel ein Zitat von Karl Valentin (LG München I, Urteil v. 08.09.2011 – 7 O 8226/11) und von Lorient (LG Braunschweig, Urteil v. 16.01.2013 – 9 O 1144/12). Mindestens genauso unwahrscheinlich ist es, dass einem Kommentar Werkcharakter beigemessen werden muss.

<sup>17</sup> Auf das Problem weist hin: Christian Keitel, Digitale Bestandserhaltung, in: Archivrecht in der Praxis. Ein Handbuch (Berliner Bibliothek zum Urheberrecht 10), München 2017, S. 99–103, bes. S. 101 Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, Erläuterung 25, 27, S. L 130/97–98 und Artikel 6, S. L 130/114. Die Richtlinie gilt allerdings erst für Werke, die ab dem 07.06.2021 entstehen, und sie muss bis zu diesem Datum in nationales Recht umgesetzt werden.

<sup>18</sup> Vgl. zu der Einschätzung Mark Alexander Steinert, Urheber- und andere Schutzrechte an Bildern im Archiv, in: Archivpflege in Westfalen-Lippe 67 (2007), S. 54–57, v. a. 56–57.

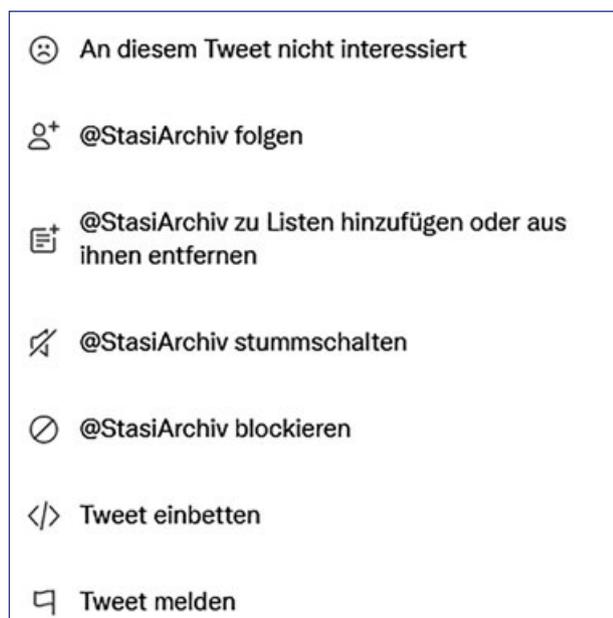


Abb. 1: Kontextmenü zu einem Tweet, mit dem Interaktionen innerhalb von Twitter möglich sind.

Orientiert man sich grob am NESTOR-Leitfaden Bestands-erhaltung<sup>19</sup> kann man folgende Kriterien für den *performance*-Abgleich identifizieren:

**Sinnliche Gesamtwahrnehmung / Anmutung:** Hierunter wird der optische Gesamteindruck des Twitter-Accounts gefasst. Das klingt zunächst klar und eindeutig, doch auf den zweiten Blick muss man sich fragen, welche Sicht auf Twitter soll als Bezugswert dienen? Die Internet-sicht über den Browser (wenn ja in welcher Bildschirmauflösung) oder die Sicht über die Twitter-App aus der Android oder der Apple-Welt oder gar von einem der fast 30 Clients, die die Twitter-Programmierschnittstelle (API) nutzen? Weitere Darstellungsunterschiede entstehen je nachdem, ob man eine Seite als nicht Angemeldeter, als angemeldeter Twitter-Nutzer oder als Eigentümer des Twitter-Accounts besucht. Für die nachfolgenden Überlegungen soll die Desktop-Sicht von twitter.com im Firefox-Browser als Referenz gelten, wobei der Autor bei Twitter angemeldet war, jedoch nicht als Eigentümer der Seite fungierte.

**„Look-and-Feel“:** Gemeint sind die Bedienbarkeit und die Funktionalitäten des Originalauftritts, soweit sie Sinn in der Archivwelt machen. Zum Beispiel sind die Interaktionsmöglichkeiten eines Twitter-Nutzers zu einem Tweet im Kontext des Langzeitarchivs nicht mehr sinnvoll (Abb. 1). Andere Funktionen wie die Mouse-Over-Einblendungen von Accounts (Abb. 2) oder von vollständigen Links, auf die verwiesen wird, dass man diese(-n) bei Klick erreicht oder bei Klick auf einen Hashtag alle Tweets dazu angezeigt bekommt, sind für die Langzeitarchivierung nur schwer realisierbar.

**(Menschen-)Lesbarkeit:** Die Texte der Tweets und die Kommentare sind für Menschen vollständig lesbar.

**(Maschinelle) Recherchierbarkeit:** Die Tweets können z.B. nach einem bestimmten Datum/Hashtag/Begriff maschinell durchsucht werden. Diese Suche basiert

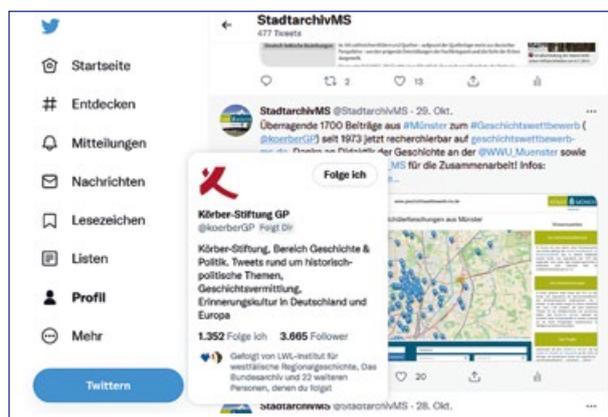


Abb. 2: Automatische Einblendung eines Kurzprofils beim Bewegen der Maus über eine @-Verlinkung (hier @koerberGP = Körber Stiftung).

auf Metadaten oder den als codierter Information vorliegenden Tweets.

**Weiterverwendbarkeit von Texten:** Die Twitterinhalte liegen weitgehend als codierte Informationen vor, so dass man den Text zumindest per Kopieren und Einfügen an anderer Stelle weiternutzen kann.

**Weiterverwendbarkeit von bildlichen Informationen:** In den Tweets enthaltene Bilder lassen sich zumindest per „Speichern-unter“-Befehl in eine andere Programmumgebung transferieren.

**Weiterverwendbare Video-Inhalte:** In den Tweets enthaltene Bewegtbilder, die dort als GIF oder in Video-Formaten vorliegen, können in einer anderen Programmumgebung nachgenutzt werden.

**Maschinelle Verarbeitbarkeit der Informationen:** Primär- und Metadaten liegen in strukturierter Form so vor, dass man sie technisch auswerten und z. B. in eine neue Programmumgebung (halb-)automatisiert importieren kann (csv/xml/json).

**Versteckte Informationen / detaillierte Metadaten zum Account:** Es werden Daten ausgegeben, die nur für den an seinem Twitter-Account angemeldeten Benutzer sichtbar sind. Das umfasst die (Meta-)Daten, die die Erstellung des Accounts betreffen, die Informationen über geblockte oder stumm geschaltete andere Twitter-Konten sowie ggf. die Privatnachrichten und die gespeicherten Daten zum Klick- und Suchverhalten.

Ob und welche dieser Merkmale das Archiv als signifikant erachtet, hängt ganz wesentlich von den Erhaltungszielen ab, die man abbilden möchte, und auch von den Nutzungsgewohnheiten des Tweeters. Viele von diesen nicht oder sehr selten genutzten Funktionen in Twitter sind in aller Regel auch nicht relevant und müssen nicht erhalten werden. Diese binnendifferenzierte Bewertung sollte stets dokumentiert und zur Beschreibung der Akzession genommen werden.

<sup>19</sup> Leitfaden zur digitalen Bestandserhaltung. Vorgehensmodell und Umsetzung. Version 2.0. Frankfurt <sup>2</sup>2013 <http://nbn-resolving.org/urn:nbn:de:kobv:11-100213246>, S. 31 ff.

## Vergleich möglicher technischer Lösungswege

Im Folgenden werden eine Reihe von möglichen technischen Wegen skizzieren, mit denen sich ein Twitter-Konto archivieren lässt. Sie haben keinen Anspruch auf Vollständigkeit, es mag noch eine Reihe von weiteren geben, die z. B. die Twitter-API verwenden. Entscheidend für die Aufnahme in die Überlegungen waren:

- der Weg lässt sich ohne tiefere Informatik-Kenntnisse beschreiben,
- es entstehen langzeitarchivierbare Daten,
- die Lösung ist nicht (über die für die Langzeitarchivierung ohnehin anfallenden Aufwände) mit dauerhaften oder hohen einmaligen Kosten verbunden.

