

Aus dem Netz!¹

Weblogs und Netzliteratur im Archiv mit einem Ausblick auf die Archivierbarkeit Sozialer Medien

Laura Pohlmann/Karin Schmidgall/Jochen Walter

Im Juli 2021 verkündete das Auktionshaus Sotheby's die Versteigerung von 9.555 Zeilen Programmcode für HTML und HTTP – den Quellcode des WWW – samt Bedienungsanleitung für 5,4 Millionen US-Dollar.² Bei dem digitalen Artefakt handelte es sich um Originaldateien mit Zeitstempel und Unterschrift von Tim Berners-Lee, der mit seinem Programmcode den Grundstein für das World Wide Web legte.

Während sich das Internet, das vor nunmehr 30 Jahren mit dem erfolgreichen Quellcode seinen Anfang nahm, als allumgebende Größe quasi dauerhaft in unserem Lebensalltag etabliert hat, ist das, was seither in ihm produziert oder ausgestellt wurde, bekanntermaßen alles Mögliche, nur nicht von

-
- 1 Einige Textabschnitte des vorliegenden Beitrags wurden teilweise übernommen und paraphrasiert aus dem Artikel »Literatur im Netz. Sammeln, Erschließen, Archivieren« von Karin Schmidgall und Jochen Walter aus dem Jahre 2010. Der Artikel erschien u.a. in: *Siegener Periodicum zur Internationalen Empirischen Literaturwissenschaft (SPIEL)* 29/1-2 (2010) (= Sonderheft: Florian Hartling/Beat Suter (Hg.): *Archivierung von digitaler Literatur: Probleme – Tendenzen – Perspektiven*), S. 223-234. Es sei außerdem auf den in Kürze erscheinenden Beitrag »Die Sammlung Literatur im Netz im Deutschen Literaturarchiv Marbach« im Rahmen des SDC4lit-Projekts verwiesen, welcher ebenfalls die Archivierung von Netzliteratur und Weblogs am Deutschen Literaturarchiv thematisiert. Vgl. o.V.: »Science Data Center for Literature«. In: *SDC4Lit*. <https://www.sdc4lit.de/>, o.J. [zuletzt eingesehen am 2.5.2022].
 - 2 Vgl. o.V.: »Versteigerung bei Sotheby's. 5,4 Millionen Dollar für die World-Wide-Web-Gründungsdokumente«. In: *FAZ.net*. <https://www.faz.net/aktuell/wirtschaft/digitec/sotheby-s-5-4-millionen-dollar-fuer-die-world-wide-web-gruendungsdokumente-17416379.html>, 1.7.2021[zuletzt eingesehen am 2.5.2022].

dauerhafter, geschweige denn kontrollierter Verfügbarkeit.³ Davon sind auch Literatur- und Kulturproduktionen, die (allein) für das Internet konzipiert und in ihm ausgestellt werden und die – denkt man an ihre papiernen Artgenossen – in besonderen Maße als archivierungswürdig gelten, nicht ausgenommen.⁴ Wer aber kümmert sich um die Bewahrung? Die Bewahrung von was genau? Und wie? Während die literaturwissenschaftliche Forschung nunmehr vermehrt den Fokus darauf lenkt, was – z.B. im Sinne neuer »Schreibweisen der Gegenwart« – im Netz produziert wird, hat die Frage nach einer adäquaten nachträglichen Archivierung der Artefakte seit 30 Jahren kaum mehr als gleichermaßen an Form wie an Dringlichkeit gewonnen.

Das Deutsche Literaturarchiv sammelt, erschließt und bewahrt handschriftliche, gedruckte, audiovisuelle und eben elektronische Materialien zur deutschsprachigen Literatur. Die Bibliothek des Deutschen Literaturarchivs versteht sich dabei schwerpunktmäßig als Quellenbibliothek, in deren Zentrum die Autorendokumentation (von der Erstausgabe bis zum Taschenbuch) steht, flankiert durch umfangreiche Sammlungen zur Verlagsproduktion (Buchumschläge, Verlagsprospekte) sowie die einschlägigen Forschungspublikationen zu den Autor*innen und zu kultur- und literaturwissenschaftlichen Forschungsbereichen allgemein. Hinzu kommen audiovisuelle Quellen, wie Autorenlesungen und die Dokumentation weiterer Rezeptionszeugnisse, die Theaterprogrammhefte, Drehbücher, Hörfunkmanuskripte und Zeitungsartikel enthält. Das Literaturarchiv konnte für das Einsammeln und Erhalten von so »flüchtigem« Material, wie Zeitungsausschnitten, also bereits eine gewisse Erfahrung einbringen, als sich in den 1990er Jahren ein weiteres Phänomen im Kreis der transitorischen Zeugnisse und Artefakte abzeichnete. Mit dem literarischen und auch literaturwissenschaftlichen Material, das sich im Laufe der 1990er Jahre nicht nur mit, sondern vielmehr durch Berners-Lees HTML-Code verbreitete, stand das Archiv jedoch vor einer gänzlich unbekannten Herausforderung, die nicht allein in der kurzen Verfügbarkeit des literarischen Materials, sondern ebenso in der technischen Herausforderung einer Archivierung der Quellen bestand, die mit neuen und häufig dynamischen Darstellungsmöglichkeiten experimentierten und ein (auch für das Archiv noch) ungewohntes Medium bedienten.

3 Vgl. bspw. Michael Hollmann/Andre Schüller-Zwierlein (Hg.): *Diachrone Zugänglichkeit als Prozess. Kulturelle Überlieferung in systematischer Sicht*. Berlin u.a. 2014.

4 Vgl. bspw. Caroline Y. Robertson-von Trotha/Robert Hauser (Hg.): *Neues Erbe. Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung*. Karlsruhe 2011.

Als das Deutsche Literaturarchiv in den 2000er Jahren die Einsendungen des Pegasus-Wettbewerbs für Netzliteratur (s.u.) auf CD-ROM erwarb, wusste man zunächst nichts damit anzufangen. Zufallsfunde von literarischen Beiträgen im Internet druckte man aus und archivierte das Papier in einem der für Marbach typischen grünen Archivkästen. Zu diesem Zeitpunkt fehlte dem Archiv schlicht noch eine Infrastruktur für die medienspezifische Archivierung von Quellen dieser Art, die sich erst zu entwickeln begonnen hatte.⁵

Als 2008 schließlich mit der Archivierung von Werken digitaler Literatur durch das Archiv begonnen wurde, war das Internet längst im »Zeitalter seiner Kommerzialisierung« angekommen, in dem nicht mehr die »anarchistischen und utopischen Kommunikationsformen des Anfangs«, sondern »neue Vertriebswege für altbekannte Produkte« wie »Informationen, Flugtickets, Autos, Immobilien, Musik, Bücher« im Mittelpunkt standen.⁶ Ältere, experimentelle Literaturformen aus der Frühzeit des Internets waren zu diesem Zeitpunkt kaum noch erhalten, dafür – vor allem mit den Weblogs – aber andere entstanden.

