

Aufsatz

Marco Braghieri, Tobias Blanke und Jonathan Gray

Eine Analyse von *Longform.org*

Wie journalistische Sammelportale als Orte von Datafizierung und Kuratierung wirken

Abstract: Welche Bedeutung haben Sammelportale für großformatige Inhalte im heutigen Journalismus? Dieser Artikel vertritt die These, dass sie als Studienobjekt der Journalismusforschung wichtiger werden.^[1] Am Beispiel von *Longform.org*, einem führenden Sammelportal für großformatige Inhalte im Netz, werden unter Zuhilfenahme digitaler Methoden empirische Analysen und theoretische Schlussfolgerungen präsentiert. Unserem Verständnis nach nutzt *Longform.org* die Datafizierung von Nachrichteninhalten, um den sogenannten Long Tail von archiviertem Material aufzuwerten. Mittels digital herausgegriffener Archivdaten nehmen wir eine eingehende Analyse von Longform-Portalen vor. *Longform.org* weist zwar ein gewisses Maß an kuratorischer Vielfalt auf, doch kommen ältere Nachrichtenmedien tendenziell häufiger vor. Einer der Hauptfaktoren, um auf *Longform.org* zu erscheinen, ist die Zugänglichkeit der digitalen Medienarchive. Unsere Analyse zeigt die relevante Rolle kleinerer, rein digitaler Publikationen, die eine einzigartige Mischung von Quellen bieten. Anhand einer Netzwerkanalyse herausgegriffener Markierungen untersuchen wir die Zusammensetzung der Themen, einschließlich der Bereiche Persönliches, Weltpolitik, Prominente, Technologie und Kultur. Die Daten und die kuratorischen Praktiken von Longform-Portalen können als aktuelle Form heutiger Informationsübermittlung verstanden werden, durch die frühere Sichtweisen leichter im Netz verfügbar, erfahrbar und programmierbar werden.

1 Eine ähnliche Version dieses Artikels ist als Buchkapitel erschienen in: Marco Braghieri (2021): *Yesterday's News. The future of long-form journalism and archives*. Oxford et al.: Peter Lang.

Einführung


Während beim heute üblichen Konsum von digitalen Nachrichtenmedien »Medieninhalte die Aufmerksamkeit für längstens eine Woche beschäftigen und dann vollkommen vergessen werden« (Fisher 2009: 59)², richtet sich ein nachhaltigeres Nutzungsinteresse im World Wide Web auf die digitalen Archive von journalistischen Medien (vgl. Elliot 2012). Solche Archive werden hier verstanden als Sammlungen »journalistischer Publikationen, Produktionen oder verwandter Inhalte [...], die in digitaler Form gespeichert und verfügbar gemacht werden« (Bødker 2018: 1114).

Großformatiger Journalismus kann auch als langsamer (vgl. Le Masurier 2015) oder literarischer Journalismus definiert werden. Seine »Berichte aus dem wahren Leben lassen sich wie ein Roman oder eine Novelle lesen« (Hartsock 2000: IX). In der heutigen digitalen Nachrichtenproduktion und -konsumtion setzt er sich vom beschleunigten Nachrichtenumschlag ab (vgl. Le Masurier 2016); legt Wert auf journalistische Darstellungsformen und Recherche (vgl. Le Masurier 2015); sucht nach Originalität (vgl. Belt/South 2016); und stellt vielseitige Produktionstechniken des digitalen Storytelling »ins Herz der erzählerischen Struktur« (Hiippala 2017: 421).

Großformatiger oder erklärender Journalismus ist als Teil des digitalen Journalismus zu betrachten, »der sich wandelnden sozialen Praxis des Auswählens, Interpretierens, Herausgebens und Verbreitens faktischer Informationen von angenommenem öffentlichen Interesse für verschiedene Publika in besonderen, aber sich verändernden Genres und Formaten. Digitaler Journalismus prägt neue Technologien und Kommunikationsprogramme und wird gleichzeitig durch sie geprägt. Er befindet sich zunehmend in einem symbiotischen Verhältnis zu den diversen Zielgruppen« (Steensen et al. 2019: 338). Wichtig ist deshalb, wie Seaton (2016) betont, dass sich in der gegenwärtigen Kommunikationsarchitektur zwei Pole gegenüberstehen: »Es gibt eine überwältigende Fülle an Informationen und Kommunikationsformen, die viele Facetten haben und sowohl multilateral als auch multinational genutzt werden«, und dennoch können »enge ‚Silos‘ an Information und Meinung entstehen« (Seaton 2016: 808). Das von Seaton (2016) beschriebene Szenario ist außerdem durch die ökonomischen Schwierigkeiten der herkömmlichen Medien geprägt, die sich, wie Franklin betont, »abnehmenden Zuhörer-, Zuschauer- und Leserschaften sowie Werbeeinnahmen« (Franklin 2014: 482) gegenübersehen. Ein Teil der Medienindustrie hat darauf »mit neuen Kommunikationstechnologien und sorgfältiger Diversifikation ihres Produktangebots« reagiert, allerdings mit ungewissem Erfolg.

2 Dieses und die nachfolgenden Zitate wurden aus dem Englischen übersetzt.

Screenshot Longform.org (18. Juni 2021)



Home

- ★ Best Articles
- ▶ Podcast
- 📁 Sections
- 📁 Collections
- 📄 Reprints

✂ Random Article

📋 Lists

- Best of 2020
- Best of 2019
- Best of 2018
- Best of 2017
- Best of 2016
- Best of 2015
- Best of 2014
- Best of 2013
- Best of 2012

📁 Archive

- Publications
- Writers
- Tags

⋮ More

- Newsletter
- Suggest Article
- About
- Contact
- Advertise

📧 📧 📧 📧

Get our Newsletter
Great articles, every Saturday.

Subscribe

Read Later
Choose service...

Today, June 18



SCIENCE HEALTH

Is There Something Wrong With the Air in South Portland, Maine?

Residents have lived near more than 100 massive petroleum storage tanks for decades, never really knowing if they're breathing in dangerous chemicals. Now they're fighting to find out.

KATHRYN MILES BOSTON GLOBE MAGAZINE JUN 2021 15min 00

Yesterday, June 17

My Father Vanished When I Was 7. The Mystery Made Me Who I Am.

My dad was a riddle to me, even more so after he disappeared. For a long time, who he was—and by extension who I was—seemed to be a puzzle I would never solve.

NICHOLAS CASEY NEW YORK TIMES MAGAZINE JUN 2021 35min 00



SCIENCE

The Lab Leak Theory Doesn't Hold Up

The rush to find a conspiracy around the COVID-19 pandemic's origins is driven by narrative, not evidence.

JUSTIN LING FOREIGN POLICY JUN 2021 20min 00



HISTORY

If We Can Soar: What Birmingham Roller Pigeons Offer the Men of South Central

In 1970 South Central, pigeon fancying was serious business. But there's a deeper story behind why these Black Angelenos are entering their fifth and sixth decade raising Birmingham Roller pigeons.

SHANNA B. TIAYON PIPE WRENCH JUN 2021 30min 00

»Es ist noch nicht klar, ob die Medienkonzerne wieder die Art von Gewinnen erzielen werden, die ihnen vorschwebt, oder ob die Konsumenten neuartige Medienprodukte so eifrig annehmen, wie es von ihnen erwartet und vorhergesagt wird« (Chyi/Chadha 2012: 432). Während mobile Nutzung attraktiver wurde, wie Nel und Westlund unterstrichen haben (2012), sah sich die Medienindustrie wieder einmal vor einem Scheideweg, nämlich der Wahl zwischen einer grundlegenden Neuausrichtung ihres Produktkonzepts oder einer eher passiven Haltung. Letztere könnte dazu führen, dass »unabhängige Softwareentwickler [...] rasch die Gelegenheit ergreifen, Apps zu kreieren, welche die reichhaltigen Inhalte einsammeln, die Pressemedien gratis im Web zur Verfügung stellen« (Nel/Westlund 2012: 751).