Neben den im Folgenden skizzierten Lösungsansätzen gibt es kostenpflichtige Angebote von öffentlichen und privaten Dienstleistern, die Archive bei der Webarchivierung unterstützen. Hier erhält man – für ein entsprechendes Entgelt – eine Lösung für die Archivierung von Webseiten und Social-Media-Kanälen, die keine oder nur geringe IT-Kenntnisse bei den nutzenden Archiven voraussetzen.<sup>20</sup>

## Screenshots der Twitterseiten

Mit Screenshot-Tools kann man Bildschirminhalte „abfotografieren“ und als Bild speichern. Vorteil dieser Variante ist, dass die ursprüngliche Twitter-Ansicht erhalten bleibt, insbesondere die Kombination von Bild- und Textinformationen.

Nachteilig ist, dass viele Screenshot-Tools nur eine verlustbehaftete JPEG-Speicherung anbieten und nur wenige eine Speicherung als verlustfreie PNG-Grafik ermöglichen.<sup>21</sup> Erfasst man eine große Zahl von Twitter-Nachrichten untereinander in einem Screenshot, erzeugt man damit sehr stark hochformatige Bilder, die ggf. mehrere 1000 Bildpunkte Höhe besitzen und nicht mehr von jedem Viewer angezeigt werden können. Alternativ teilt man die Tweet-Abfolge in viele kleine Screenshots, um den gesamten Twitter-Inhalt des Accounts abzubilden, was wiederum die deutlich aufwändigere Methode darstellt. Bei vielen Tools werden ab einer bestimmten Bildhöhe Seitenumbrüche eingebaut, die die grafische Wiedergabe stören. Weitere wesentliche Nachteile sind, dass keine kodierten Informationen mehr vorliegen, sondern nur noch *non-coded information*, also Bildinformation, und dass auf diese Weise nur die Timeline (also z. B. keine Kommentare zu Tweets) gesichert wird. Daraus entstehen Beschränkungen hinsichtlich der Durchsuchbarkeit und auch des Aussagewerts.

Screenshot-Tools gibt es als Browsererweiterungen oder als externe Software. Während einige browser-basierte Tools das Blättern der Seite während der Bilderfassung erlauben, ist das bei den externen Programmen eher selten der Fall. Eine Reihe von Softwarelösungen ist kostenfrei (oder für den nichtkommerziellen Gebrauch kostenlos nutzbar), aber es gibt auch kostenpflichtige Spezialprogramme auf dem Markt, die das beschriebene Vorgehen vereinfachen.<sup>22</sup>

## Harvesting der Einträge

Es existieren eine Reihe von unterschiedlichen kostenpflichtigen und Freeware-Lösungen, die Internet-Seiten offline verfügbar machen und sich grundsätzlich eignen, auch Twitter-Accounts auszuwerten und lokal nachnutzbar zu machen. Man spricht in diesem Zusammenhang oft von „harvesten“, also „abernten“.<sup>23</sup>

Der Vorteil einer solchen Lösung ist, dass man die Einträge in der Regel als kodierte Information erhält und außerdem noch eine Reihe interessanter Metadaten, die über den Verbreitungsgrad der Nachricht und die Medienwirksamkeit des Twitternden Auskunft geben, ausgelesen werden.

Ein wesentlicher Nachteil ist bei den schlichteren Varianten, dass der bildliche Anteil der Tweets nicht mit gesichert wird, sondern nur als Link auf die Online-Ressource erhalten bleibt. Auch das „Look-and-Feel“ des Twitter-Accounts wird nicht mit gesichert.

Es können beispielsweise reine Online-Tools genutzt werden. Der Anbieter All My Tweets wirbt mit dem Slogan: „View all your tweets on one page“. Er erzeugt dafür eine einseitige HTML-Repräsentation aller Einträge, die über t.co-Kurzlinks mit dem ursprünglichen Tweet verbunden sind (Abb. 3). Das Tool wird weiterentwickelt, sodass in der neuesten Version die Tweets eines Accounts, die Timeline dieses Accounts, Like-Listen und Following- und Follower-Listen ausgegeben werden können. War in früheren

20 An dieser Stelle möchte ich Johannes Schuck (Stadt- und Stiftsarchiv Aschaffenburg) für seine wertvollen Hinweise u. a. auf die kommerziellen Anbieter danken. Vgl. dazu auch die BKK-Empfehlung vom 14.04.2011 (wie Anm. 23), Teil 2: Technik, bes. S. 4 zu Verbundlösungen und Schucks Beitrag in diesem Heft, S. 17–20.

21 Während JPEG-Dateien der ersten Generation durchaus von guter Qualität sein können, handelt es sich bei den hier erzeugten Screenshots um neukomprimierte Bilder, sodass ggf. eine leichte Artefaktbildung auftreten kann. Dieses Phänomen untersuchten Kai Naumann/Christoph Schmidt, Chancen und Risiken verlustbehafteter Bildkompression in der digitalen Archivierung, in: Informationswissenschaft 5/1 (2018) online abrufbar unter <https://doi.org/10.18755/iw.2018.7> und kommen dabei zu sehr interessanten Ergebnissen, die JPEG in deutlich besserem Licht dastehen lassen, als sein Ruf unter Archivarinnen und Archivaren wegen der verlustbehafteten Komprimierung vermuten lässt.

22 Erwähnenswert ist die Software Pablo der Firma Startext, die aus den Internetseiten Bilder macht und den enthaltenen Text und die Struktur und gewisse Funktionalitäten in XML-kodierter Form abbildet und erhaltbar macht. Ob diese Technik bei Twitter funktioniert, wurde im Zuge dieser Untersuchung nicht weiterverfolgt. Eine andere, für die ersten 10 Seiten kostenlose Möglichkeit ist der Online-Dienst <https://perma.cc>, der die Ausgabe als PNG unterstützt und keine Seitenumbrüche setzt. Es wurde bei meinem Versuch eine 1302 x 14912 Pixel große Datei erzeugt, die nicht den gesamten Twitterumfang abbildete.

23 Den deutschen Stand der Diskussion hat zusammengefasst Kai Naumann, Gemeinsam stark – Web-Archivierung in Baden-Württemberg, Deutschland und der Welt, in: ARCHIVAR 65 (2012), S. 33–41 online abrufbar unter [https://www.archive.nrw.de/sites/default/files/media/files/Archivar\\_1\\_2012.pdf](https://www.archive.nrw.de/sites/default/files/media/files/Archivar_1_2012.pdf) sowie zuvor Jens Niederhut, Internetarchivierung. Überlegungen für das Landesarchiv Nordrhein-Westfalen, in: Volker Hirsch (Hrsg.), Golden die Praxis, hölzern die Theorie? Ausgewählte Transferarbeiten des 41. und 42. wissenschaftlichen Kurses an der Archivschule Marburg (Veröffentlichungen der Archivschule Marburg 52), Marburg 2011, S. 123–156 mit einer Vorstellung von den damals verbreitetsten Harvestern HTrack und Heritrix sowie die beiden Empfehlungen der Bundeskonferenz der Kommunalarchive beim Deutschen Städtetag vom 28.09.2010 Speicherung von kommunalen Webseiten – Teil 1: Bewertung online unter [http://www.bundeskonferenz-kommunalarchive.de/empfehlungen/Empfehlung\\_Webarchivierung\\_Teil1\\_Bewertung.pdf](http://www.bundeskonferenz-kommunalarchive.de/empfehlungen/Empfehlung_Webarchivierung_Teil1_Bewertung.pdf) und vom 14.04.2011 Teil 2: Technik online [http://www.bundeskonferenz-kommunalarchive.de/empfehlungen/Empfehlung\\_Webarchivierung\\_Teil2\\_Technik.pdf](http://www.bundeskonferenz-kommunalarchive.de/empfehlungen/Empfehlung_Webarchivierung_Teil2_Technik.pdf).



Abb. 3: Die Ausgabe von All My Tweets erlaubt ein rudimentäres Sichern eines Twitter-Kontos.