In Marbach entschied man sich schließlich für zwei Archivierungslinien. Zum einen die *Literatur im Netz*, womit sowohl Texte gemeint waren, die nicht zwingend eine digitale und vernetzte Umgebung benötigten (z.B. Online-Zeitschriften), als auch literarische Weblogs. Beides war »üblicherweise mit standardisierter Software erstellt und daher leichter zu archivieren«.⁷ Die zweite Linie verfolgte die Archivierung der *Netzliteratur*, jenen »anarchistischen und utopischen Kommunikationsformen des Anfangs«, ⁸ die mehr Vorlauf benötigte und nur als finanziell gefördertes Drittmittelprojekt durchgeführt werden konnte – und mit Unterstützung der Netzautor*innen.

Schon früh wurde die Netzliteratur als »Strohfeuer in den Medien« bezeichnet, es sei »kein Zufall, dass mit dem Niedergang der Netzliteratur [...]

5 Noch im selben Jahr begannen Mitarbeiter*innen des Archivs jedoch gemeinsam mit Oliver Gassner, dem Mitorganisator der Netzliteraturwettbewerbe in Ettlingen, über die Möglichkeit zu diskutieren, das Literaturarchiv Marbach als Standort einer Archivierung von Quellenliteratur aus dem Internet auszuweiten.

6 Roberto Simanowski: »Autorenschaft in digitalen Medien«. In: *Digitale Literatur* Band 152 *Text+Kritik* (2001), S. 3-21, hier S. 3.

7 Jutta Bendt: »Im Netz! Digitale Quellen im Archiv«. In: Jutta Bendt (Hg.): *Netzliteratur im Archiv. Erfahrungen und Perspektiven*. Göttingen 2017, S. 7-11, hier S. 8. Zur Auswahl der Quellen, den technischen Voraussetzungen für die Archivierung sowie zu den rechtlichen Aspekten vgl. Schmidgall/Walter: »Literatur im Netz«.

8 Simanowski: »Autorenschaft in digitalen Medien«, S. 3.

die Erfolgsgeschichte der Blogs beginnt«. ⁹ Nur knapp zehn Jahre seit ihrer Entstehung waren für die Bereitstellung von Netzliteraturprojekten aus der Frühphase schon Ausgrabungsarbeiten notwendig, die sich sicher auch dadurch verstärkten, dass die Produktionen kaum je kommerzialisiert wurden, geschweige denn eine institutionelle Anbindung erfahren hatten. Letztendlich waren es am ehesten die Netzautor*innen selbst, die sich für eine Erhaltung ihrer Werke bemühten.

Heute sind auf der Plattform Literatur im Netz ¹⁰ ca. 500 archivierte Quellen für die Forschung frei zugänglich. Die Sammlung umfasst: 93 literarische Zeitschriften, 334 literarische Blogs und knapp 60 Netzliteratur-Objekte. Das DLA verfolgt den Anspruch, die Authentizität und Integrität der Quellen im Archiv, also das »Look and Feel« der originalen Internetquellen zu erhalten.

Die Sicherung der Quellen stellt ein Archiv vor eine Reihe von Anforderungen: urheberrechtliche, organisatorische, bibliothekarische, in erster Linie aber technische Anforderungen, die der jeweiligen Quelle immanent sind. Da bei digitaler Literatur nicht mehr zwischen Inhalt und Software unterschieden werden kann, muss eine Archivierung immer dem Werk und der Präsentationsumgebung angemessen sein.

Am Beispiel der Netzliteratur und der Weblogs werden Einblicke in den Aufbau der Sammlung und die praktische Tätigkeit der (Web)archivierung gegeben, daraus folgt eine Einschätzung, ob die gewählten Methoden sich auch zur Archivierung sozialer Medien eignen.

Die Archivierung von Netzliteratur

Zwischen 2013 und 2016 konnte das Deutsche Literaturarchiv in dem von der Deutschen Forschungsgemeinschaft geförderten Projekt »Netzliteratur authentisch archivieren« in enger Zusammenarbeit mit Netzautor*innen und Forscher*innen einen in technischer und ästhetischer Hinsicht typologisch

9 Don Alphonso: »Warum diese Blogger? Warum nicht die anderen?«. In: Don Alphonso/ Kai Pahl (Hg.): *Blogs! Text und Form im Internet*. Berlin 2004. S. 17–21, hier S. 19. Für die Blogs wurden zunehmend Weblog-Dienste wie Wordpress (ab 2004) oder Tumblr (ab 2007) genutzt. Insgesamt beruht ihr Erfolg auch darauf, dass sie die Einstiegshürden absenkten. »Sie sind einfach zu bedienen, praktisch kostenlos, und jeder Nutzer kann ohne Vorkenntnisse alles veröffentlichen«, so Alphonso. Ebd.

10 o.V.: »Literatur im Netz«. In: *DLA Marbach*. <http://literatur-im-netz.dla-marbach.de/>, 2009 [zuletzt eingesehen am 2.5.2022].

beispielhaften Grundstock an Quellen aufbauen.¹¹ Das Corpus von über 50 Werken ist zwar keinesfalls vollständig, vereinigt aber Beispiele sämtlicher literarischer Gattungen aus allen technischen und ästhetischen Entwicklungsstufen der frühen bis späten Netzliteratur: Monographisch orientierte, sowie kollaborativ entstandene Werke, hypertextuelle ›Eintagsfliegen‹ ebenso wie verschiedene, gleichermaßen anspruchsvolle Werke einzelner Netzautor*innen.¹² Einen Orientierungspunkt für die Quellenauswahl lieferten verschiedenen Wettbewerbe¹³ und ihre Preisträger, sofern man als Archiv der einflussreichsten Werke noch ›habhaft‹ werden konnte.