Diese Vorhersage findet sich im Longtail-Modell der Web-Ökonomie von Anderson (2009) wieder, denn Plattformen und Portale sind in der heutigen digitalen Medienwelt für die Verbreitung von Inhalten effektiver als Produzenten. Anderson sieht das Longtail-Modell charakterisiert durch eine »unbegrenzte Auswahl. Reichliches und billiges Angebot bedeutet reichliche, billige und unlimitierte Vielfalt« (Anderson 2009: 180). Die Verkaufskosten müssten so niedrig wie möglich gehalten werden, da sonst die Teilnehmenden am Markt zu »Gefangenen« würden (Anderson 2009: 185). Anderson nennt die Informationsmedien als Beispiel so einer »gefangenen Branche« und betont, welches wertschöpfende Potential vermittelnde Plattformen in diesem Bereich haben (Anderson 2009: 88). Unser Beitrag baut auf dieser Idee von Anderson auf und untersucht, wie großformatiger Journalismus durch Sammelportale verbreitet wird, zumal durch darauf ausgerichtete Gründungen wie *Longform.org*.

Longform.org war schon bisher Gegenstand intensiver Aufmerksamkeit, sei es als Lieferant von Nutzerdaten (vgl. Boynton 2013), durch Beschreibung seiner Gründung und Struktur (vgl. Albalad Aiguabella 2015) oder die Analyse des digitalen Marktes für großformatigen Journalismus (vgl. Dowling/Vogan 2015; Longhi/Winques 2015) und seiner heutigen Leserschaft (vgl. Jacobson et al. 2018). Unser Artikel unternimmt den Versuch, die Aktivitäten von *Longform.org* im Rahmen von Andersons Longtail-Theorie (vgl. Anderson 2009) in Bezug auf digitale Informationsanbieter und einen ihrer wichtigsten Typen von Inhalt zu analysieren, nämlich großformatigen Journalismus. Darüber hinaus bietet er durch digitale Methoden (vgl. Rogers 2013), besonders die innovative Anwendung »medienspezifischer Verfahren« wie Web-Scraping und Hyperlink-Analysen, einen frischen Blick auf Funktionsweise und Bedeutung von Informationsportalen. Wir arbeiten mit einem Datensatz von über 1000 Einträgen auf *Longform.org* aus dem Jahr 2016.

Hintergrund: Archive von Informationsmedien und Sammelportale als Verbreiter von großformatigem Journalismus

Dieser Abschnitt soll zeigen, wie aus Medienarchiven im Zusammenspiel mit digitalen Plattformen eine Netzwerkpraxis für die Verbreitung von Informationen werden kann. Es ist wichtig darauf hinzuweisen, dass digitale Archive von Informationsmedien starkes Interesse im World Wide Web finden können. Laut Chris Elliott, dem früheren Ombudsmann des *Guardian*, »umfasst das digitale Archiv des *Guardian* mehr als eine Million Artikel und ist sehr populär. Fast 40 Prozent der Inhalte, die auf der Website angeklickt werden, sind älter als 48 Stunden« (Elliott 2012). Zudem hat eine Studie von Smith, Connor und Stanton (2015) zum Thema Longform-Journalismus gezeigt, dass innerhalb ihres Korpus von 5,2 Millionen Beiträgen für die dauerhafte Relevanz ausschlaggebend war, dass »großformatige Artikel externe Links als Faktor des Lese-Interesses länger verfügbar halten als gewöhnliche Nachrichten« (Smith et al. 2015: 2115).

Der Abrufprozess stellt allerdings besondere Ansprüche an Content-Management-Systeme (CMS), die in Redaktionen für eine Reihe von Aufgaben verwendet werden, die vom Erstellen von Inhalten bis zu ihrem Redigieren, Herausgeben und Verbreiten reicht (vgl. Barker 2016). Deshalb stellt das Abrufen von Inhalten generell und besonders von großformatigen Beiträgen eine schwierige Herausforderung beim Verwenden von CMS dar. Longform-Journalismus ist ein Inhalt, bei dem die Fähigkeit, Lesende zu binden, über längere Zeit erhalten bleibt und deshalb von Informationsmedien mit höherer Wahrscheinlichkeit verfügbar gehalten wird. Wir können deshalb erklärenden Journalismus als einen der Faktoren betrachten, die das Interesse an den Archiven von Informationsmedien fördern. Allgemeiner gesagt: In der digitalen Gegenwart werden einzelne Beiträge in digitalen Archiven zusammengefasst, die als Vertriebsinstrumente fungieren. Daher werden wir uns nun darauf konzentrieren, wie man sich Archive von Informationsmedien als Verbreiter von Inhalten vorzustellen hat.

Anderson beschreibt den »Longtail« als Modell, das auf »unendlichen Wahlmöglichkeiten« basiert (Anderson 2009: 180). Es hat sich als eines der führenden Produktions- und Verteilungsmodelle in der digitalen Wirtschaft und Gesellschaft etabliert. Ihm gemäß sinken die Distributionskosten, während gleichzeitig die Vielfalt steigt, wobei das Publikum sich tendenziell so weit vergrößert, »wie es die Wahlmöglichkeiten zulassen« (Anderson 2009: 180). Anderson (2009) zufolge wurde dieses neue Modell bereitwilliger von intermediären Akteuren als von traditionellen Produzenten angenommen. Als eines der Hauptbeispiele für diese Dynamik nennt Anderson die Medienbranche: Digitale Plattformen funktionieren für die Verbreitung von Inhalten effektiver als die traditionellen Informationsmedien. Außerdem gibt es im Longtail-Modell andere wichtige Faktoren wie die Demokratisierung der Produktions- und Vertriebsinstrumente

sowie die abgleichende Verknüpfung von Angebot und Nachfrage durch Filter (vgl. Huang/Wang 2014).

In der digitalen Gegenwart sind großformatiger, erklärender Journalismus und digitale Medienarchive noch nicht vollständig datafiziert (vgl. Mayer-Schönberger/Cukier 2013). Mayer-Schönberger und Cukier definieren »Datafizierung« als Prozess der Organisation und Darstellung eines Phänomens »in einem quantifizierten Format, so dass es tabelliert und analysiert werden kann« (Mayer-Schönberger/Cukier 2013: 78). Während sich digitale Medienarchive einer wachsenden Präsenz im Netz erfreuen, was auf die Digitalisierung physischer Archive, die Abstimmung der Archive aufeinander und die Praxis der digitalen Archivierung zurückzuführen ist, lässt die Datafizierung ihrer Inhalte noch auf sich warten. Wie Blanke und Prescott hervorheben, ist der Prozess der Datafizierung »zu unterscheiden vom Prozess des Herstellens digitaler Surrogate, der auf der Digitalisierung ursprünglich analoger Inhalte basiert, [zum Beispiel] dem Überführen eines Mikrofilms von einem Buch in eine digitale Form oder dem Herstellen einer MP3-Version von einem aufgenommenen Interview« (Blanke/Prescott 2016: 192). Datafizierung bedeutet demgegenüber, dass aus dem Transformationsprozess ein quantitatives Format resultiert, das auf unterschiedliche Weise genutzt werden kann.

Datafizierung ist allerdings kein neutraler Vorgang, da einige Aspekte an Probleme erinnern, die durch das aufgeworfen werden, was Derrida als »Entpapierungs«-Prozess (vgl. Derrida 2005) bezeichnet hat. Derrida hat auf dessen potentiellen Nutzen hingewiesen, gleichzeitig aber auch Themen wie »unsichtbare Vermachtungen und Aneignungen« (Derrida 2005: 55ff.) betont. Ein Beispiel für die Nicht-Neutralität der Datafizierung liefern in diesem Zusammenhang Mayer-Schönberger und Cukier, die das Google-Books-Projekt als erfolgreiche Datafizierung loben (vgl. Mayer-Schönberger/Cukier 2013). 2004 begann Google mit dem Scannen von Büchern und baute so allmählich eine digitale Bibliothek auf. 2015 waren in diesem Projekt »mehr als 25 Millionen Bände von Texten in 400 Sprachen aus mehr als 100 Ländern gescannt worden« (Heymann 2015). Aus diesem Prozess ging aber nicht nur eine immense digitale Bibliothek hervor, sondern er hatte daneben auch ökonomische Vorteile für Google (vgl. Pybus et al. 2015). Während der Datafizierungsprozess aus der Sicht von Google erfolgreich war, weil er hohe Vermögenswerte und Raum für zahlreiche weitere Anwendungen hervorgebracht, führte er gleichzeitig zu den »unsichtbaren Vermachtungen und Aneignungen« (Derrida 2005: 55ff.), vor denen Derrida als Risiken der »Entpapierung« gewarnt hat.