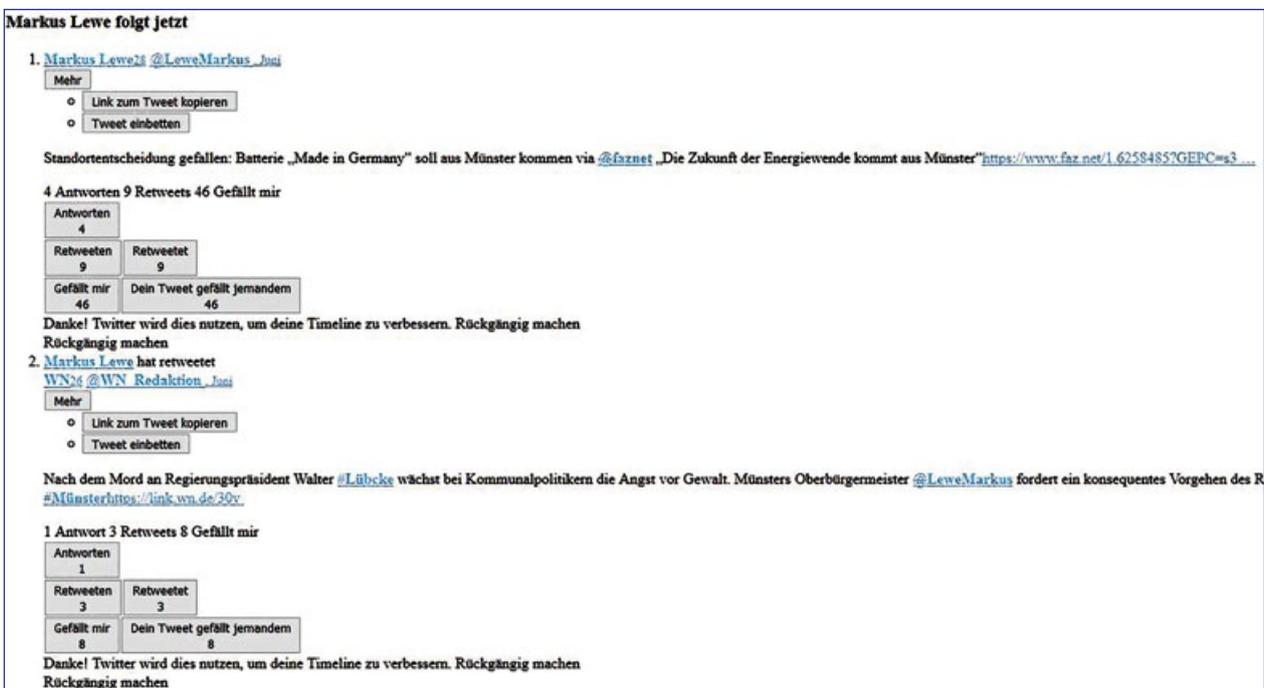


Abb. 4: Der mit Wget erzeugte WARC-Container, wie er im Webrecorder angezeigt wird.

Versionen keine Anmeldung notwendig, benötigt man inzwischen die Login-Daten des Twitter-Accounts, um einen Bericht zu erhalten. Da die Kommentare nicht ausgegeben werden, stellt dieser Weg für das Stadtarchiv Münster keine optimale Lösung dar. Bei weniger ambitionierten Erhaltungszielen (nämlich nur die Inhalte recherchierbar zu halten) kann dieser Webdienst sinnvoll eingesetzt werden und die erzeugte HTML-Seite für die Langzeitarchivierung lokal abgespeichert werden.

Eine andere und professionellere Art der lokalen Auswertung erzeugt der Twitter Archiver, der mit dem Dienst IFTTT arbeitet und die gewonnenen Daten in eine Google Tabelle abspeichert.<sup>24</sup> Die einzelnen Tweets werden als Listeneinträge abgebildet. Die mit EXCEL vergleichbare Anwendung enthält die folgenden Spalten: *Date*, *Screen Name*, *Full Na-*

*me*, *Tweet Text*, *Tweet ID*, *Link(s)*, *Media*, *Location*, *Retweets*, *Favorites*, *App*, *Followers*, *Follows*, *Listed*, *Verified*, *User Since*, *Location*, *Bio*, *Website*, *Timezone*, *Profile Image*. Diese an und für sich interessante Form der Webarchivierung hat einen wesentlichen Nachteil: Es können damit keine alten Tweets abgeerntet, sondern nur prospektiv seit der Einrichtung neu angelegte Tweets gespeichert werden. Damit ist sie in dieser Form für die Archivierung ungeeignet.

Ein weiteres Harvesting Tool, Wget, diente ursprünglich unter Linux dazu, Internetseiten offline anzeigen zu können, also eine lokale Kopie der Seite zu erzeugen. Mit der Version 1.21.2 lassen sich sog. WARC-Container aus Internetseiten generieren.<sup>25</sup> Das Format ist inzwischen als ISO 28500:2017 standardisiert und erlaubt die komprimierte Speicherung einer Internetseite in einer einzigen Datei. Die Bedienung erfolgt über die Befehlszeile (also ohne grafische Oberfläche), so dass man sich zunächst an ein gu-

tes Ergebnis herantasten muss. Je nach Konfiguration erhält man einen dem Twitter Account ähnlichen Output,

24 Zur Chrome-Erweiterung Twitter Archiver vgl. [https://workspace.google.com/marketplace/app/tweet\\_archiver/976886281542](https://workspace.google.com/marketplace/app/tweet_archiver/976886281542); zum Diensteanbieter IFTTT vgl. <https://de.wikipedia.org/wiki/IFTTT>. Andere z. T. kostenpflichtige Analysetools greifen ebenfalls auf die Twitter-API zurück und erlauben auch hier Teile der Twitter-Inhalte herunterzuladen und zu sichern – in der Regel werden diese Daten in JSON ausgegeben. Diese Lösungen sind v. a. für die massenhafte Auswertung von Tweets als Forschungsgrundlage interessant, vgl. Katrin Weller, Big Data & New Data: Ein Ausblick auf die Herausforderungen im Umgang mit Social-Media-Inhalten als neue Art von Forschungsdaten, in: Uwe Jensen/ Sebastian Netscher/Katrin Weller (Hgg.), Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten Grundlagen und praktische Lösungen für den Umgang mit quantitativen Forschungsdaten, Berlin/Toronto 2019, S. 193–210, bes. S. 198–199, online unter <https://doi.org/10.3224/84742233.12>.

25 Die Entstehungsgeschichte und die Links zu den aktuellen Versionen enthält <https://de.wikipedia.org/wiki/Wget>.



Abb. 5: Das Ergebnis eines Harvesting-Laufs mit dem Webrecorder.io.

den man nur mit einer geeigneten Software, einem sogenannten WARC-Viewer wie dem Webrecorder, wieder zur Ansicht bringen kann. Meine Tests waren leider nicht sehr erfolgreich (Abb. 4), was vermutlich auch an meiner unzureichenden Bedienung des Programms über die Befehlszeile liegt.

Als alternatives Harvesting Tool zur Erzeugung von WARC-Dateien kann der zuvor nur als Viewer angesprochene Webrecorder.io selbst eingesetzt werden. Er lässt sich über die grafische Oberfläche einfach bedienen und das Ergebnis überzeugt durch die grafische Qualität und die ans Original heranreichende Anmutung, die u. a. durch eine vollständige Einbettung von Medien-Files erreicht wird (Abb. 5). Die Qualität des WARC-Containers hängt bei eingebetteten Video- und Foto-Dateien davon ab, dass man beim Aufzeichnen der Seite ausreichend langsam und manuell durch die entsprechende Seite geht und ggf. davon, dass Videos abgespielt werden, damit der hinterlegte Video-File Teil des WARC-Containers wird. Auch die Sicherung von Kommentaren anderer Tweetenden geschieht nur, wenn während des Aufzeichnens der Seiten die Vollansicht eines Tweets angefordert wird. Hieraus ergibt sich ein aufwändiger und zum Teil fehleranfälliger Prozess, der eine intensive Qualitätssicherung des erzeugten Containers nach sich zieht.

Die Mühe lohnt sich auf den ersten Blick, wenn man die sehr nah am Original liegende Anmutung der archivierten Seiten sieht. Gerade bei der Ausgabe aus dem Webrecorder fragt man sich als Anwender ab und zu, ob man sich gerade im Original oder in der Archivsicht von Twitter befindet. Die Tweets bleiben als *coded information* erhalten, die Fotos und Videos werden in den WARC-File integriert. Einige Kernmetadaten zu Tweets und dazu, wer einem folgt und wem man selbst folgt, werden mitgesichert, da sie auf den aufgezeichneten Seiteninhalten zu sehen sind. Der WARC-Container ist für die Langzeitarchivierung aus der Browseranwendung des Webrecorder.io heraus exportierbar.