Da sich seit Beginn der Archivierung schon viele Webseiten nicht mehr im Netz befanden und (mehr noch) auch weitere Webseiten seither nicht mehr zugänglich sind, hat das Archiv nicht nur die Funktion einer zentralen Nachweisstätte, sondern wird zur zuletzt verbliebenen ›primären‹ Umgebung, zum Schauplatz des Werks. Für das Archiv eine ungewöhnliche Position, gilt doch hier in besonderem Maße die Anforderung, dass »die Archivierung das Ereignis im gleichen Maße hervorbringt, wie sie es aufzeichnet.«¹⁴ Wie aber können die Werke der Netzliteratur authentisch archiviert und wiedergegeben werden, wenn die Authentizität der archivierten Objekte vor allem davon abhängt, was Autor*innen und Nutzer*innen als deren relevante Eigenschaften (›Significant Properties‹) ansehen? Und wie kann es gelingen, diese im Archivierungsprozess und auch zukünftig zu bewahren und korrekt wiederzugeben? In der Praxis bedeutet dies vor allem eines: dass der Aufwand, Quellen möglichst ›authentisch‹ in das Archiv einzubringen, sehr hoch ist.¹⁵ Und sie ›authentisch‹ zu erhalten auch.

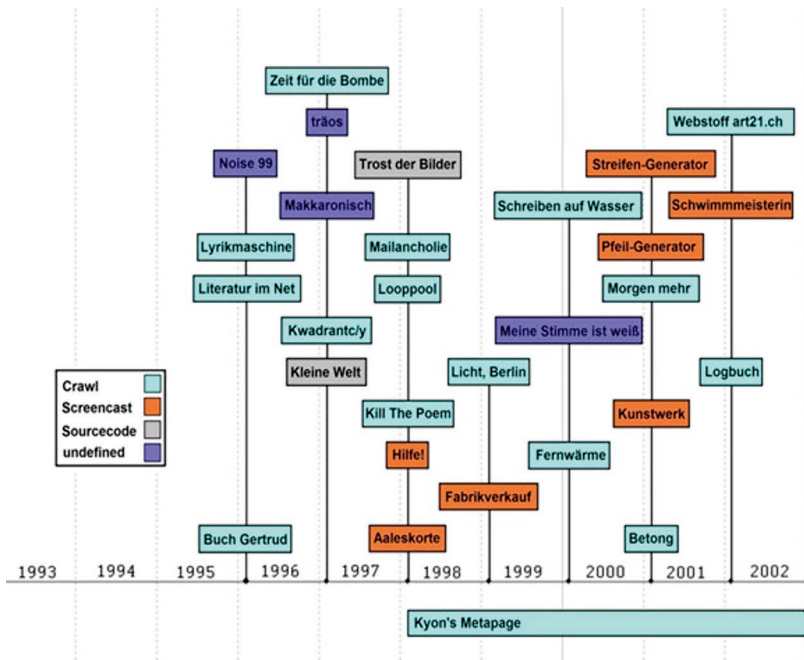
11 Am 2. Dezember 2015 fand in Marbach die Abschlusstagung des Projekts statt, vgl. hierzu den Sammelband Bendt (Hg.): *Netzliteratur im Archiv*.

12 Weitere 25 Werke konnten im Wiki (s.u.) erfasst und beschrieben, aufgrund fehlender Genehmigungen aber nicht archiviert werden.

13 Wie der erste Pegasus-Wettbewerb für Internet-Literatur 1996 oder der Junggesellenpreis für Netzliteratur von 2005.

14 Jacques Derrida: *Dem Archiv verschrieben. Eine Freudsche Impression*. Berlin 1997, S. 35.

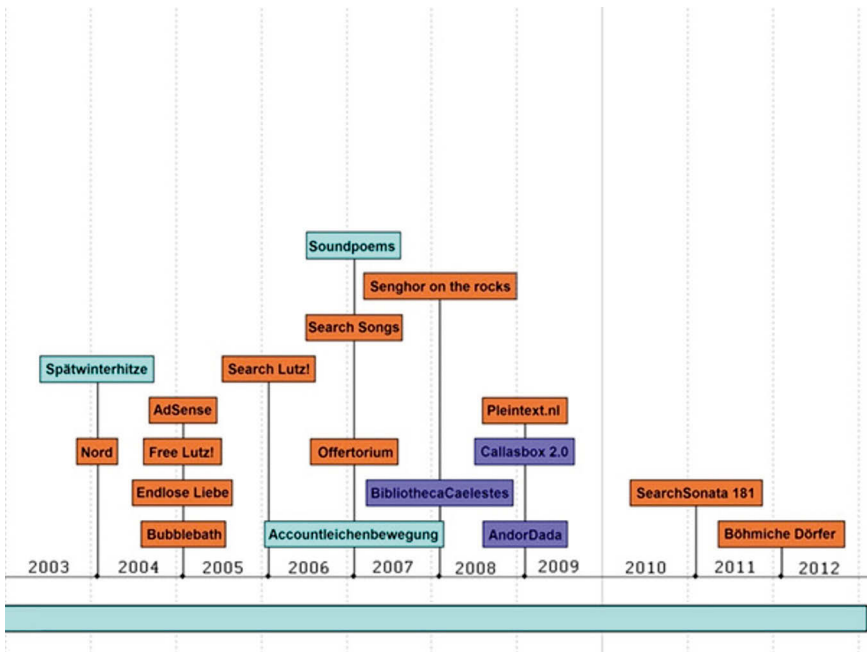
15 Im Projekt wurde ein Workflow für das Einbringen der Quellen ins Archiv entwickelt, als Archivformat ist BagIt gewählt worden, vgl. Steffen Fritz: »Die Anwendung des BagIt-Formats im Deutschen Literaturarchiv Marbach«. In: Bendt (Hg.): *Netzliteratur im Archiv*, S. 63-75. Die für den Workflow entwickelten modularen Softwarekomponenten liegen auf GitHub: o.V.: »NetLit DLA Marbach«. In: *GitHub*. <https://github.com/netzliteratur/>, 2015 [zuletzt eingesehen am 2.5.2022].

Abbildung 16a: Das Corpus Netzliteratur (Zeitstrahl), Foto: DLA Marbach¹⁶

Die Fragen, welche Eigenschaften für ein Werk relevant sind und welche Methode sich eignet, um Layout, Texte, Bilder, einzelne Features oder die Möglichkeit zur Interaktivität im Archiv zu erhalten, wurden schließlich mit Autor*innen anhand ausgewählter Werke auf einem Projekt-Workshop diskutiert.¹⁷ Die Netzautorin Susanne Berkenheger, bekannt für ihre preisge-

16 Vgl. o.V.: »Zeitgeschichte«. In: *DLA Marbach Wiki*. <https://wwik.dla-marbach.de/line/index.php/Zeitgeschichte>, 2015 [zuletzt eingesehen am 2.5.2022].