Bevor wir unsere Analyse eines einzelnen Sammelportals für großformatigen Journalismus über ein Jahr hinweg beginnen, ist es nützlich, sich an die Ergebnisse der Forschungen von Smith, Connor und Stanton (2015) zum Longform-Journalismus zu erinnern. Die Autoren weisen darauf hin, dass die Produktion von großformatigem Journalismus in der digitalen Gegenwart zwar zunimmt,

dies jedoch in einem besonderen Verlauf vor sich geht. Ihre Studie zeigt, dass die Zahl der Medien wächst, die Longform-Journalismus produzieren, und dass es unter ihnen eine starke Präsenz solcher Angebote gibt, die von Anfang an nur digital verfügbar sind (vgl. Smith et al. 2015).

Die Ergebnisse dieser Untersuchungen (vgl. Smith et al. 2015) charakterisieren die digitale Gegenwart also als ein Milieu, in dem großformatiger Journalismus in diversen Medien zu finden ist, von denen jedoch nur eine relativ kleine Zahl die ökonomische und organisatorische Stärke hat, die Produktion von Longform-Journalismus nachhaltig zu garantieren, worauf Bruns, Highfield und Lind (2012) hinweisen. »Nur wenige journalistische Organisationen können es sich leisten, viel in großformatigen, kostenintensiven investigativen Journalismus zu investieren« (Bruns et al. 2012: 2). Der Anstieg der journalistischen Longform-Produktion in der digitalen Gegenwart ist das Ergebnis einer großen Zahl von Medien, die jeweils nur eine kleine Zahl von großformatigen Stücken produzieren. Während er im Vergleich mit der standardisierten Nachrichtenproduktion in langsamerem Tempo wächst, hat dieser besondere Typ von Journalismus das Potential, für längere Zeit relevant zu bleiben (vgl. Smith et al. 2015). In dem von Smith, Connor und Stanton beschriebenen Szenario ist deshalb die Konzentration auf großformatigen Journalismus durch die Analyse eines Sammelportals wie *Longform.org* ein wirksamer Weg, um die Produktion und die Verbreitung dieser spezifischen Art des Journalismus zu untersuchen.

Longform.org: ein Beispiel für das digitale Sammeln und Aufbereiten von großformatigem Journalismus

Wie wir im vorigen Abschnitt gesehen haben, erreichen in Übereinstimmung mit dem Longtail-Modell Plattformen mit Inhalten Dritter Vertriebsziele effektiver als traditionelle Einheiten. Ein Beispiel für dieses Phänomen ist die Informationsbranche (vgl. Anderson 2009). Im folgenden Abschnitt beschränken wir uns auf die datenbasierte Analyse eines einzelnen digitalen Sammelportals für großformatigen Journalismus, nämlich *Longform.org*.

Longform.org (vgl. *Longform.org* 2010) wurde 2010 gegründet und startete mit aktuell publizierten, digital archivierten Stücken von über 2000 Wörtern Länge, die im World Wide Web verfügbar waren. Der Unterschied zwischen *Longform.org* und ähnlichen Plattformen liegt nach Auffassung von Shapiro, Hiatt und Hoyt (2015) im Ziel der Aktivitäten. *Longform.org* »will nicht den Traffic an sich reißen, sondern schiebt die Lesenden einfach auf die Seiten der Anbieter weiter« (Shapiro et al. 2015: 175).

Die erste Online-Version von *Longform.org* geht auf das Jahr 2010 zurück. Sie wurde parallel zu Apples iPad von zwei Einzelpersonen auf den Markt gebracht:

Max Linsky und Aaron Lammer. Seitdem hat das Portal sein Angebot erweitert. 2012 fügte es eine Belletristik-Rubrik hinzu, begann mit einer intensiven Podcast-Produktion und entwickelte seine erste iPad-App zum Preis von fünf US-Dollar, die 2014 dann 60.000-mal verkauft wurde (vgl. Bercovici 2014). Der Podcast erreichte laut einem 2014 im Magazin New York erschienenen Artikel 50.000 Hörer (vgl. Kachka 2014). Im September des gleichen Jahres brachte *Longform.org* seine erste iPhone-App heraus, bei der Lesende die Möglichkeit erhielten, Listen mit besonderen Autorinnen oder Autoren aufzustellen, denen bei verschiedenen Informationsmedien gefolgt werden konnte. Zudem wurden Algorithmen entwickelt, die regulieren, welche Artikel von der App angeboten werden sollen (vgl. Mullin 2014).

Seit April 2017 ist die App von *Longform.org* weder auf iOS- noch auf Android-Basis mehr erhältlich (vgl. Longform.org 2017b). Die Gründer Max Linsky und Aaron Lammer erklärten dazu, dass die von ihnen entwickelte neue Version der App von Apple abgelehnt worden sei (vgl. Longform.org 2017b).

Linsky und Lammer heben hervor, dass die Instrumente, die es Lesenden erlauben, auf großformatige Inhalte zuzugreifen, im Laufe der Zeit erheblich weiterentwickelt wurden, und betonen außerdem, dass sie – über *Longform.org* – »seit 2012 mehr als 100 Millionen ausgehende Links an Verlage geschickt haben« (Longform.org 2017b). Im Übrigen ist zu beachten, dass *Longform.org* sich »mehr als Technologie-Unternehmen denn als Kommunikationsmedium versteht« (Albalad Aiguabella 2015: 18).

Weitere Einsichten in die Praxis von *Longform.org* liefert Robert S. Boynton. Boyntons Daten zufolge müsste die demografische Zusammensetzung des Publikums von *longform.org* das Herz eines jeden Werbetreibenden höher schlagen lassen: »Jung (50% der Leserinnen und Leser sind unter 34), mobil (30% lesen hauptsächlich auf Smartphones und Tablets) und gebildet (42% haben einen Hochschulabschluss)« (Boynton 2013: 130). Außerdem weist Boynton darauf hin, dass die »beste erzählende Sachliteratur – im Gegensatz zu im Grunde allen anderen Web-Inhalten – ihre Attraktivität nicht verliert, wenn sie altert. 10 Prozent des Longform-Publikums lesen lieber eine ältere Geschichte als eine neue. Der Publikationszeitpunkt hat kaum Gewicht. Für die Lesenden ist das Thema eines Artikels wichtiger als die Frage, wie neu er ist« (Boynton 2013: 130off.). Boyntons Daten zufolge kommen bei großformatigem Journalismus und Medienarchiven in der digitalen Gegenwart also nicht nur die nachhaltige Relevanz zusammen, sondern auch mögliche Sammel- und Aufbereitungs-Praktiken, die, wenn wie bei *Longform.org* lesefreundlich angewendet, einen ständigen Publikumszustrom mit Material aus Archiven von Informationsmedien versorgen können, das sonst ungenutzt bliebe.

Die Aktivitäten von Sammelportalen wie *Longform.org* zeigen, dass es in den digitalen Archiven der Informationsmedien Ressourcen gibt, besonders großformatige Inhalte, die weniger genutzt werden als es möglich wäre, weil sie nicht gebündelt und aufbereitet werden, um die Leserschaft zu verstärken. Das gilt

nicht nur für Traditionsmedien, die über große Mengen archivierter Inhalte verfügen, Sammelportale sind auch wichtig für die Absatzchancen von Medien mit kürzerer Lebensdauer. Das wurde von Boynton unter Bezug auf eine von *Longform.org* initiierte Leserschaftsumfrage so beschrieben: »Ein bekannter Verlagsname bewegt den Zeiger fast nicht [...], unbekannte Verlage haben oft mehr Erfolg als bekannte Markennamen, denn Lesende sind begierig, etwas Anderes, für sie Neues zu erfahren« (Boynton 2013: 131).

Nachdem wir einen Blick auf die Entwicklung von *Longform.org* und auf damit zusammenhängende allgemeine Daten geworfen haben, stellen wir dem nun einen spezifischen Datensatz gegenüber, den wir der Tätigkeit von *Longform.org* im Jahr 2016 entnehmen. Wir wollen prüfen, ob sich die identifizierten wesentlichen Faktoren, wie die nachhaltige Relevanz von großformatigen Stücken und die Bedeutung von Archiven als Ressourcenpools, auch bei der Datenanalyse im folgenden Abschnitt zeigen.