Die Verschmelzung aller Teile einer Twitterseite in einem WARC-Container stellt gleichzeitig einen Nachteil dar, da die einzelnen Objekte durch spätere Nutzer nicht einfach exportiert und nachverwendet werden können, sondern dafür zunächst mit einem geeigneten WARC-Viewer zur Ansicht gebracht werden müssen. Dann lassen sich viele Multimedia-Inhalte über die „Speichern unter“-Funktion im Kontextmenü als Datei extrahieren. Die langfristige Erhaltung des WARC-Containers und vor allem der enthaltenen unterschiedlichen Datenformate kann trotz des ISO-Standards nicht hundertprozentig ga-

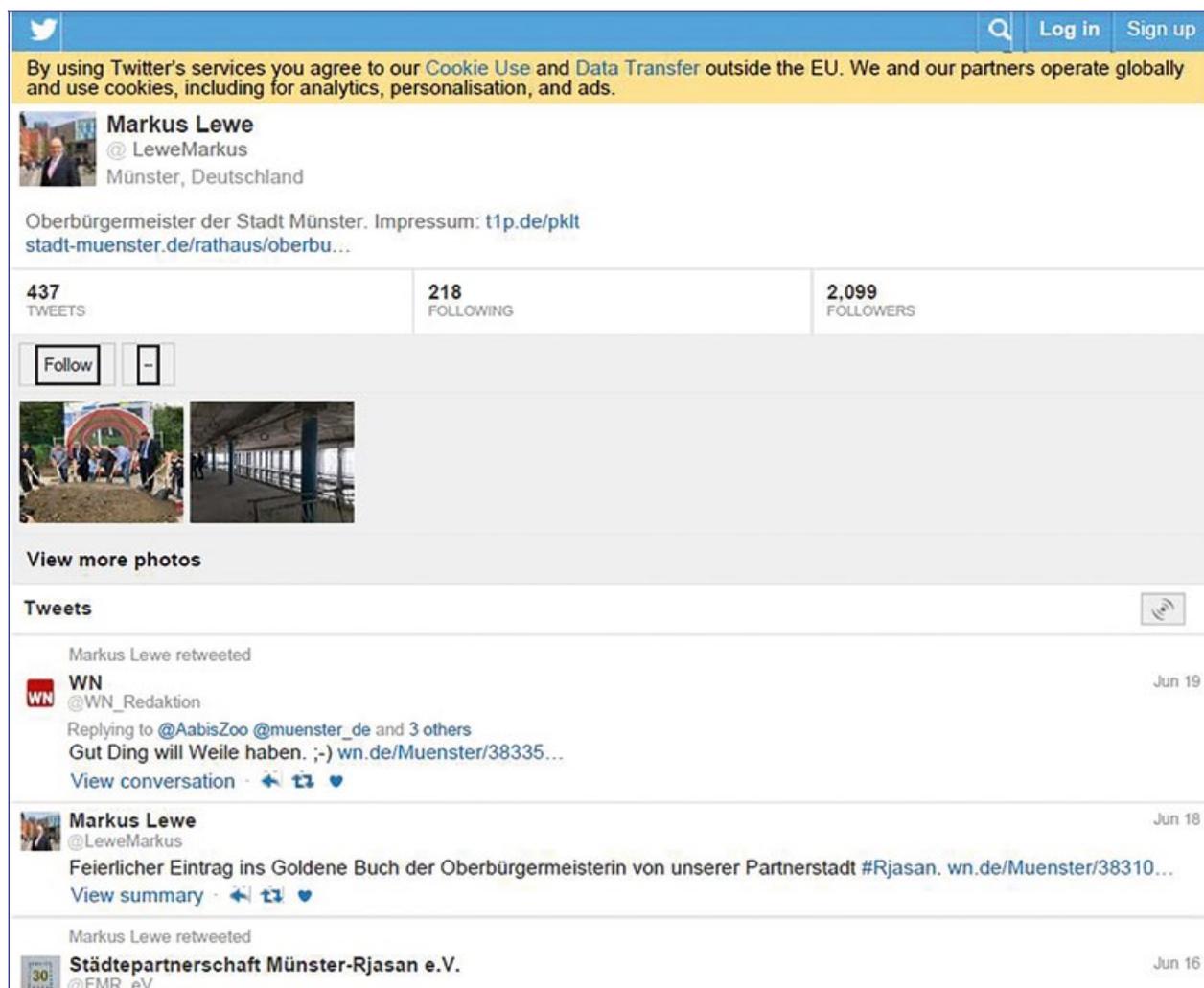


Abb. 6: Die mit dem Acrobat erzeugte PDF-Ausgabe des Twitter-Accounts vermag nicht zu überzeugen.

rantiert werden.<sup>26</sup> Als problematisch könnte sich erweisen, dass im Zuge der Verarbeitung einer WARC-Datei durch das Langzeitarchiv (Ingests) zwar die Wohlgeformtheit (Validität) dieses Containerformats, aber nicht jedes im Container enthaltene Objekt geprüft wird. Da zur Anzeige multimedialer Inhalte bestimmte Codecs (Algorithmenpaare, mit denen Daten oder Audio- und Video-Signale digital kodiert und dekodiert werden) notwendig sind, müssten diese detailliert untersucht und in den Metadaten des Archivinformationspakets (AIP) hinterlegt werden. Nur so ist es möglich, bei einer drohenden Obsoleszenz eines dieser Codecs mit einer Datenmigration in aktuelle Kodierungen zu reagieren. Trotz dieser Schwächen genießt WARC gerade in der anglo-amerikanischen Fachwelt bei Bibliotheken und Archiven eine große Verbreitung und hohe Akzeptanz, sodass nach heutigem Stand – auch mangels entsprechender Alternativen – nichts gegen die Verwendung als (Übergangs-)Langzeitarchivformat spricht.

## PDF-Erstellung

Der Adobe Acrobat besitzt über sein Browser-Plugin die Möglichkeit, Internetseiten in PDF zu überführen. Dabei wird in der Regel auf die Druckausgabe der Seiten zurück-

gegriffen, die ggf. deutlich von der Bildschirmanzeige abweichen kann. Wendet man diese Ausgabefunktion auf einen Twitter-Account an, so vermag das erzeugte PDF nicht recht zu überzeugen: Weder ist es ein vollständiges bildliches Pendant zum Twitter-Inhalt noch ein vollständiges textliches Pendant (Abb. 6).

Vorteilhaft ist, dass die Seiten als kodierte Information erhalten bleiben und manche Fotos in das PDF integriert werden. So wird späteren Nutzern zumindest ein grober Eindruck von der Text-Bild-Zusammenstellung vermittelt. In der PDF-Ausgabe kann man eine Reihe von Kernmetadaten zum Account finden (jeweils die Anzahl der Tweets, der Personen, denen man folgt und von denen man verfolgt wird, werden ausgegeben). Gerade diese „Zwitterfunktion“ macht es unattraktiv, hierauf eine Archivierung

<sup>26</sup> Vgl. Katalog archivischer Dateiformate, Version 6.1, August 2020, online unter <https://kost-ceco.ch/cms/warc.html>; The National Archives. The technical registry PRONOM, online unter <http://www.nationalarchives.gov.uk/pronom/fmt/1281>. Die Schweizer Kolleginnen und Kollegen positionieren sich klar: „Obwohl WARC das eingeführte Format der Webarchivierung ist, kann es als Langzeitformat nicht empfohlen werden: Die Vielzahl möglicher eingebetteter Formate und die Schwierigkeit, diese zu migrieren, sprechen dagegen. Für die Archivierung eines gesamten Webauftritts empfiehlt sich längerfristig eindeutig eine Umwandlung in PDF/A“, vgl. <https://kost-ceco.ch/cms/hypertext-einleitung.html>.

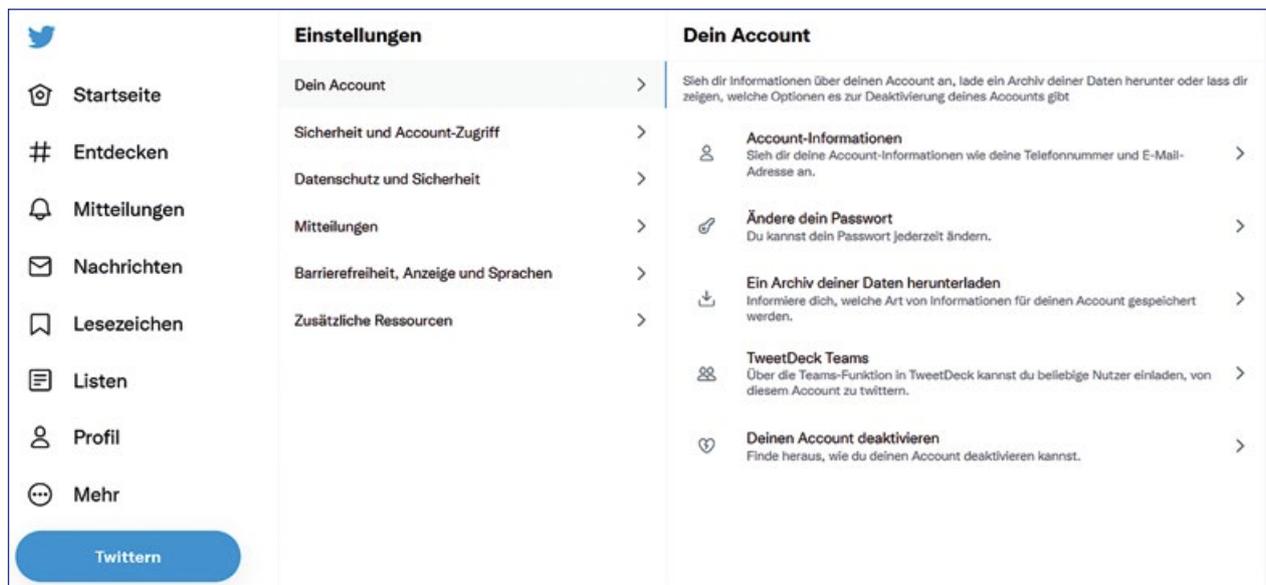


Abb. 7: Oberfläche in Twitter, mit der der Export des Profils möglich ist („Ein Archiv deiner Daten herunterladen“).

aufzusetzen. Es entsteht ein ganz eigener Eindruck, der deutlich von der ursprünglichen Erscheinung der Seite abweicht. Darüber hinaus werden viele Bilder nicht ins PDF integriert, sondern wie evtl. vorhandene Links nur als externe Verweise übernommen. Gerade diese verlinkten Informationen werden langfristig kaum erhalten werden können.