17 Der Workshop »Netzliteratur im Archiv: Identifikation signifikanter Eigenschaften« fand am 11.2.2015 im Deutschen Literaturarchiv Marbach statt, das Protokoll ist veröffentlicht: Steffen Fritz/Stephanie Kuch: »Protokoll Workshop Netzliteratur: Identifikation signifikanter Eigenschaften«. In: *DLA Marbach*. https://wwik-prod.dla-marbach.de/line/images/7/7c/Protokoll-workshop_final.pdf, 19.3.2015 ebenso die Projektpräsentation: o.V.: »Workshop Netzliteratur im Archiv: Identifikation signifikanter Eigenschaften«. In: *DLA Marbach*. https://wwik-prod.dla-marbach.de/line/images/8/8b/Praesentation_Workshop_Projekt.pdf, 2015 [zuletzt eingesehen am 2.5.2022].

Abbildung 16b Das Corpus Netzliteratur (Zeitstrahl), Foto: DLA Marbach¹⁸

krönten Werke »Hilfe!« (1999), »Zeit für die Bombe« (1997) und »Die Schwimmmeisterin« (2002), brachte für ihre Workshop-Präsentation »Der Alptraum des Archivars: Die Schwimmmeisterin – das Original« ihr historisches Laptop (Windows 98) mit, auf dem das Werk 2001/2002 entstanden war. Die signifikanten Eigenschaften des Werks waren aus ihrer Sicht allerdings schwer zu definieren, vielmehr könnte die Vernachlässigung einer jeden Eigenschaft unerwartete Folgefehler nach sich ziehen, beispielsweise beeinträchtigte die Änderung des Fonts oder der Schriftgröße die Pop-Ups und die »Leseerfahrung«. Gemeinsam mit der Autorin konnte jedoch festgestellt werden, dass

18 Vgl. o.V.: »Zeitgeschichte«. In: *DLA Marbach Wiki*. <https://wwik.dla-marbach.de/line/index.php/Zeitgeschichte>, 2015 [zuletzt eingesehen am 2.5.2022].

aktuelle Hard- und Software dem Treiben im Schwimmbad eine zu hohe Geschwindigkeit verliehen, so dass für die Erstellung eines Screencasts eine besser geeignete, langsamere Umgebung emuliert wurde.¹⁹

Im ›Online-Musical‹ »Endlose Liebe« (2005), von Frank Klötgen gehören, wie unten im Text beschrieben, die Texte und Musikstücke, die graphische Gestaltung und die Interaktion zu den signifikanten Eigenschaften.²⁰

Endlose Liebe ist ausgestattet mit 19 Liebes- und Leidensliedern, die Klötgen selbst mit der Band Marilyn's Army singt. Klangvolle und skurrile Requisiten ergänzen die reduzierte Bühnenlandschaft des als ›Trashical‹ bezeichneten Online-Musicals. [...] **Endlose Liebe – Endless Love** ist in mehrfacher Hinsicht ein doppelbödiges Theaterstück; es benutzt die performativen Mittel des Computers und inszeniert den Bildschirm auf einfache Weise als Guckkastent Bühne. Hier nun spielen sich die Szenen und Akte des Stücks ab. Die Figuren sind allesamt lediglich weiß auf schwarzem Grund skizziert. Einzelne Dialoge erhalten ihren eigenen Rahmen in Pop-up-Fenstern. Die sieben Protagonisten des Stücks liefern sich amüsante Liebeskämpfe und Verwicklungen bis zum Mord. Der Ausgang aus dem Irrgarten der Gefühle ist nicht leicht zu finden. Die Geschichte kennt mehrere unterschiedliche Verläufe.²¹

-
- 19 Über den Link www.berkenheger.netzliteratur.net/ouargla/websprudel/ gelangt man zu einer noch abspielbaren Version mit Warnhinweisen: »Sollten Sie ein Faible für digitale Ruinen haben, können Sie auch direkt von Ihrer aktuellen Plattform aus über den Zaun springen und versuchen, die herumfliegenden Trümmer der ›Schwimmeisterin‹ ohne archäologischen Schutzanzug zu genießen.« Susanne Berkenheger: »Die Schwimmeisterin«. In: *Berkenheger.Netzliteratur.net*, 2002 [zuletzt eingesehen am 2.5.2022]. In ihrem Artikel »Die Schwimmeisterin oder ›Literatur im Netz ist eine Zumutung« geht Susanne Berkenheger auf die Browserabhängigkeit des Werks ein: In: *Dichtung-Digital*. www.dichtung-digital.de/2002/modemfieber/berkenheger.htm, 2002 [zuletzt eingesehen am 2.5.2022].
- 20 Wie auch bei der »Schwimmeisterin« spielen Fonts und ursprüngliche Hardware hinsichtlich der Auflösung eine wesentliche Rolle. Vgl. Fritz/Kuch: »Protokoll Workshop Netzliteratur«.
- 21 Aus der lobenden Erwähnung der Jury des 1. Jungesellenpreis für Netzliteratur (s.o.), Florian Höllerer u.a.: »1. Jungesellenpreis für Netzliteratur: Begründung der Jury«. In: *Jungesellenpreis.de*. www.jungesellenpreis.de/jury.html, 2005 [zuletzt eingesehen am 2.5.2022].

Die Live-Version²² von »Endlose Liebe« ist im Netz spielbar, allerdings:

Die Browser-Versionen mobiler Geräte schicken *Endlose Liebe* fast ausnahmslos in frühe Sackgassen. [...] Kurzum, der Spielspaß bei *Endlose Liebe* kommt innerhalb kürzester Zeit zum Erliegen. [...] In den Standardbrowsern der aktuellen mobilen Endgeräte werden Flash-Filme im blinden Frame nicht mehr alleine abgespielt, wodurch das Musical *Endlose Liebe* zu einem Musical ohne Musik wird.²³

Frank Klötgen schildert hier sehr bildhaft die Herausforderungen, vor denen nicht nur Leser*innen, sondern besonders auch Archivar*innen stehen, wenn sie eine Internetquelle wenige Jahre nach dem Entstehen des Werks authentisch ins Archiv einbringen wollen. Die ›Flüchtigkeit‹ der Netzquellen zeigt sich hier gleich in doppelter Hinsicht: Zum einen in der Fragilität ihrer Komposition, die in einer sich rasch wandelnden technischen Infrastruktur schnell ihre Haltbarkeit verliert. Zum anderen in ihrer Komposition selbst: die geplante (bzw. programmierte) Inszenierung von Handlungsalternativen qua Navigation, macht jeden ›Durchgang‹ durch das Werk zu einem nahezu einmaligen Ereignis, das sich einer Wiederholbarkeit entzieht oder dieser widerspricht.