Die Tätigkeit von *Longform.org* 2016 – eine quantitative Analyse

Wir werden nun eine quantitative Analyse aller Einträge vornehmen, die *Longform.org* 2016 produziert hat. Das Herstellen unseres Datensatzes begann mit dem Sammeln aller *Longform.org*-Einträge aus der Zeit zwischen dem 1. Januar und dem 31. Dezember 2016. Die Website von *Longform.org* ermöglicht ein lückenloses Navigieren bis zurück zum 1. April 2010. Während das Design der Website seit ihrer Gründung erneuert wurde, blieb die Anordnung um eine zentrale Spalte erhalten, die eine Liste von Artikeln präsentiert.

Wir haben die erforderlichen Daten für alle *Longform.org*-Einträge von 2016 aus dem World Wide Web mit Hilfe eines »web scraping« genannten Verfahrens gewonnen, das definiert wird als »Praxis des Datensammelns, wobei jedes andere Mittel benutzt werden kann außer einem Programm, das mit einer API verbunden ist (oder natürlich durch einen Menschen, der einen Webbrowser nutzt)« (Mitchell 2015: VIII). Darüber hinaus ist »Scraping nicht nur eine Verfahrenstechnik, sondern auch eine spezielle Art und Weise, mit Informationen und Kenntnissen umzugehen: Es ist auch eine analytische Praxis« (Marres/Weltevrede 2013: 317). Als eine solche wird Web-scraping als eine der digitalen Methoden angewendet, die als »Techniken für die voranschreitende Erforschung der Inhalte von Online-Medien« (Venturini/Bounegru et al. 2018: 4) geeignet sind und benutzt werden, um »an Informationen zu gelangen, die auf Internet-Plattformen zugänglich sind« (Venturini/Bounegru et al. 2018: 2). Um unser Web-Scraping-Vorgehen zu realisieren, haben wir eine Browser-Erweiterung von Google Chrome namens Data Miner verwendet, »die Ihnen hilft, Daten, die Sie in Ihrem Browser sehen, zu extrahieren und in einer Excel-Tabellen-Datei abzuspeichern« (Data Miner 2016).

Um die von uns benötigten Daten von *Longform.org* zu erhalten, mussten wir eine Reihe von »Extraktionsanweisungen« entwickeln, »die Data Miner zum Extrahieren von Daten aus Websites« (Data Miner 2016) vorschreibt und die »Rezepte« genannt werden. Wir konnten so 50 Seiten aus der Website von *Longform.org* mit Hilfe von Web-scraping bearbeiten. Die Daten wurden uns in einer CSV-Datei vorgelegt, die wir dann in OpenRefine eingegeben haben, ein interaktives Datentransformations-Tool (vgl. Verborgh/De Wilde 2013). Mit OpenRefine nahmen wir eine Qualitätsprüfung [»data profiling«] und Bereinigung [»data cleaning«] der Daten vor. Das Data-profiling diente dazu, die »tatsächliche Struktur, Inhalt und Qualität« der Daten herauszufinden; das Data-cleaning, um kleine mögliche Fehler in »halbautomatischer Weise« (Verborgh/De Wilde 2013: 6) zu korrigieren. Wir analysieren nun den Datensatz, der das Ergebnis des Web-scraping-, des Data-profiling- und des Data-cleaning-Prozesses ist.

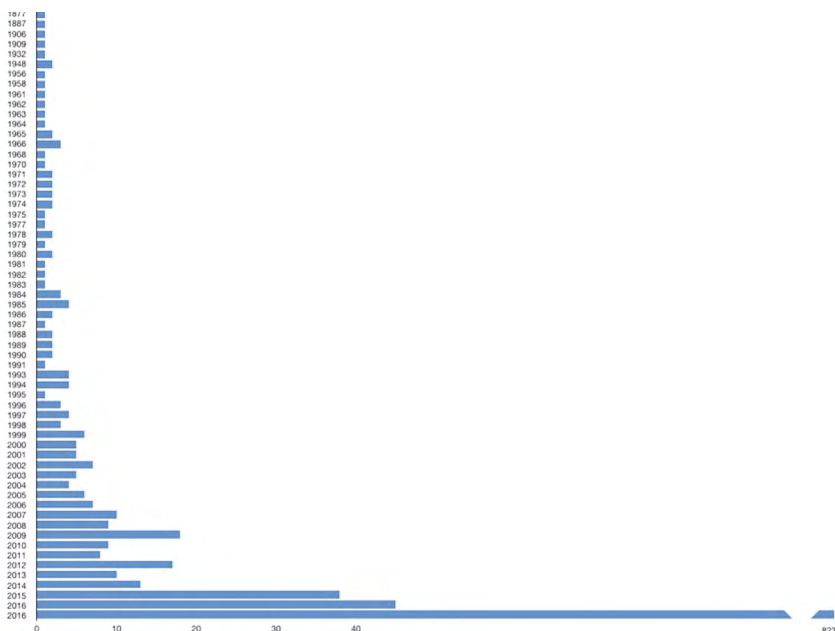
Insgesamt haben wir 1225 Beiträge (»posts«) in der Zeit von 1. Januar 2016 bis zum 31. Dezember 2016 untersucht. Typischerweise erstellt *Longform.org* Einträge, die einen Link zu einem großformatigen journalistischen Stück, dessen Zusammenfassung, die Biographie und weitere Informationen zur Person der Autorin oder des Autors, die Angabe des betreffenden Informationsmediums und das Datum enthalten, an dem das Stück zuerst publiziert wurde. 2016 veröffentlichte *Longform.org* 1074 Einträge zu einzelnen Stücken. Neben diesem Haupttyp hat *Longform.org* im Laufe der Jahre noch weitere Arten von Einträgen entwickelt. Die erste ist der »longform guide entry«, der in aller Regel großformatige Stücke aus verschiedenen Medien zusammenstellt, die sich auf das gleiche Thema konzentrieren (30 im Jahr 2016). *Longform.org* veröffentlicht auch Einträge, die einer Person gewidmet sind, die schon mehrmals als Autor(in) in den Einträgen der Website zu einzelnen Stücken zu finden war (15 im Jahr 2016). Neben den thematisch orientierten und personenbezogenen gibt es noch wöchentliche Einträge, die sich dem belletristischen Schreiben widmen (51 im Jahr 2016). Außer den sammelnden und kuratierenden Aktivitäten hat *Longform.org* eine bedeutende originale Multimedia-Produktion in Form einer Podcast-Serie entwickelt (55 Stücke im Jahr 2016).

Wie sich zeigt, besteht die überwiegende Mehrheit aus Einträgen zu einzelnen Stücken, die wir nun tiefergehend analysieren wollen. Wir versuchen die von *Longform.org* getroffene Auswahl von Informationsmedien zu beurteilen und richten unser Augenmerk auf die Erscheinungsdaten, an denen die jeweiligen Longform-Stücke zuerst veröffentlicht wurden. Da es sich bei *Longform.org* um einen kuratierenden Dienst handelt, den wir als Sammelportal im Sinne des Longtail-Modells verstehen (vgl. Anderson 2009), ist es wichtig, die Selektionskriterien im Detail zu untersuchen, zumal sich die Aktivität speziell um großformatigen Erklärjournalismus und die digitalen Archive von Informationsmedien dreht.

Wir konzentrieren unsere Analyse also auf die 1074 Einträge zu Einzelstücken, die *Longform.org* 2016 publiziert hat. Da das erste Erscheinungsdatum jeder Geschichte aus dem jeweiligen Eintrag ersichtlich ist, lässt sich feststellen, wie *Longform.org* seine Auswahl auf die verschiedenen Informationsmedien und Zeitphasen verteilt hat. Die 1074 Einträge von 2016 sind einem bemerkenswert langen Zeitraum entnommen, die älteste großformatige Geschichte, zu der ein Eintrag gefunden wurde, ist zuerst im Jahr 1877 erschienen. Über die Jahre betrachtet ist die Gruppe der aufgenommenen Einzelbeiträge am wichtigsten, die zuerst 2016 publiziert wurden, dem Jahr also, auf das wir unsere Analyse hauptsächlich richten. Einträge, die auf großformatige Stücke hinweisen, welche zuerst im Jahr 2016 veröffentlicht wurden, machen 77 Prozent aller Einträge aus.