### Twitter Export als Grundlage

Verfügt man über Admin-Rechte für einen Twitter-Account, bietet der Online-Service die Möglichkeit an, die Daten komplett herunterzuladen (Abb. 7).

Auf das Ergebnis muss man eine Weile warten, aber man wird über die mit dem Account verknüpfte E-Mail-Adresse informiert, sobald ein Download als ZIP-Archiv möglich ist.<sup>27</sup> Die erzeugte ZIP-Datei enthält einen vollumfänglichen Export aller Daten des Twitter-Kontos, der auch die nicht frei verfügbaren Inhalte wie private Messages (DMs), Momente (moments), Followerlisten, Adressbücher und andere selbst erstellte Listen, ja sogar verknüpfte Werbeeinhalte und demografische Kennwerte berücksichtigt. Nicht vom Konto-Eigentümer verantwortete Elemente, wie die von anderen

```

window.YTD.account.part0 = [ {
  "account" : {
    "phoneNumber" : "+4900000000000000",
    "email" : "00000000@stadt-muenster.de",
    "createdVia" : "oauth: 000000",
    "username" : "000000000000",
    "accountId" : "000000000000000000000000",
    "createdAt" : "2010-00-00T09:32:49.213Z",
    "accountDisplayName" : "00000000000000"
  }
} ]

```

Abb. 8: Als Beispiel der anonymisierte Inhalt der Datei „account.js“, der durchaus sowohl menschen- als auch maschinenlesbar die zentralen Daten zur Erzeugung, der Bezeichnung und den bei der Erzeugung verwendeten Kontaktdaten eines Twitter-Kontos aufweist.

Tweetern angebrachten Kommentare, werden leider nicht ausgegeben. Die exportierten Inhalte werden als JSON Dateien (Javascript Object Notation) ausgegeben (Abb. 8).<sup>28</sup>

Bei der Detailbewertung der exportierten JSON-Dateien<sup>29</sup> enthalten die folgenden Dateien archivwürdige Informationen:

<sup>27</sup> Eine detaillierte Anleitung bietet Twitter auf den Hilfeseiten unter <https://help.twitter.com/de/managing-your-account/accessing-your-twitter-data>. Im README.TXT des Exports schreibt Twitter: „This archive consists of machine-readable JSON files containing information associated with your account. We've included the information we believe is most relevant and useful to you, including your profile information, your Tweets, your DMs, your Moments, your media (images, videos and GIFs you've attached to Tweets, DMs, or Moments), a list of your followers, a list of accounts following you, your address book, Lists that you've created, are a member of, or are subscribed to, interest and demographic information that we have inferred about you, information about ads that you've seen or engaged with on Twitter, and more. The information contained in this archive reflects the state of the account at the time when the archive was created. In addition, if we do not have any data associated with your account for a particular category (e.g., if you have never created a List), then this archive will not include a file for that category.“

<sup>28</sup> Will man für Erhaltungs- oder Auswertungszwecke aus dem ausgelieferten JSON eine valide CSV oder EXCEL-Datei erzeugen, so lässt sich das über webbasierte Tools ohne Programmierkenntnisse vornehmen, z. B. mit

Hilfe der Internetseite <https://json-csv.com/>. Als Feldtrenner ist in diesem Fall ein geeignetes Zeichen – zum Beispiel der sog. Pipe (|) – zu wählen. Der Feldtrenner sollte so gewählt sein, dass er in den Tweets selbst oder in den exportierten sonstigen Daten nicht enthalten ist.

<sup>29</sup> Die Daten des Twitter-Accounts werden in den folgenden Einzeldateien ausgegeben (alphabetisch sortiert): account.js, account-creation-ip.js, account-suspension.js, account-timezone.js, ad-engagements.js, ad-impressions.js, ad-mobile-conversions-attributed.js, ad-mobile-conversions-unattributed.js, ad-online-conversions-attributed.js, ad-online-conversions-unattributed.js, ageinfo.js, block.js, connected-application.js, contact.js, direct-message.js, direct-message-group.js, direct-message-group-headers.js, direct-message-headers.js, email-address-change.js, follower.js, following.js, ip-audit.js, like.js, lists-created.js, lists-member.js, lists-subscribed.js, moment.js, mute.js, ni-devices.js, periscope-account-information.js, periscope-broadcast-metadata.js, periscope-comments-made-by-user.js, periscope-expired-broadcasts.js, periscope-followers.js, periscope-profile-description.js, personalization.js, phone-number.js, profile.js, protected-history.js, saved-search.js, screen-name-change.js, tweet.js, verified.js.

account.js	Die Metadaten zum Konto selbst und zu den bei der Erzeugung verwendeten Kontaktdaten
follower.js	Die Account-Nummern der Twitter-Konten, von denen man verfolgt wird
following.js	Die Account-Nummern der Twitter-Konten, denen man folgt
profile.js	Die im Profil verwendeten Beschreibungsdaten zur twitternden Person
profile_media (Verzeichnis)	Die Profilbilder in einem weiteren ZIP-Container (...profile_media\profile-media-part1.zip)
tweet.js	Die textlichen Inhalte der Tweets, jedoch ohne Kommentare!
tweet_media (Verzeichnis)	Die in Tweets verwendeten Bilder und Videos in einem weiteren ZIP-Container (...tweet_media\tweet-media-part1.zip)

Genauer zu bewerten sind:

block.js	Blockierte Twitter-Accounts, die von der Interaktion mit dem Tweeter komplett ausgeschlossen waren (wird häufig auf sogenannte Trolle <sup>30</sup> angewandt)
like.js	Tweets, die als Favoriten gekennzeichnet oder mit „Like“ versehen wurden
lists-member.js	Öffentliche Listen, in denen das archivierte Konto aufgeführt wird
lists-subscribed.js	Listen, in denen sich der Tweeter mit seinem Konto angemeldet hat
moment.js	Eine Sammlung an Tweets, die via Twitter geteilt werden kann (es ergibt sich notwendiger Weise eine Redundanz zu den Einzeleinträgen – zu klären ist hier, ob die Informationszusammenstellung archivwürdig erscheint)
moments_media (Verzeichnis)	Die in den sog. Moments verwendeten Medien
mute.js	Stumm geschaltete Twitter-Accounts, deren Tweets nicht mehr angezeigt werden sollten

Die Tweets und ihre zugehörigen Medien lassen sich leicht miteinander verknüpfen: der Dateiname der Mediendatei folgt der Konvention [Tweet-ID]-[Dateiname auf dem Twitter-Server].[Endung]].

Die Exporte aus Twitter enthalten z. T. kodierte Informationen. So werden bei geblockten und verfolgten Accounts die Account-IDs angegeben statt des Klarnamens (*username*). Aus Sicht der Langzeitarchivierung sollten diese Informationen in Klartext überführt werden, z. B. unter Zuhilfenahme von <https://tweeterid.com/>. Gerade bei intensiv vernetzten Personen wie Kommunalpolitikern kann dieses Identifizieren erhebliche Mühe machen, so dass unter Umständen zu empfehlen ist, sich die Informationen – wie unten für die Follower-Listen vorgeschlagen – im Nutzerprofil<sup>31</sup> anzeigen zu lassen und dann über die Kopieren-und-Einfügen-Funktion zu sichern, da hier die Klartextinformation dargestellt wird.