Im Projekt wurden letztendlich vier (Web-)Archivierungsmethoden angewendet: Crawl (Spiegelung mittels eines Webcrawlers),²⁴ Screencast, Screenshot und die Archivierung des Quellcodes und ggf. der Datenbank selbst.²⁵

22 Frank Klötgen: »Endlose Liebe – Endless Love«. In: *Hirnpoma.de* www.hirnpoma.de/trashical/, 2005 [zuletzt eingesehen am 2.5.2022].

23 Frank Klötgen: »Die Endlose Liebe in aller Kürze – unter besonderer Berücksichtigung der Endlichkeit hypermedialer Gestaltungsmerkmale«. In: Bendt (Hg.): *Netzliteratur im Archiv*, S. 63–73.

24 Mit Hilfe eines Webcrawlers, z.B. Heritrix werden von einer Ausgangs-URL ausgehend die Inhalte einer Webseite eingesammelt und ein Archivobjekt erzeugt, für das sich als internationaler Standard das WARC-Format durchgesetzt hat. Vgl. Andreas Rauber/Hans Liegmann: »17.9 Web-Archivierung zur Langzeiterhaltung von Internet-Dokumenten«. In: Heike Neuroth u.a. [Hg.]: *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0*. E-Book. Göttingen 2009, o.S. Die von Rhizome entwickelten Webrecorder-Technologie ermöglicht es auch Interaktivität (klicken, scrollen etc.) aufzuzeichnen und damit die Archivierung von Social Media zu verbessern, vgl. o.V.: »Conifer«. In: *Rhizome.org*. <https://conifer.rhizome.org/>, o.J. [zuletzt eingesehen am 2.5.2022].

25 Insgesamt konnten zu 25 Werken die Quelltexte archiviert werden.

Allerdings stieß der Crawl sehr schnell an seine Grenzen, wenn die Aufführbarkeit eines Werks serverseitige Funktionen, wie Deep Web-Abfragen benötigte (wie der »Assoziationsblaster« (seit 1999) von Dragan Espenschied und Alvar Freude), bestimmte Systemanforderungen stellte (wie bei »Endlose Liebe«) oder wenn externe Datenquellen einbezogen wurden (wie bei »search lutz!« (2006) von Johannes Auer).²⁶

Im Projektwiki werden im sogenannten Corpus I insgesamt 71 Werke beschrieben,²⁷ die zum Projektzeitpunkt noch online waren. Für 51 Werke erhielt das Archiv von den Autor*innen die Rechte zur Speicherung der archivierten Fassung und zur öffentlichen Präsentation auf der Archivierungsplattform »Literatur im Netz«.²⁸ Insgesamt 43 Werke konnten über einen Crawl als Spiegelung ins Archiv eingebracht werden, allerdings ergab eine Prüfung der Spiegelfähigkeit,²⁹ dass nur 24 Werke mittels der Spiegelung vollständig erfasst werden konnten und auch bei diesen die Darstellung unter Verwendung aktueller Hard- und Software oft nicht mehr als authentisch bewertet werden kann, das »Look & Feel« des Originals in der Archivkopie fehlte.

Aus diesem Grund wurde im Projekt ein Schwerpunkt auf die Dokumentation der einzelnen Werke gelegt: Jeder Eintrag enthält eine kurze Beschreibung des Werks, die durch die jeweilige Autorin bzw. den jeweiligen Autor selbst erstellt wurde. Ergänzt wurden außerdem Hinweise auf Sekundärliteratur,³⁰ sowie Metadaten nach internationalen Standards.³¹

26 In seiner Präsentation »Netzliteratur in Archiven: Von der technischen Analyse zur Emulation« skizziert Steffen Fritz alternative Lösungsansätze und nennt Beispiele für die Crawl-Grenzen. In: *DLA Marbach*. https://wwik-prod.dla-marbach.de/line/images/8/87/Projekt_Workshop_20140611.pdf, 2014 [zuletzt eingesehen am 2.5.2022].

27 o.V.: »Kategorie: Corpus I«. In: *DLA Marburg Wiki*. https://wwik.dla-marbach.de/line/index.php/Kategorie:Corpus_I, 2015 [zuletzt eingesehen am 2.5.2022].

28 o.V.: »Literatur im Netz«.

29 Zu jeder Quelle wurde ein Bericht mit Beurteilung der Spiegelfähigkeit erstellt, vgl. Stefanie Kuch: »Überprüfung der Spiegelfähigkeit von Internetressourcen mit dem Online-Tool ArchiveReady und Bewertung der dabei analysierten Facetten«. In: *DLA Marbach*. <https://wwik-prod.dla-marbach.de/line/images/9/9b/ArchiveReady.pdf>, 2016 [zuletzt eingesehen am 2.5.2022].

30 Das Projekt-Wiki bietet einen guten Einstieg in das Genre der Netzliteratur. Jeder Eintrag eines Werkes enthält Angaben zur Entstehungsgeschichte, zu Rezeption des Werks in der Literatur, zu benötigter Hard- und Software sowie eine Liste der Sekundärliteratur zum Werk.

31 Für die bibliographisch-deskriptive Beschreibung wie Titel, Autor, Entstehungsjahr, URLs, wird MODS verwendet. Technische Metadaten, wie Dateiformat, Größe, Hash-

Zur Dokumentation zählt auch die Erstellung eines Screenshots der Startseite, vor allem aber von Screencasts. Für Werke wie die »Schwimmmeisterin« von Susanne Berkenheger oder »Endlose Liebe« von Frank Klötgen wurde für die Videoaufzeichnung die technische Abspielumgebung ihrer Entstehungszeit auf einer virtuellen Maschine emuliert. Nur so war es auch möglich die komplexen Handlungsstränge des Werks »Endlose Liebe« zu analysieren und in Zusammenarbeit mit dem Autor denjenigen Handlungsstrang zu identifizieren, der den Grundgedanken hinter dem Werk am besten vermittelt.³² Auch wenn der Screencast für die Leseerfahrung des Werks uninteressant ist, eignet sich das Verfahren zur Dokumentation und für eine spätere authentische Rekonstruktion auf Grundlage des Quellcodes.