Schaubild 1

Zahl der Einzeleinträge auf *Longform.org* im Jahr 2016 pro Ersterscheinungsjahr des Bezugsbeitrags



Wie Schaubild 1 zeigt, ist der Zeitraum, aus dem *Longform.org*-Geschichten ausgewählt wurden, um in Einzeleinträgen berücksichtigt zu werden, wie erwähnt sehr weit. Ungefähr ab 2009 aufwärts wird die Auswahl umfangreicher. *Longform.org* sammelt und kuratiert ausschließlich Beiträge aus Archiven von

Informationsmedien, die keine hohe Bezahlschranke errichten (Brock 2013: 155). Diese Vorbedingung wirkt sich offensichtlich auf den gesamten Pool von Informationsmedien aus, auf die dieses spezialisierte Portal zurückgreift.

Nachdem wir Einblick in die Ersterscheinungsdaten von Geschichten genommen haben, auf die sich die Gesamtproduktion von *Longform.org*-Einträgen zu Einzelbeiträgen bezieht, blicken wir jetzt auf die Einträge, deren Bezugsstücke vor 2016, dem Jahr unseres Untersuchungsmaterials, zuerst erschienen sind. Wie aus Schaubild 1 hervorgeht, betrifft diese Analyse 23 Prozent aller Einzeleinträge. Um einen prägnanteren Eindruck von ihrer Verteilung zu erhalten, haben wir alle vor 2016 erschienenen Stücke, auf die sich die Einträge beziehen, in Dekaden aufgeteilt und die Einträge zu von *Longform.org* ausgewählten Geschichten, die vor 1960 publiziert wurden, zu einer Gruppe zusammengefasst.

Schaubild 2

Zahl der Einzeleinträge auf *Longform.org* von 2016 mit Ersterscheinungsdaten der Bezugsbeiträge vor 2016

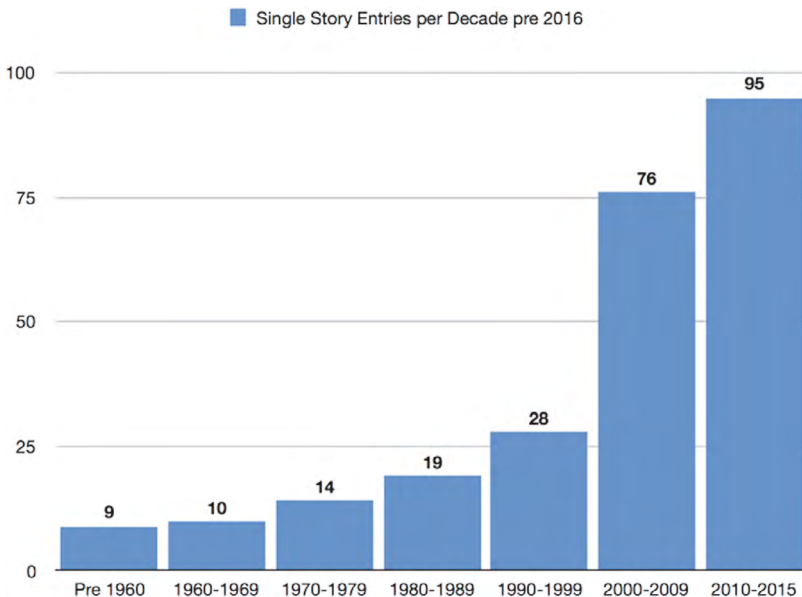
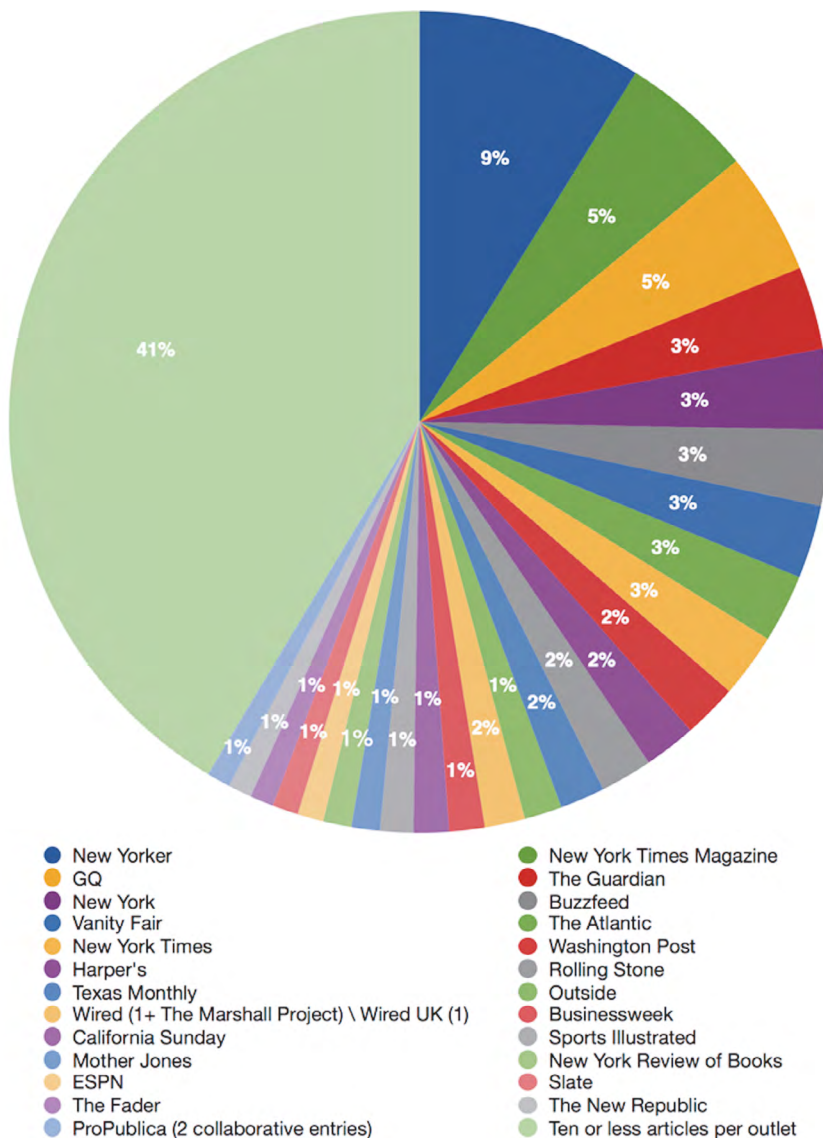


Schaubild 2 zeigt, wie die Zahl der Einzeleinträge, die sich um großformatige Stücke drehen und die vor 2016 publiziert wurden, von Dekade zu Dekade gewachsen ist. Die Zunahme ist allerdings in der Phase von 2000 bis 2009 mit 171 Prozent (und in der folgenden Dekade ab 2010) besonders stark.

Schaubild 3

Journalistische Medien, die in *Longform.org*-Einzeleinträgen im Jahre 2016 berücksichtigt wurden



Der Digitalisierungsprozess in den Redaktionen hatte einen starken Einfluss auf die digitalen Medienarchive, deren Menge und Bestände erweitert wurden, was sich in der großen Zahl an Stücken spiegelt, die in der Dekade zwischen 2000 und 2009 entnommen wurden. Während die digitalen Archive von Informationsmedien, die auf physische Bestände zurückgehen, mit intensivem Arbeitsaufwand erstellt werden mussten, hat die Digitalisierung in den Redaktionen selbst einen breiteren Zugang zu den journalistischen Archiven ermöglicht.

Wie sich aus Schaubild 3 ersehen lässt, sind Informationsmedien, die ihre digitalen Archive erheblich aufgestockt haben, tendenziell in höherem Maße repräsentiert. Schaubild 3 zeigt, dass etwas mehr als vier von zehn großformatigen Stücken, die von Longform.org für Einzeleinträge ausgewählt wurden, zuerst von Medien publiziert wurden, die insgesamt zehn Mal oder seltener in den 1074 Einträgen vertreten sind. Die Gesamtzahl an Informationsmedien, die von Longform.org 2016 bei Einzeleinträgen berücksichtigt wurden, beträgt 221, und von diesen journalistischen Medien kamen nur 25 bei zehn oder mehr Einträgen vor. Unter den 196 Medien, die weniger als zehn Mal berücksichtigt wurden, bilden diejenigen die größte Gruppe, die nur einmal berücksichtigt wurden. Sie umfasst 108 Informationsmedien.