Wirkungsvolle Tests dieser Überlieferungsmethoden kann das Archiv nur in Zusammenarbeit mit dem Inhaber des Twitter-Kontos durchführen. Dann hat diese Ausgabe den Vorteil, dass man mutmaßlich alle Inhalte des Accounts in strukturierter Form als *coded information* bekommt und somit eine optimale technische Nachnutzbarkeit für spätere Nutzer erreicht. Die Auswahl der Meta- und Primärdaten kann, wie die vorstehende Detailbewertung zeigt, nach individuellen Kriterien pro Twitter-Account erfolgen. Die Form erlaubt es darüber hinaus, leicht und präzise Zeitschnitte festzusetzen – sei es nach Wahlperioden, nach Jahren oder bei stark genutzten Twitter-Konten auch nach Monaten, um so eine leichte Zugänglichkeit für spätere Nutzer sicherzustellen. Jedoch geht bei der Methode das ursprüngliche *Look-and-Feel* und die Anmutung von Twitter vollständig verloren. Will man eine komfortablere Ansicht als der pure JSON- oder ein daraus generierter CSV-File bieten, müssen diese Files vor der Nutzung nachbearbeitet und in eine Oberfläche eingespielt werden, die die Daten strukturiert sichtbar macht und mit zugehörigen Mediendateien verknüpft. Ein weiterer erheblicher Nachteil ist, dass die Kommentare der anderen Tweeter nicht mit ausgegeben werden, wodurch eine Nutzung als Quelle für die Bürgerbeteiligung wegfällt.

Fasst man die technischen Lösungswege mit ihren Vor- und Nachteilen zusammen, erhält man folgende Übersicht:

30 „Als Troll bezeichnet man im Netzjargon eine Person, die im Internet vorsätzlich mit „zündelnden“ Flame-Kommentaren einen verbalen Disput entfachen oder absichtlich Menschen im Internet verärgern will. (...) Ihre Kommunikation in diesen Communities ist auf Beiträge beschränkt, die auf emotionale Provokation anderer Gesprächsteilnehmer zielen.“ Vgl. dazu: [https://de.wikipedia.org/wiki/Troll\\_\(Netzkultur\)](https://de.wikipedia.org/wiki/Troll_(Netzkultur)).

31 Auf twitter.com kann sich der Eigentümer des Accounts unter „Mehr“ / Rubrik „Datenschutz und Sicherheit“ / Unterrubrik „Sicherheit“ die „Stummgeschalteten“ und „Blockierten Accounts“ auflisten lassen.

Lösungsweg Kriterium	Screen- shot	All My Tweets	IFTTT- basiert	WARC via Web- recorder	PDF	Twitter- Export
<b>Sinnliche Gesamtwahrnehmung / Anmutung</b>	Ja	Nein	Nein	Ja	±	Nein
<b>„Look-and-Feel“</b>	±	Nein	Nein	Ja	Nein	Nein
<b>(Menschen-)Lesbarkeit</b>	Ja*	Ja*	Ja*	Ja	Ja*	Ja*
<b>Recherchierbarkeit, z. B. nach einem bestimmten Datum / Hashtag / Stichwort</b>	Nein	Ja	Ja	±	Ja	Ja
<b>Weiterverwendbarkeit von textlichen Informationen</b>	Nein	Ja	Ja	Ja	Ja	Ja
<b>Weiterverwendbarkeit von bildlichen Informationen</b>	±	Nein	Nein	±	±	Ja
<b>Weiterverwendbare Video-Inhalte</b>	Nein	Nein	Nein	±	Nein	Ja
<b>Maschinelle Verarbeitbarkeit der Informationen</b>	Nein	Nein	±	Nein	Nein	Ja
<b>Versteckte Informationen / detaillierte Metadaten zum Account</b>	Nein	Nein	±	±	Nein	Ja

\* jedoch nicht vollständig, da ohne Kommentare.

Den höchsten Abdeckungsgrad an signifikanten Eigenschaften erhält man, wenn man auf die Kommentare von anderen Tweetenden verzichten kann, durch die Kombination eines nutzerbasierten Exports der Twitter-Daten, ergänzt um Screenshots des Auftritts.<sup>32</sup> Dabei sollte dafür nach Möglichkeit ein unkomprimiertes oder nicht neu komprimiertes Bildformat genutzt werden und eine möglichst gute Auflösung der Bilder erzielt werden. Der Bildausschnitt sollte möglichst exakt definierbar sein und über die aktuelle Bildschirmsicht hinausgehende Seitenausschnitte erlauben. Umbrüche in den Screenshots sollten vermieden werden. Für diese Varianten kann überlegt werden, ob eine grafische Entsprechung für den gesamten Tweet-Bestand notwendig ist, oder es genügt, exemplarisch zu zeigen wie die Handhabung und das Aussehen von Twitter zum Archivierungszeitpunkt war.

Benötigt man, wie hier in den Vorüberlegungen dargestellt, die Kommentare, um die Interaktion des Wahlbeamten mit den Bürgerinnen und Bürgern abzubilden, würde ein valider WARC-Export gegebenenfalls in Kombination mit dem Twitter-Export die notwendigen signifikanten Eigenschaften besser abbilden. Der Umstand, dass es weder mit der Browserversion noch mit der Stand-Alone-Installation des Webrecorders gelungen ist, einen validen WARC-Container zu exportieren, verhinderte jedoch, diesen Weg zu beschreiten.<sup>33</sup>

### TWINT als erfolgversprechender Ansatz

Mit diesem nicht befriedigenden Fazit endete der im Februar 2020 veröffentlichte Blogartikel. Er war bewusst als Werkstattbericht gedacht, der zunächst die grundsätzliche Frage der Archivwürdigkeit und dann signifikante Eigenschaften sowie Erhaltungsziele diskutierte. Der Abgleich mit den damals vorhandenen Lösungswegen machte eines klar: Es gab keinen „Königsweg“, mit dem alle Erhaltungsziele vollständig abgedeckt wurden. Die am Schluss skizzierten technischen Maßnahmen lieferten zumeist provisorische Ergebnisse oder mussten miteinander kombiniert werden, um alle Erhaltungsziele abzudecken

Ein alternativer Lösungsweg musste her. Hier half eine Reaktion auf den Archive 2.0 Beitrag von Michael Schmalenstroer weiter.<sup>34</sup> Er verwies auf gute Erfahrungen mit dem in der Programmiersprache Python geschriebenen Harvesting-Tool TWINT<sup>35</sup>. Nach kurzer Orientierung über die Möglichkeiten ließen sich damit bereits überzeugende

32 Zu einem ähnlichen Befund kommt Justin Littman, *Web archiving and/or vs social media API archiving*, in: *Social Feed Manager* am 13.12.2017, online abrufbar unter <https://gwu-libraries.github.io/sfm-ui/posts/2017-12-13-web-social-media-archiving>.

33 Auf das Problem machte dankenswerter Weise Heike Beimfohr von der LWL.IT im Rahmen des DiPS.kommunal Kundensupports aufmerksam.

34 Unter der Anmeldung „Michael“ vom 4. Februar 2020, <https://archive20.hypotheses.org/8710#comment-30285>.

35 TWINT steht für Twitter Intelligence Tool und ist quelloffen und kostenfrei (MIT License) verfügbar unter <https://github.com/twintproject/twint>.

1277968987150618624	Markus Lewe	2020-06-30	16:15:24	Vier neue E-Busse sind der nächste Schritt der Stadtwerke, Münster ein Stück grüner zu machen. Sie fahren ab diesem Sommer auf der Linie 11 von Gievenbeck durch die Innenstadt bis zur Mondstraße. Die Busse werden mit Ökostrom betankt und fahren leise. <a href="https://pic.twitter.com/e5WQeeHTU">pic.twitter.com/e5WQeeHTU</a>		<a href="https://pbs.twimg.com/media/EbWAZNBUBAYjGW.jpg">https://pbs.twimg.com/media/EbWAZNBUBAYjGW.jpg</a>	
1278082791431954434	Almanó	2020-06-30	23:47:37	🤔🤔	lewemarkus		
1279781866342678528	Sabarned Oia lie oia	2020-07-05	16:19:08	Hoppla, da brauchen wir ja bald ein eigenes Kraftwerk...	lewemarkus		
1277529325802328066	Markus Lewe	2020-06-29	11:08:21	<a href="https://pic.twitter.com/6Ldo0Vgg1G">pic.twitter.com/6Ldo0Vgg1G</a>		<a href="https://pbs.twimg.com/media/Eb9yCDWwAcaKWS.jpg">https://pbs.twimg.com/media/Eb9yCDWwAcaKWS.jpg</a>	
1278377043240116224	Jutta Manufeld	2020-07-01	19:16:52	Lieber Herr Lewe, haben Sie ebenfalls eine gute Zeit und bleiben Sie gesund.	lewemarkus		
1276146476612288512	Markus Lewe	2020-06-25	15:33:24	<b>#Spatenstich</b> für 247 neue Wohnungen im Zentrum Nord, von denen fast 80% öffentlich gefördert und somit preisgebunden sind. Neben den großen Vorhaben auf den <b>#Konversionsflächen</b> , sind es auch Projekte wie dieses, die zur <b>#Verbesserung der Lage</b> auf dem <b>#Wohnungsmarkt</b> beitragen. <a href="https://pic.twitter.com/HgZcJQCF">pic.twitter.com/HgZcJQCF</a>		<a href="https://pbs.twimg.com/media/EbXlcuXgAAa3NR.jpg">https://pbs.twimg.com/media/EbXlcuXgAAa3NR.jpg</a> <a href="https://pbs.twimg.com/media/EbXlcw4WoAlFYH.jpg">https://pbs.twimg.com/media/EbXlcw4WoAlFYH.jpg</a> <a href="https://pbs.twimg.com/media/EbXlcw30IAADwO6.jpg">https://pbs.twimg.com/media/EbXlcw30IAADwO6.jpg</a>	<b>#spatenstich</b> <b>#konversionsflächen</b> <b>#verbesserung</b> <b>#wohnungsmarkt</b>
1276146878845988864	pflegekraft-muenster.de 🇩🇪	2020-06-25	15:35:00	Bitte auch an altersgerechten Wohnraum denken...	lewemarkus		
1276385736095404034	Markus Lewe	2020-06-26	07:24:08	Machen wir. Ein Teil der Wohnungen wird auch altersgerecht ausgestattet sein. Auch Begegnungsräume mit Betreuung sind Teil dieses Projektes. <a href="https://pic.twitter.com/HgZcJQCF">pic.twitter.com/HgZcJQCF</a>	pflegekraft_ms		