Das aufwändige Verfahren der Rekonstruktion der technischen Abspielumgebung und Screencast-Erstellung musste aus Zeitgründen auf wenige Werke beschränkt bleiben. Wobei sich rückblickend erweist, dass ein Screencast auch für weitere Quellen hätte erstellt werden müssen, die zum Zeitpunkt des Einbringens ins Archiv noch anders beurteilt wurden. Denn der Zahn der Zeit nagt auch an den Archiv-Versionen und produziert neue Fragen: passt der Archivierungsbericht nach wenigen Jahren noch mit der archivierten Quelle überein?³³

summe, Speicherort, benötigte Hard- und Software etc. und administrativ-rechtliche Metadaten werden mit PREMIS beschrieben. Vgl. Stephanie Kuch: »Netzliteratur und Metadaten. METS, MODS & PREMIS und ihre Anwendung im Projekt »Aufbau eines Quellenrepertorios für die seit den 1990er-Jahren entstehende Literaturgattung Netzliteratur«. In: Bendt (Hg.): *Netzliteratur im Archiv*, S. 75-92. Im Projekt wurde ein Application Profile entwickelt und veröffentlicht, das die im Rahmen des Projekts verwendeten Standards, die angewendeten Elemente und deren Belegungsmöglichkeiten genau spezifiziert. Es kann online eingesehen werden: Stephanie Kuch: »DLA – Application Profile, Version 3«. In: *DLA Marbach*. https://wwik-prod.dla-marbach.de/line/images/4/4d/Application_profile_V3.pdf, 2015 [zuletzt eingesehen am 2.5.2022].

32 Das Ablaufdiagramm der Handlungsstränge ist im Projektwiki veröffentlicht: o.V.: »Handlungsstränge«. In: *DLA Marbach Wiki*. <https://wwik.dla-marbach.de/line/index.php/Datei:Handlungsstr%C3%A4nge.jpg>, 2016 [zuletzt eingesehen am 2.5.2022].

33 Leider muss man dies verneinen, beispielsweise ist die Spiegelung des Werks »search lutz!« von Johannes Auer im Archiv nicht mehr aufrufbar: o.V.: »search lutz! [[Elektronische Ressource]]/Johannes Auer«. In: *DLA Marbach. Literatur im Netz*. <http://literatur-im-netz.dla-marbach.de/bsz396881904.html>, 2015 und der im Wiki dokumentierte Bericht zur Beurteilung der Spiegelung ist nicht mehr nachvollziehbar: o.V.: »Search lutz!«. In: *DLA Marbach Wiki*. https://wwik.dla-marbach.de/line/index.php/Search_lutz!, 2016 [beide zuletzt eingesehen am 2.5.2022].

Das Corpus II enthält vier Werke, die zum Zeitpunkt ihrer Archivierung nur noch offline verfügbar waren.³⁴ Hier war als Methode die Archivierung des Quellcodes selbst relevant und die Rekonstruktion extrem aufwändig. Exemplarisch wurde das Werk »Schwamm« von Detlev Fischer, das bereits Ende der 1980er Jahre mit der Apple-Software Hypercard entwickelt wurde, in einer lokalen Emulation des Betriebssystems 7 von Apple nahezu komplett wieder abspielbar gemacht. Der Autor übersandte dafür dem DLA Disketten, von denen man befürchtete, dass sie zu diesem Zeitpunkt schon nicht mehr lesbar wären.³⁵ Für »Die Quotenmaschine« (1995) von Norman Ohler wurden dagegen die Texte digitalisiert, die (mittlerweile) auch in Buchform publiziert waren, und die gemeinsam mit dem Autor neu angeordnet und mit einer interaktiven Karte verknüpft wurden.³⁶ Es war ein Probelauf dafür, welches Vorgehen geeignet ist, um ein Werk, von dem nur noch unvollständige Daten vorhanden sind, wieder »herzustellen« und neu zu beleben.

Auch wenn im Projekt nicht jede Quelle mit den heutigen technischen Mitteln authentisch archiviert werden konnte, dient die systematische Analyse und Dokumentation der Entstehungs-, Abspiel- (Aufführungs-) und Spielumgebung nach festgelegten Standards als Basis für zukünftige Forschungsprojekte im Bereich der Entwicklung von Emulations-Strategien im Rahmen der Langzeitarchivierung.

34 o.V.: »Kategorie:Corpus II«. In: *DLA Marbach Wiki*. https://wwik.dla-marbach.de/line/index.php/Kategorie:Corpus_II, 2015 [zuletzt eingesehen am 2.5.2022].

35 Ein Teil des Werkes auf den Datenträgern ist auch nach Einschätzung des Archivs unwiederbringlich beschädigt, ein Teil konnte jedoch rekonstruiert werden. Vgl. auch Detlev Fischer: »Schwammfragmente wieder ansehen«. In: *Oturn.net*. www.oturn.net/work/schwamm/schwammwiederansehen.pdf, 2006 [zuletzt eingesehen am 2.5.2022]. Das Werk im Archiv: o.V.: »Schwamm«. In: *DLA Marbach Wiki*. <https://wwik.dla-marbach.de/line/index.php/Schwamm>, 2016 [zuletzt eingesehen am 2.5.2022].

36 Das Ergebnis ist veröffentlicht unter: o.V.: »Quotenmaschine 2.0 [[Elektronische Resource]]/Norman Ohler«. In: *DLA Marbach. Literatur im Netz*. <http://literatur-im-netz.dla-marbach.de/bsz444605606.html>, 2015 veröffentlicht.

Die Archivierung von Weblogs

Neben dem Corpus der *Netzliteratur* befinden sich über 330 Weblogs im Archiv.³⁷ Aus technischer Sicht scheint die Archivierung von Weblogs weit weniger anspruchsvoll. Weblogs basieren immer auf einem Content-Management-System (CMS) und entsprechenden Datenbanken, sie haben statische und dynamische Bestandteile. Zu den statischen gehören die Beiträge und die Gestaltungsanweisungen, zu den dynamischen Bestandteilen die Kommentar- und die Suchfunktionen sowie andere interaktive Funktionen, wie z.B. der scrollabhängige automatische Seitenaufbau, wie man ihn auch von Twitter her kennt. Des Weiteren binden Blogs oft externe Inhalte ein, wie Youtube-Videos (die aus rechtlichen Gründen nicht archiviert werden dürfen). Was sich für den Benutzer als Einheit präsentiert, ist unter der Oberfläche ein Mosaik aus verstreuten Teilchen, die für eine Archivkopie zusammengeführt werden müssen.