Während sich der Großteil der von Longform.org behandelten Artikel auf eine bestimmte Auswahl von Informationsmedien konzentriert, ist Vielfalt der Informationsmedien insgesamt, aus denen die Plattform ihre Geschichten bezieht, ein wichtiger Faktor für die von Longform.org getroffenen Auswahlentscheidungen. Diese Vielfalt wird nicht nur dadurch erreicht, dass die Plattform die Zahl der Informationsmedien, aus denen sie ihre Geschichten bezieht, generell erweitert, sondern auch dadurch, dass sie gezielt Informationsmedien adressiert, die nur in geringem Maße berücksichtigt wurden. Diese Art der Auswahl deutet darauf hin, dass eins der wichtigsten Auswahlkriterien der Plattform die Vielfalt ihrer Angebote ist. Allerdings wurde 2016 bei den journalistischen Medien, die durch Einzeleinträge vorgestellt wurden, ein besonderes Augenmerk auf Artikel der Zeitschrift *The New Yorker* gelegt, die in 95 Einträgen berücksichtigt wurden.

Um die Konsistenz der Zahl der Top-Informationsmedien, die 2016 in Einzeleinträgen berücksichtigt wurden, prüfen zu können, haben wir unseren Datensatz mit den gesamten Häufigkeiten verglichen, mit denen einzelne Medien von Longform.org berücksichtigt wurden. Während diese Daten, die sich auf den 22. Mai 2017 beziehen, bei Longform.org heruntergeladen werden können (vgl. Longform.org 2017a), wird die Methode, mit der Longform.org sie gewonnen hat, nicht spezifiziert. Die entsprechende Seite zeigt lediglich an, aus wie vielen Medien eine festgelegte Mindestanzahl von Beiträgen veröffentlicht wurden.

Schaubild 4

Journalistische Medien, die in *Longform.org*-Einzeleinträgen im Jahre 2016 berücksichtigt wurden

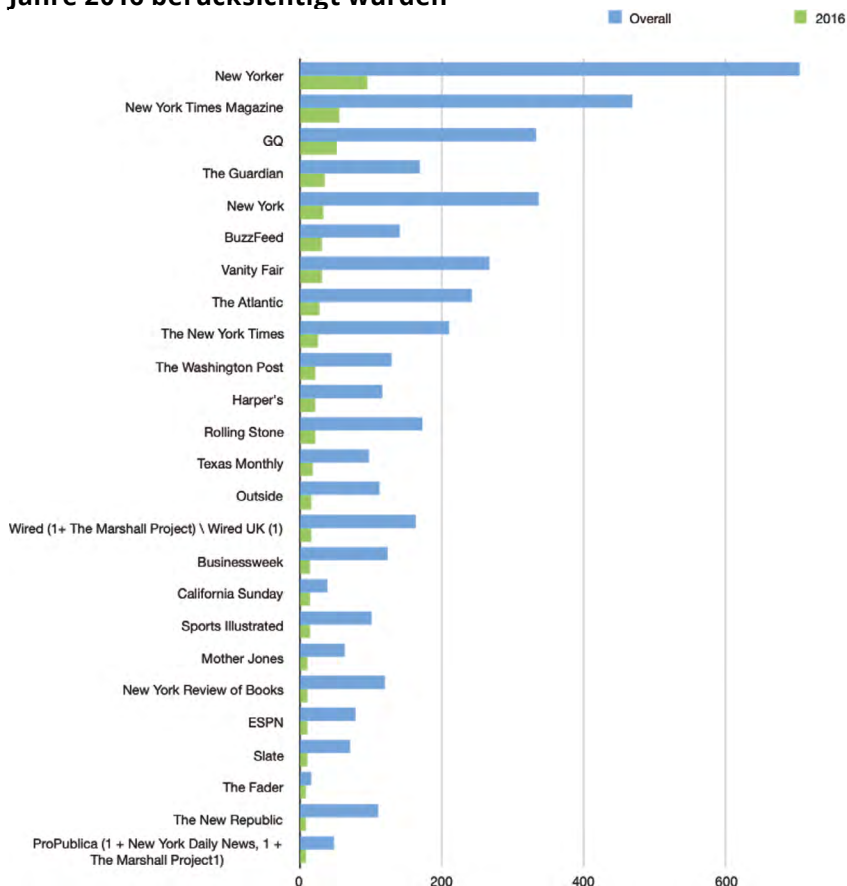


Schaubild 4 zeigt eine Übereinstimmung zwischen den Daten, die wir den Einzeleinträgen im Jahr 2016 einerseits und den Gesamtangaben von *Longform.org* andererseits entnommen haben. Wir beobachten zwar, dass es eine kleine Zahl von Medien gibt, die 2016 nicht so oft berücksichtigt wurden, während man sie in der Vergangenheit ausgiebig berücksichtigt hat, zum Beispiel *The New Republic*, *Slate*, *The New York Review of Books*, *WIRED* und *Rolling Stone*. Es ist aber anzunehmen, dass die 25 Top-Lieferanten von Einzeleinträgen im Jahr 2016 repräsentativ für die gesamte Medienauswahl von *Longform.org* sind.

Wir können also sehen, dass neben der Vielfalt die Tradition der zweite bestimmende Faktor in der Medienauswahl von *Longform.org* ist. Da die Medienpräsenz in beiden Datensätzen weitgehend übereinstimmt, richtet sich die generelle Präferenz von *Longform.org* für seine hauptsächlichen Zulieferer offenbar auf führende Informationsmedien, die für längere Zeit aktiv gewesen sind. So finden wir die Magazine *New Yorker*, *The New York Times Magazine*, *New York*, *GQ* und *Vanity Fair* unter den Hauptlieferanten.

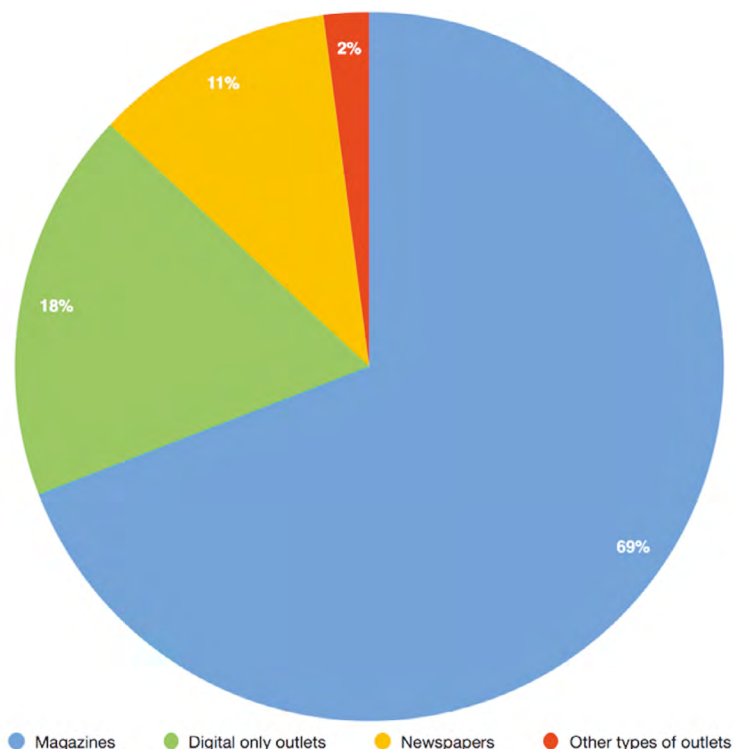
Um die Bedeutung der Informationsmedien genauer abzuschätzen, richten wir den Fokus nun auf Medien, die 2016 in mindestens zwei Einzeleinträgen von *Longform.org* auftauchen. Deren Zahl beträgt 111, die für 965 der insgesamt 1.074 Einträge verantwortlich sind. Wie wir in Schaubild 4 gesehen haben, stimmt die Auswahl der 2016 am häufigsten berücksichtigten journalistischen Medien mit der Gesamtauswahl von *Longform.org* seit der Gründung im Jahr 2010 überein. Eine Analyse des Typs von Informationsmedien, die für diese 965 Einträge verantwortlich waren, kann also generellere Einsichten darüber vermitteln, welche Art von Informationsmedien *Longform.org* bevorzugt.