Abb. 9: Erste Exportversuche mit dem TWINT-Skript: die Ausgabe erfolgt als HTML-File.

Ausgaben von Follower- und Following-Listen als JSON erzeugen.<sup>36</sup> Dafür kann Twint mit den beiden folgenden Befehlszeilen angesteuert werden:

```
twint -u [username] -following -user-full -o following_[username].json -json
twint -u [username] -follower -user-full -o follower_[username].json -json
```

Es stellte sich die Frage, ob mit dieser Software nicht auch die Timeline inklusive der Kommentare zu den Tweets gesichert werden könnte. Nativ war das zunächst nicht möglich – in der Grundfunktionalität

```
twint -u [username]
```

fehlten bei den Exporten die Bilder und Multimediadateien genauso wie die Kommentare. Aber es gab drei Lichtblicke:

1. TWINT nutzt zur Erstellung seiner Berichte nicht die Twitter-API, die eine strenge Zuordnung der Informationen zu den Accounts des Tweetenden vornimmt, und eine Verknüpfung der Kommentare zu den Tweets, auf die sie sich beziehen, nicht zulässt, sondern eine alternative Methode.
2. Die Software ist quelloffen und kann durch eigene Skripte angepasst und erweitert werden.
3. Ein neuer Mitarbeiter der LWL.IT, Marc Malwitz, nahm seine Tätigkeit beim Kundensupport von DIPS.kommunal auf und war bereit, sich in die Frage und in die Programmierung einer geeigneten Exportschnittstelle einzuarbeiten. Malwitz beschäftigte sich zu dem Zeitpunkt bereits mit der Frage einer Web-Archivierung und prüfte, ob Werkzeuge wie Heritrix oder Brozzler erfolgreich zum Einsatz kommen könnten. Auch Twitter-Scraper wie GetOldTweets3<sup>37</sup>, twitter-scraper<sup>38</sup> oder das bereits erwähnte Twint wurden evaluiert. Letztendlich fiel die Wahl wegen der guten Dokumentation, der aktiven Entwicklergemeinschaft hinter dem Projekt und nicht zuletzt wegen der guten Erfahrungen seitens des Stadtarchivs Münster auf Twint.

Schon die ersten Versuche ermöglichten außer dem Tweet selbst eine Ausgabe der Kommentare (zumindest der ers-

ten Ebene) und der angehängten Bilder, die in einem separaten Ordner abgelegt wurden (vgl. Abb. 9). Aus technischer Perspektive erfasst das von Marc Malwitz entwickelte Python-Skript zunächst mittels Twint die Daten der Tweets des zu archivierenden Nutzers. Mit diesen „Rohdaten“ erzeugt das Skript dann einen Zwischenstand in JSON-Dateien, über die anschließend ein Bericht erzeugt und in HTML ausgegeben wird. Die Darstellung ist nicht ganz unähnlich zur Twitter-Timeline und hält sich vor allem strikt an ihre inhaltliche Logik.

Die weitere Entwicklung konzentrierte sich darauf, die Antworten der zweiten (und tieferen) Ebene(n) rekursiv auszulesen und die zentrale JSON-Datei zu einer Baumstruktur mit sämtlichen Tweets auszubauen. Da während des Harvestens viele separate Twint-Durchläufe erforderlich sind, musste die Performanz des Prozesses optimiert werden. Dabei war hilfreich, dass das Skript nun mit einem „Cache“ aus JSON-Dateien arbeitete, um die eingelesenen Inhalte nicht immer erneut erfassen zu müssen.

Eine wesentliche Hürde war die korrekte Zuordnung der Antworten zu den vorherigen Autoren. Da Twint die Abhängigkeiten nicht anhand von Parent-IDs aus dem Quelltext auslesen kann, ist ein Abruf aller Antworten an den entsprechenden Nutzer sowie eine Filterstrategie aus Vergleich der Tweet-ID, der Benutzernamen und der Zeitstempel erforderlich.

36 Für die Langzeitarchivierung könnte man in einem zweiten Schritt aus den JSON-Dateien CSV-Dateien generieren, die leichter verständlich und vermutlich auch leichter zu erhalten sind. Die direkte Erzeugung von CSV aus Twint war wenig überzeugend, da das Komma zur Feld-Separierung eingesetzt wird. Da in den Kommentarfeldern der Twitter-Profilen Satzzeichen verwendet werden (und natürlich auch Kommas!) erhält man „Versprünge“ bei der Spaltenzahl, da die Satzzeichen irrtümlich ebenfalls als Feldtrenner interpretiert werden. Der „Schlenker“ über JSON verhindert das; vgl. dazu Anm. 28.

37 Der Python3-basierte Twitter-Scraper GetOldTweets3 ist eine verbesserte Version der Projekts GetOldTweets von Jefferson Henrique und steht offen kostenfrei unter der MIT-Lizenz zur Verfügung: <https://github.com/Mottl/GetOldTweets3>.

38 Auch das ebenfalls Python3-basierte Projekt twitter-scraper steht kostenfrei unter der MIT-Lizenz bei GitHub zur Verfügung: <https://github.com/bisguzar/twitter-scraper>.

**Markus Lewe** @LeweMarkus

OB Lewe begleitete in seiner Doppelfunktion als Vizepräsident des Deutschen Städtetages und als OB von Münster den deutschen Bundespräsidenten Steinmeier bei einer Kurzreise nach Mailand. Ziel der Reise war die Vertiefung der kommunalen Zusammenarbeit innerhalb der EU, 🇪🇺🇪🇺🇪🇺

6:05 nachm. · 19. Sep. 2020 · Twitter for iPhone

2 Retweets 8 „Gefällt mir“-Angaben

**Lucky Look** @realeyes\_00 · 19. Sep. 2020  
Antwort an @LeweMarkus  
Schwätzen kann er ja. Nur handeln halt nicht. Ganz die Kanzlerin. 🇪🇺

**Tephtor** @tephtor · 19. Sep. 2020  
Antwort an @LeweMarkus  
Mailand oder Madrid - Hauptsache Italien

**Twitter-Export des Nutzers „lewemarkus“**

September 2020 4 Tweets

**Markus Lewe** lewemarkus 1307350050012639236  
19.09.2020, 18:05:15

OB Lewe begleitete in seiner Doppelfunktion als Vizepräsident des Deutschen Städtetages und als OB von Münster den deutschen Bundespräsidenten Steinmeier bei einer Kurzreise nach Mailand. Ziel der Reise war die Vertiefung der kommunalen Zusammenarbeit innerhalb der EU, 🇪🇺🇪🇺🇪🇺 <https://t.co/UzKY7JnrIt>

1307350050012639236.EGj8FYWwAqKtCt.jpg

8 „Gefällt mir“-Angaben · 2 Retweets

**Lucky Look** realeyes\_00 13073509526632673  
19.09.2020, 18:08:52 · Antwort an LeweMarkus

@LeweMarkus Schwätzen kann er ja. Nur handeln halt nicht. Ganz die Kanzlerin. 🇪🇺

1 „Gefällt mir“-Angabe

**Tephtor** tephtor 1307353415593361408  
19.09.2020, 18:18:38 · Antwort an LeweMarkus

@LeweMarkus Mailand oder Madrid - Hauptsache Italien

1 „Gefällt mir“-Angabe

Abb. 10: Original-Tweet mit Kommentaren und der Vergleich zum Twint-Export

### Features des TWINT-basierten Exports

Der verbesserte Export erreicht nun aus archiverischer Sicht sehr gute Ergebnisse, die nahezu alle signifikanten Eigenschaften abdecken:

- Tweets und deren Struktur werden vollständig rekonstruiert.
- Die Darstellung des Beitragstextes mit Hashtags und Links sowie den Gefällt-mir-Angaben und Retweets stellt kein Problem dar.
- Grafiken und Fotos weist das Skript durch den eindeutigen Verweis zur Twitter-ID mit Dateinamen dem jeweiligen Beitrag zu. Notfalls können die Bilder, welche durch den Export im Datenverzeichnis separiert zusammengestellt sind, durch Kennzeichnung der UUID auch manuell einem Tweet zugeordnet werden.
- Kommentare können nun in allen Ebenen chronologisch eingruppiert und durch Einrückung abgebildet werden.