Weblogs werden mit dem Webcrawling-Tool Heritrix eingesammelt. Bei der Durchführung eines Crawls ist zu beachten, dass das Ergebnis immer durch Parameter, wie die Ausgangsadresse der Quelle, Filter für ein- bzw. auszuschließende Adressen, zeitliche und quantitative Limitierungen, interne und externe Spiegelungstiefe beeinflusst wird. Daraus folgt, dass die zu spiegelnde Vorlage im Gegensatz zu konventionellen Medien in ihrem Umfang nicht definiert ist. Vielmehr werden ihre Grenzen erst von der*dem Archivierenden durch die oben genannten Einstellungen gesetzt, z.B. wenn aus rechtlichen Gründen und/oder aufgrund der Begrenztheit des Speicherplatzes externe Links nicht verfolgt werden. Jede Spiegelung einer Netzressource kann daher als neues Werk oder neue Version betrachtet werden. Häufig nähert man sich beim Webharvesting lediglich einer authentischen Kopie. Trotz aller Bemühungen stößt man an Grenzen, die den Quellen immanent sind: Webquellen werden aus dem Netz »ausgeschnitten« und damit zu etwas, was sie nicht sind: abgeschlossen. Eine hundertprozentige 1:1-Kopie ist bei Weblogs unmöglich und das Ziel der Archivierung liegt darin, das Original in einer Art abzubilden, die möglichst viele seiner wesentlichen Eigenschaften erhält und die für den Crawl verwendete Konfiguration objektiv nachvollziehbar macht.

37 o.V.: »Browsen über DLA-Dokumentarten«. In: *DLA Marbach. Literatur im Netz*. <http://literatur-im-netz.dla-marbach.de/menu.do?browser, o.> [zuletzt eingesehen am 2.5.2022].

Die Rechte-Einholung für die Archivierung von Weblogs gestaltet sich aufwändig, viele Blogger*innen schreiben unter Pseudonym. Es wird in der Regel nur die hauptverantwortliche Person angeschrieben. Eine vollständige Ermittlung und Kontaktierung aller Beteiligten (weitere Autor*innen, Fotograf*innen, Illustrator*innen, Verfasser*innen von Stylesheets und nicht zuletzt die Diskussionsteilnehmende) ist, auch aufgrund der Wahrung der Pseudonyme, nicht machbar, so dass hier urheberrechtliche Bedenken nicht vollständig ausgeräumt werden können. Die Herausforderungen, die sich bei der Archivierung von Weblogs stellen, wurden bereits an anderer Stelle ausführlich dargelegt.³⁸ Als »Schreibweisen der Gegenwart« sind literarische Blogs interessant und die Quellen im Archiv bieten für die Jahre 2008-2018 der Forschung eine gute Grundlage.³⁹

Die Archivierbarkeit Sozialer Medien – Ausblick

Das Deutsche Literaturarchiv sammelt Literatur, die im Internet entsteht, nicht vollständig, sondern dokumentiert Phänomene sowie neue »Schreibweisen der Gegenwart« mit Corpus-Bildung. Die Sammelkriterien werden der Öffentlichkeit transparent gemacht.⁴⁰ Als Quelleninstitut wird das Archiv gerade auch im Bereich der Sozialen Medien versuchen seinen Anspruch an das »Look and Feel« der Quellen im Archiv zu erhalten und diese in ihrer Umgebung zu dokumentieren. Der Ansatz wird durch eine Reihe von Studien zu neuen Literaturplattformen und der wieder neu geführten Diskussion zum Thema »Autorschaft in Sozialen Medien« unterstützt. Auf der Plattform »Netzliteraturwissenschaft« formuliert es ein Beitrag so:

Zudem ist es unabdingbar, die unterschiedlichen Webplattformen wie Facebook, Twitter oder Instagram, ihr spezifisches Funktionieren sowie ihre je-

38 Vgl. Schmidgall/Walter: »Literatur im Netz«.

39 Im Archiv sind drei der in einer aktuellen Publikation beschriebenen Blogs gesichert: »Auf der Borderline nachts um halb ein« von Joachim Lottmann, »Dschungel Anderswelt« von Alban Nikolai Herbst und »Aleatorik« von Claus Heck unter dem Pseudonym Aléa Torik. Vgl. Marcella Fassio: *Das literarische Weblog. Praktiken, Poetiken, Autorenschaften*. Bielefeld 2021.

40 o.V.: »Auswahlkriterien und Verfahren«. In: *DLA Marbach*. <https://www.dla-marbach.de/bibliothek/literatur-im-netz/auswahlkriterien-und-verfahren>, o.J. [zuletzt eingesehen am 2.5.2022].

weiligen literarischen Formate wie Facebook-Romane, Twitteratur oder Instapoesie möglichst genau und im Anschluss an literarische Gattungstraditionen zu analysieren. Aber auch neue Medienformate wie Litblogs, Fanfiction oder narrative und interaktive Computerspiele müssen in ihrer jeweiligen Form beschrieben werden.⁴¹

Ein 2016 bei der DFG eingereichtes Vorhaben zur »Archivierung von deutschsprachigen literarischen Tweets und Twitter-Literatur« konnte nicht weiterverfolgt werden, da die Entwicklung der im DLA eingesetzten Software SWBcontent aufgeben werden musste. Geplant war der Aufbau eines repräsentativen Corpus, die Auswahlliste enthielt 70 Einträge. Anfang 2021 hat Twitter sein komplettes Archiv für die Forschung geöffnet. Bisher konnten Wissenschaftler*innen über eine Programmierschnittstelle nur auf die Tweets der letzten sieben Tage zugreifen. Nun bietet die Plattform einen kostenfreien Zugang zum kompletten Archiv, auf alle Tweets seit Bestehen des Netzwerks im Jahr 2006.⁴² Ein Schritt, der das geplante DFG-Vorhaben obsolet machen würde? Nur bedingt, denn über die API können zwar alle Tweets ab 2006 abgezogen werden und der Text mit computergestützten Methoden analysiert werden, allerdings fehlt der Kontext der Plattform mit ihren sozialen Funktionen und die Eingrenzung auf literarische Werke.

Genau hier liegt die Aufgabe eines Literaturarchivs bei der Archivierung Sozialer Medien. Am Beispiel von Twitter verdeutlicht: die Einbettung des kompletten Accounts und des einzelnen Tweets in seiner Umgebung mit weiteren eigenen Tweets, eigenen und fremden Retweets, Followers und Likes machen das zu erhaltende und dauerhaft zu präsentierende Archivobjekt aus. Mit herkömmlichen Werkzeugen wie sie im Bereich Weblogs verwendet werden, ist die ursprüngliche Nutzersicht mit ihrem Geflecht an Bezügen, Verweisungen und Gestaltungselementen nicht zu archivieren. Eine von Rhizo-

41 o.V.: »Plattformen, Medienformate und Gattungen der Netzliteratur«. In: *Netzliteraturwissenschaften*. <https://netzliteraturwissenschaft.net/medienformate-und-gattung-en/>, o.J. [zuletzt eingesehen am 2.5.2022].