Unter den insgesamt 111 Medien, die im Jahre 2016 mindestens zweimal in den Einzeleinträgen berücksichtigt wurden, sind 68 Magazine, die eine gedruckte Ausgabe behalten haben, 32 sind reine Online-Medien, 8 sind Tageszeitungen und 3 sind von Informationsmedien entwickelte Websites, die sich hauptsächlich auf einen anderen Medientyp beziehen, zum Beispiel Fernsehen oder Radio. Die Menge der Magazine, die eine auf Papier gedruckte Ausgabe behalten haben, ist also sehr wichtig, sowohl, wenn man sie mit den anderen Typen der 2016 berücksichtigten Medien und mit den wichtigsten Zulieferern für Longform-org generell vergleicht. Es gibt jedoch auch die nennenswerte Zahl (insgesamt 32) von reinen Online-Medien, die die zweitgrößte Gruppe unter den Typen Informationsmedien sind. Die Zahl der nur digitalen Publikationen erscheint im Übrigen auch deshalb wichtig, weil diese Gruppe viermal größer ist als die acht Tageszeitungen, zu denen vor allem der britische *Guardian*, die US-amerikanischen Blätter *The New York Times*, *The Washington Post*, *The Boston Globe*, *The Los Angeles Times*, *The Tampa Bay Times* sowie das norwegische *Dagbladet* gehören. Neben den Magazinen, reinen Online-Publikationen und Zeitungen gibt es drei Websites, die sich vor allem auf TV und Radio beziehen, nämlich ESPN, MTV NPR und Fusion.

Schaubild 5 zeigt die Prozentzahl jeder Medien-Kategorie an den 2016 erschienenen Einzeleinträgen. Magazine sind der wichtigste Typ von Informationsmedien. Sie machen 69 Prozent aller Einzeleinträge aus einem Pool von 68 verschiedenen Medien aus, die 61 Prozent aller berücksichtigten Informationsmedien darstellen.

Reine Online-Medien machen 18 Prozent aller veröffentlichten Einzeleinträge aus, repräsentieren aber 29 Prozent der von *Longform.org* ausgewählten Informationsmedien. Außerdem haben Tageszeitungen einen Anteil von 11 Prozent an

Schaubild 5

Anteile der *Longform.org* Einträge 2016 pro Medientyp

den Einzeleinträgen und repräsentieren gerade einmal 7 Prozent aller berücksichtigten Medien. Ähnlich lässt sich feststellen, dass die übrigen Arten von Medien (Websites von Rundfunksendern), die 4 Prozent der Einzeleinträge ausmachen, nur 2 Prozent der Informationsmedien repräsentieren.

Die prozentuale Differenz zwischen den Medientypen und der aus jeder Gruppe ausgewählten Zahl von Artikeln zeigt, dass reine Online-Medien durchschnittlich für weniger Artikel als Quelle genutzt werden. Das zeigt, dass *Longform.org* die Vielfalt des Spektrums an Informationsmedien dadurch erreicht, dass großformatige Stücke vor allem aus diversen Online-Medien berücksichtigt werden. Im Vergleich zu ihrem Anteil an den berücksichtigten Informationsmedien sind Magazine und Tageszeitungen an der Menge der Einzeleinträge überrepräsentiert. Wie Schaubild 3 gezeigt hat, bevorzugt *Longform.org* tendenziell traditionelle Informationsmedien. Und es gibt einen weiteren Faktor im Auswahlverhalten

von *Longform.org*, der klar identifiziert werden kann, weil mit Ausnahme von elf Medien alle anderen in den USA beheimatet sind. Diese elf Ausnahmen sind in Großbritannien, Deutschland, Norwegen, Kanada und Australien ansässig. Es handelt sich um *The Guardian*, *Der Spiegel*, *London Review of Books*, *Dagbladet*, *The Globe and Mail*, *Canadian Business*, die BBC, *The Economist*, *The Sidney Morning Herald*, *Toronto Star* und *Toronto Life*.

Unter diesen Medien ist *The Guardian* das einzige mit einem signifikanten Einfluss auf die Auswahl von *Longform.org*, da die Zeitung 2016 mit 35 Einzeleinträgen berücksichtigt wurde. Insgesamt beträgt die Zahl der Einzeleinträge, die von nicht in den USA ansässigen Quellen beschafft wurden, nur 56 von insgesamt 1074.

Um genauer herauszufinden, wie bestimmte Informationsmedien von *Longform.org* berücksichtigt werden, richten wir das Augenmerk noch auf die ersten Erscheinungsdaten von großformatigen Stücken, die 2016 von den zehn wichtigsten Medien beschafft wurden.

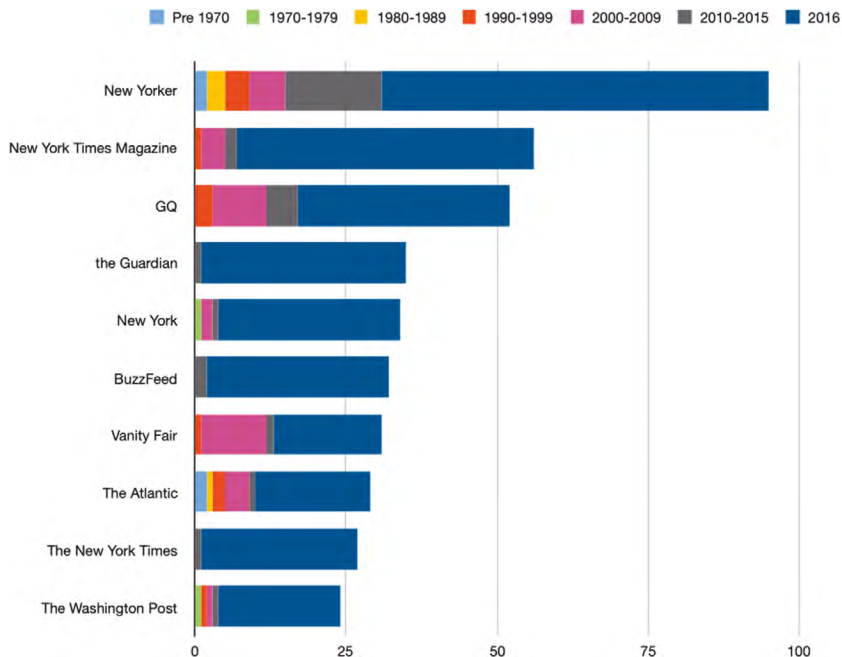
Wie wir in Schaubild 3 gesehen haben, wurden im Jahr 2016 Einzeleinträge einer großen Vielfalt von Informationsmedien entnommen. Von den 25 Medien, die 2016 häufiger als zehn Mal berücksichtigt wurden, betrachten wir die zehn wichtigsten. Um besser zu verstehen, wie großformatige Stücke aus diesen Medien ausgewählt wurden, haben wir für jede Geschichte das Erscheinungsdatum ermittelt. Die Beiträge erschienen in den folgenden zehn Magazinen: *New Yorker*, *The New York Times Magazine*, *GQ*, *The Guardian*, *New York Magazine*, *BuzzFeed*, *Vanity Fair*, *The Atlantic*, *The New York Times* und *The Washington Post*. Die ersten Erscheinungsdaten haben wir in sieben Zeitphasen unterteilt, die erste für Stücke, die vor 1970 publiziert wurden, gefolgt von fünf Dekaden und dem Ersterscheinungsjahr 2016. Wie Schaubild 6 zeigt, ist Letztere die weitaus am häufigsten besetzte Kategorie.