Lediglich bei der Archivierung von bewegten Bildern stieß das Skript an seine Grenzen. Sogenannte animated GIFs werden als Bilddateien ausgegeben, aber eingebundene Video-Dateien kann man auf diese Weise nicht überführen.

Der TWINT-basierte Export liefert eine festgelegte Datenstruktur, welche letztlich die Grundlage zur Langzeitarchi-

Python-Skript „twitter-report.py“

data.json  
twitter-report.lewemarkus.2020-10-30.html  
982665602602012672.DaMhHvTW4AEFXnc.jpg  
982670227057061888.DaM1SpUX4AAN2Hr.jpg  
982670227057061888.DaM1WSFW4AE0KLy.jpg  
984148128919105536.Dahle11XUAE552-.jpg  
...  
hashtag-list.lewemarkus.2020-10-30.txt

Ergänzend

Screenshot\_Beispieltweet\_Lewe.JPG  
Screenshot\_Profil\_Lewe.JPG  
Screenshot\_Tweets\_Lewe.JPG  
Follower\_Lewe\_User.xlsx  
Following\_Lewe\_User.xlsx

Abb. 11: Die beiden für Archivzwecke erzeugten Submission Information Packages (SIP)

vierung bildet. Die Rohdaten werden als JSON gespeichert und Bilder als JPEG, PNG oder GIF separiert. Die gesammelten Hashtags des Exports stellt das Skript in Form einer einfachen Text-Datei zusammen. Diese können als Vorstufe für eine inhaltliche Erschließung beim Ingest genutzt werden. Die von TWINT abgerufenen Daten visualisiert das Skript letztlich im HTML-Format. Die Formatierung orientiert sich zwar an der Dialogstruktur von Twitter, aber die eigentliche Optik sollte durch die Archivierung zusätzlicher Screenshots der Website realisiert werden. Eine übergeordnete CSV-Datei ermöglicht schließlich die Anreicherung der Beiträge mit Metadaten für den Ingest ins Langzeitarchiv. In Abbil-

derung 11 ist der Inhalt dieses ersten für die Archivierung erzeugten Paketes (SIP) schematisch dargestellt.

Ein zweites Paket enthält begleitenden Informationen zum Account wie den Follower- und Following-Listen sowie Screenshots aus dem originalen Account, um die ursprüngliche Anmutung exemplarisch zu überliefern.

## Twitter-Archivierung – eine kontinuierliche Herausforderung

Generell ist zu sagen, dass die Archivierung komplexer Websites wie der Social-Media-Site Twitter ein anspruchsvolles Thema darstellt. Es handelt sich bei derartigen Sites nicht mehr um klassische Webauftritte, die primär der Verbreitung von Informationen dienen. Vielmehr handelt es sich um interaktive Anwendungen, bei denen die Beteiligung der Nutzer als zentrales Element zum Geschäftsmodell gehört. Die Betreiber streben ein Maximum an Bedienungsfreundlichkeit an und wollen ihre Dienste über möglichst viele Kanäle anbieten. Zugleich findet eine starke Personalisierung statt: Inhalte werden in Bezug auf individuelle Interessen der Nutzer gefiltert und sortiert ausgeliefert.

Auf die technische Ebene bezogen bedeutet dies, dass moderne Websites wie Twitter komplexe Gebilde darstellen. Die Oberflächen im Browser bestehen häufig aus modularen, javascript-basierten Frontends, welche schnell an neue Anforderungen und Erwartungen angepasst werden. Die Inhalte werden über Schnittstellen von einer Vielzahl Server bezogen. Selbst wenn keine explizite Anmeldung stattfindet, wird stets eine Sitzung gestartet, um die technische Anmeldung an den Frontend-Schnittstellen zu gewährleisten. Die gleichen Schnittstellen speisen auch Smartphone-Apps oder andere Clients. Das komplette System wird permanent weiterentwickelt und optimiert. Eine derartige Microservice-Architektur erlaubt partielle Aktualisierungen, ohne eine komplett neue Website in Betrieb nehmen zu müssen. Das Gegenstück zu diesem Konzept sind eher monolithische Web-Auftritte, welche fast immer über Content-Management-Systeme betrieben werden.

Für die Technik der Archivierung stellen Web-Architekturen wie bei Twitter eine Herausforderung dar. Das Harvesting-Modul Twint wird stets an die aktuelle Twitterversion angepasst, Änderungen bei Twitter müssen laufend analysiert, der Twint-Code angepasst werden. Dies kann zeitweise zur Fehlfunktion des Skripts führen, bis ein Update der Twint-Entwickler freigegeben und in die Python-Umgebung eingespielt ist. Insofern ist das Skript nie komplett fertig, sondern es bedarf analog zur Website permanent einer gewissen Pflege.<sup>39</sup>

## Fazit

Twitter hat in den letzten Jahren einen festen Stellenwert in der Außendarstellung von (Kommunal-)Politikerinnen und Politikern bekommen. Es ermöglicht eine direkte und hierarchiearme Kontaktaufnahme und kann als Weg der

Bürgerbeteiligung eingesetzt werden. Es lohnt sich, dieses Ausprobieren neuer (basis-)demokratischer Ansätze für die Nachwelt zu erhalten. Dabei ist die Bewertungsentscheidung nicht in Stein gemeißelt: Vielleicht verwendet der jetzige oder der nächste Oberbürgermeister/die nächste Oberbürgermeisterin Twitter nicht mehr, es wird zu einem Kanal, über den nur Pressemitteilungen der Stadt verlautbart werden, oder Twitter wird in seiner Funktion von einer anderen technischen Plattform abgelöst. Schon heute ist zu beobachten, dass Facebook und Instagram stärker eingesetzt werden als Twitter. Deshalb muss das Stadtarchiv seine Archivierungsstrategie rechtzeitig anpassen und überlegen, ob man den Microblogging-Dienst auch für die Folgejahre überliefern will.

Gleichwie diese Entscheidung ausfällt, steht nun eine leicht zu bedienende technische Lösung für jeden nachnutzbar zur Verfügung, mit der alle zentralen Erhaltungsziele abgedeckt werden können.

Weiterhin müssen wir Archive darüber nachdenken, was mit Twitter-Accounts passieren soll, die weniger dem amtlichen und mehr dem politischen Bereich zuzuordnen sind: Viele Ratsfraktionen bestücken regelmäßig ihre Twitter-Accounts; im Wahlkampf pflegen Kandidatinnen und Kandidaten diese Kanäle. Manche politischen Kräfte nutzen das Medium zur politischen Agitation und (Fehl-)Information. Hier fehlt bisher die facharchivische Diskussion, ob diese Inhalte zum Archivierungsauftrag der Kommunalarchive zählen und wie ggf. Auswahlkriterien und Erschließungsrichtlinien (nach Personen, Institutionen oder #hashtags) aussehen könnten.<sup>40</sup> ■



Dr. Peter Worm  
Stadtarchiv Münster  
worm@stadt-muenster.de

<sup>39</sup> Die LWL.IT wird, soweit es die Kapazitäten erlauben, aktualisierte Versionen des Skripts zur Verfügung stellen, ist aber auf Rückmeldungen, Anregungen und Berichten über die durchgeführten Archivierungen angewiesen. Das als ZIP gepackte Skript sowie eine Anleitung unter folgendem Link: <https://cloud.lwl.org/s/PykLmmBBL9cLRay>. Unter der Überschrift „Anleitung/Technische Umsetzung“ hat Marc Malwitz eine detaillierte Installationsanleitung und eine Schritt-für-Schritt-Anleitung für die Durchführung der Archivierung hinterlegt: <https://archive20.hypotheses.org/10031>.

<sup>40</sup> Wie solche ereignisbezogenen Dokumentationsstrategien aussehen können und wo sie ihre Grenzen haben, diskutiert Mirjam Schaap vom Stadsarchief Amsterdam, siehe in diesem Heft Markus Stumpf, Tagungsbericht, S. 4. Dass von Seiten der Zeithistorikerinnen und -historiker Interesse an der Sicherung von Inhalten aus Social Media und städtischen Internetseiten besteht belegt der Beitrag von Malte Thießen in diesem Heft, S. 40–46.