42 Vgl. die Meldung der AWW Webarchivierung: o.V.: »Twitter öffnet sein Tweet-Archiv für Forschungszwecke«. In: *AWW Webarchivierung*. <https://webarchivierung.awv-net.de/ueber-uns/aktuelles/twitter-oeffnet-sein-tweet-archiv-fuer-forschungszwecke.html>, 17.2.2021 sowie den Artikel des Mediennetzwerks The Verge: Nick Statt: »Twitter is opening up its full tweet archive to academic researchers for free«. In: *The Verge*. <https://www.theverge.com/2021/1/26/22250203/twitter-academic-research-public-tweet-archive-free-access>, 26.1.2021 [beide zuletzt eingesehen am 2.5.2022].

me⁴³ entwickelte Webrecorder-Technologie ermöglicht es Interaktivität (klicken, scrollen etc.) aufzuzeichnen und scheint für die Archivierung von Social Media geeignet zu sein. Das Tool erlaubt seinen Nutzer*innen dynamische Seiteninhalte durch Browsing zu sammeln, die Ergebnisse sofort zu überprüfen und das Material direkt herunterzuladen. Die ununterbrochene Speicherung der Serveranfragen und -antworten macht eine vollständige und authentische Archivierung der Nutzer*innensicht (Transactional Harvesting) möglich.

Die rechtliche Lage für eine Archivierung von Social Media Inhalten in Deutschland ist komplizierter als in den Vereinigten Staaten. Twitter beispielsweise überlässt den Autor*innen zwar jegliche Rechte an ihren Tweets und erlaubt ausdrücklich auch ein Crawling, doch ist der Umgang mit Retweets bisher noch nicht vollständig geklärt: Twitter-Nutzer*innen haben faktisch die Möglichkeit anonym zu bleiben, da im Gegensatz zu Facebook Twitter seine Nutzer*innen nicht zur Angabe von Klarnamen verpflichtet. Darüber hinaus belässt Twitter in seinen AGBs eindeutig die Urheberrechte bei den Nutzer*innen und erlaubt ausdrücklich die Archivierung der Inhalte.⁴⁴ Die Rechteklärung wird daher bei der Archivierung von Sozialen Medien – ob auf Twitter, Instagram, Facebook, Pinterest oder TikTok – ebenso wichtig und aufwändig sein wie bei der Archivierung von Weblogs.

Seit Ende 2019 entsteht unter Beteiligung des Deutschen Literaturarchivs Marbach im Rahmen des Projekts SDC4Lit ein verteiltes langzeitverfügbares Repositorium für Digitale Literatur und eine Forschungsplattform. In diesem Rahmen wird das Deutsche Literaturarchiv seine Sammeltätigkeit wieder aufnehmen und über eine Erweiterung des Sammelprofils auf neue, ›flüchtige‹ Schreibweisen der Gegenwart im Netz nachdenken können.

»Ein großes Archiv für deutschsprachige elektronische Literatur existiert bisher noch nicht.«⁴⁵ Auch wenn man Beat Suter 2021 noch nicht wiederprechen kann und große Zentralarchive sicher nicht mehr zeitgemäß sind,

43 Rhizome ist eine Non-Profit-Organisation, welche die Verwendung neuer Technologien für das Erzeugen und Erhalten digitaler Kunst fördert, siehe auch o.V.: »Conifer«.

44 Vgl. die allgemeinen Geschäftsbedingungen von Twitter, online verfügbar unter: o.V.: »Twitter Allgemeine Geschäftsbedingungen«. In: *Twitter*. <https://twitter.com/tos?lang=de>, 2021 [zuletzt eingesehen am 2.5.2022].

45 Beat Suter: »Digitales Schreiben und Archivierbarkeit von elektronischer Literatur«. In: *Siegener Periodicum zur Internationalen Empirischen Literaturwissenschaft (SPIEL)* 29/1-2 (2010) (= Sonderheft: Florian Hartling/Beat Suter (Hg.): *Archivierung von digitaler Literatur: Probleme – Tendenzen – Perspektiven*), S. 37-61, hier S. 46.

die Grundsteine für ein verteiltes, vernetztes Archiv sind gelegt⁴⁶ und das Interesse der Forschung wächst.⁴⁷ Aufgabe von Archiven wird es sein, die flüchtigen und von technischen Umgebungen abhängigen Primärquellen der netzbasierten Gegenwartsliteratur zu archivieren und den Zugang zu ihnen langfristig zu sichern.

46 Neben der Plattform »Literatur im Netz« des Deutschen Literaturarchivs gibt es eine Reihe von Initiativen, mit denen das Deutsche Literaturarchiv in engem Kontakt steht: Dilimag in Innsbruck in o.V.: »dilimag. Digitale Literaturmagazine«. In: *Innsbrucker Zeitungsarchiv*. <https://iza-server.uibk.ac.at/dilimag/home?conversationContext=1>, 2022 [zuletzt eingesehen am 2.5.2022], die Plattform der Autor*innen selbst in Johannes Auer u.a.: »netzliteratur.net_ Netzliteratur//Internetliteratur//Netzkunst«. In: *Netzliteratur.net*. <https://www.netzliteratur.net/>, 2021 [zuletzt eingesehen am 2.5.2022] sowie die drei Kollektionen der Electronic Literature Organization, in der sich einige wenige Netzliteratur-Beiträge aus Deutschland befinden in »Electronic Literature Collection«. In: *eliterature.com*. <https://collection.eliterature.org/>, o.J. [zuletzt eingesehen am 2.5.2022]. Das Archiv der deutschsprachigen elektronischen Literatur, ADEL, dokumentiert die Entwicklung der elektronischen Literatur in Deutschland, Österreich und der Schweiz seit den 1960er Jahren, dort ist geplant, nicht mehr lauffähige Werke durch Emulatoren nachbilden zu können. Vgl. o.V.: »Über ADEL«. In: *Archiv der deutschsprachigen elektronischen Literatur*. <https://adel.uni-siegen.de/node/5397>, o.J. [zuletzt eingesehen am 2.5.2022].

47 Beispielsweise angeregt durch die Veröffentlichungen des DFG-Projekts »Schreibweisen der Gegenwart« (2020-2022), vgl. o.V.: »DFG-Projekt ›Schreibweisen der Gegenwart««. In: *Universität Greifswald. Institut für Deutsche Philologie*. <https://germanistik.uni-greifswald.de/institut/arbeitsbereiche/neuere-deutsche-literatur/dfg-projekt-schreibweisen-der-gegenwart/>, 2020 und seit Mitte 2021 durch die Webplattform »#Netzliteraturwissenschaft« von Thomas Ernst: *Netzliteraturwissenschaft.net*. <https://netzliteraturwissenschaft.net>, 2021 [zuletzt eingesehen am 2.5.2022].