Nur bei zwei Medien beziehen sich die Einträge auf Stücke aus sechs verschiedenen Zeiträumen, nämlich *New Yorker* und *The Atlantic*. Bei beiden handelt es sich um Traditionsmedien, die bereits in den Jahren 1925 und 1857 gegründet wurden. Im Übrigen ist zu erkennen, dass *The New York Times Magazine*, *GQ* und *Vanity Fair* einem gleichen Muster folgen. Ihre Einträge stammen aus den gleichen vier Phasen (1990-1999, 2000-2009, 2010-2015 und 2016). Das einzige reine Online-Medium unter den Top Ten ist *BuzzFeed*, dessen Muster jedoch dem von *The Guardian* und *The New York Times* ähnlich ist, bei denen sich Einträge nur auf Stücke aus den Phasen 2010-2015 und 2016 beziehen. *The Washington Post* sticht durch ein anderes Muster hervor, mit Einträgen aus fünf verschiedenen Phasen. Aus all dem lässt sich schließen, dass Traditionsmedien aus einem vielfältigeren Pool mit verschiedenen Erscheinungsphasen berücksichtigt werden, mit einem Schwerpunkt auf Stücken, die zuerst vor 2016 publiziert wurden.

Schließlich haben wir die Zusammenhänge zwischen den Artikeln mit auf sie bezogenen kategorialen Markierungen untersucht, wofür wir ein Netzwerkdiagramm (vgl. Schaubild 7) auf der Basis der aus dem Web Scraping anfallenden

Schaubild 6

Die zehn im Jahr 2016 am häufigsten berücksichtigten Informationsmedien mit den Ersterscheinungsdaten der Stücke nach Dekaden und im Jahr 2016

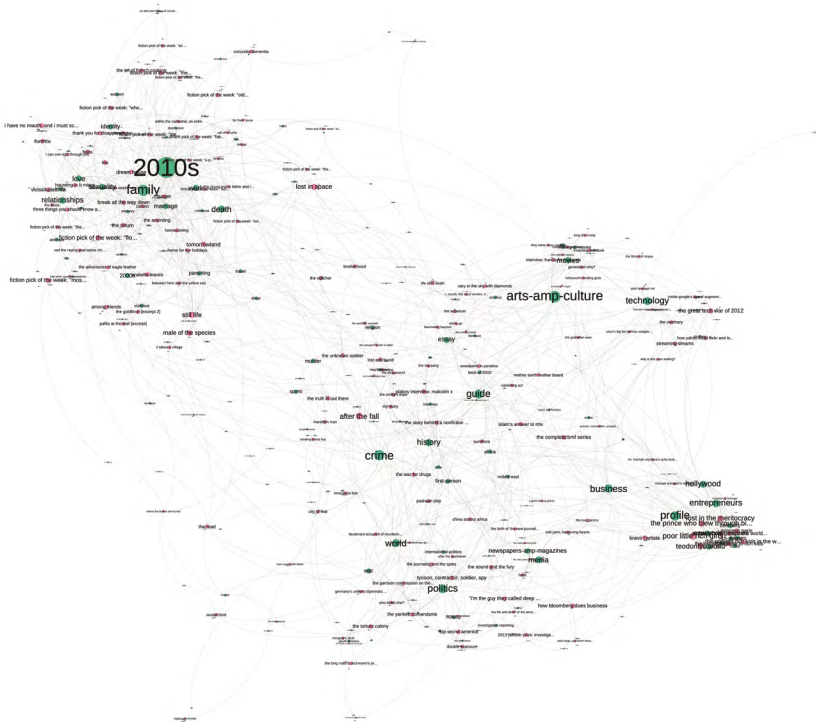


Daten mit Hilfe der Gephi-Software und des Verräumlichungs-Algorithmus Force Atlas 2 erstellen. Artikel, die ähnliche Markierungen aufweisen, rücken dabei zusammen. Indem wir dieses Netzwerkdigramm als Mittel benutzen, um in dem von *Longform.org* gesammelten und angebotenen Material inhaltliche Verknüpfungen sichtbar zu machen (Venturini, Jacomy, et al. 2018), können wir unter den Artikeln fünf Hauptcluster erkennen.

Erstens die vielleicht bedeutendste Gruppe oben links. Das häufigste Stichwort ist hier »2010s« (mit drei Vierteln aller Artikel verknüpft), das Themen wie »Liebe«, »Beziehungen«, »Sexualität«, »Identität«, »Frauen«, »Familie«, »Ehe«, »Elternschaft«, »Arbeit«, »Tod« umfasst und auf die Thematisierung von Personen und Narrativen der alltäglichen Lebenserfahrung hinweist. Zweitens befasst sich rechts unten ein Cluster mit »Hollywood«, »Unternehmer«, »Prominenz«, »Profil« sowie »Diktatoren« und »Kulte«.

Schaubild 7

Netzwerk von Artikeln und ihren zugehörigen Tags erstellt mit Hilfe der Visualisierungssoftware Gephi



Drittens geht es in einem Bereich rechts oben um »Technologie«, »Dot-coms« und »Gadgets« (technische Spielereien). Daneben gibt es viertens einen Bereich, in dem es sich um »Kino«, »Film«, und »Kunst & Kultur« dreht. Fünftens und letztens enthält eine diffusere Region zur Mitte rechts hin ein Konglomerat von Artikeln, die sich mit »Verbrechen«, »Geschichte«, »Welt«, »Politik«, »Krieg«, »Internationale Politik«, »Deutschland«, »CIA«, »Kuba«, »Weißes Haus«, »Afghanistan-Krieg« beschäftigen, was ein dauerhaftes Interesse an dramatischen Ereignissen auf der globalen Bühne anzeigt. Zu den Stichworten, die auf journalistische Genres im gesamten Netzwerk schließen lassen, gehören: Verbrechen (13%), Kunst und Kultur (8%), Essays (7%), Profile (7%), Politik (7%), Erste Person (6%), Geschäft (6%), Sport (6%) Technologie (5%), Wissenschaft (5%). Dieser kurzen Analyse ist zu entnehmen, wie die auf Longform.org markierten Artikel

die Resonanz von persönlichen, dramatischen weltpolitischen, Prominente-betreffenden, technologischen und kulturellen Themen in Archiven von großformatigem Journalismus anzeigen.

Fazit

In diesem Artikel haben wir uns auf das Portal *Longform.org* konzentriert und Daten über seine Sammel- und Kuratierungsentscheidungen vorgelegt. Indem wir an einem Beispiel ermittelt haben, wie solche Drittanbieter ihren Aktivitäten nachgehen, konnten wir eine Reihe spezifischer Praktiken identifizieren, zum Beispiel die vielfältige Auswahl von Quellenmedien und die Balance zwischen neueren und älteren Longform-Stücken. Über die journalistischen Medien, die ein Drittanbieter wie *Longform.org* auswählt, haben wir herausgefunden, dass ältere Periodika tendenziell häufiger berücksichtigt werden, da sie einen reiferen, facettenreicheren Pool von großformatigen Beiträgen bieten. Beim Sammeln und Aufbereiten von Longform-Beiträgen spielen Archive von journalistischen Informationsmedien eine wichtige Rolle, wenn es darum geht, ältere Inhalte neu zu verbreiten.

Die Sammel- und Kuratierungsaktivitäten von *Longform.org* entsprechen der Vermittlerrolle, wie sie von Anderson (2009) mit seinen Ideen zur Verteilung und zum Konsum von Longtail-Angeboten im Internet theoretisiert wurde.

Andersons Schwerpunkt lag auf der Effizienz der Verbreitung und nicht auf der Verfügung über ein bestimmtes Produkt (vgl. Anderson 2009) – in unserem Fall den des großformatigen Journalismus. Etablierte Unternehmen überflügelten in der Regel traditionelle Branchen, indem sie in großem Umfang neue Vertriebspraktiken einsetzen, wie im Fall der Nutzung digitaler Kommunikationstechnologie durch soziale Netzwerke, die – neben anderen Aktivitäten – eine Vermittlerfunktion zwischen Produzent(inn)en und der Menge der Konsument(inn)en übernehmen. Der Fall von *Longform.org* ist dafür höchst relevant, da er nicht als durchgeplantes Unternehmen begann. Wie die beiden Gründer anmerkten: »Das Publikum, das kam, wurde einfach immer größer, ohne dass wir viel taten« (McQuade 2015). Es ist daher davon auszugehen, dass in der Gegenwart das Sammeln und Aufbereiten durch digitale Vermittler entscheidende Erfolgsfaktoren sind, während die Bedeutung von Produktion und Besitz abnimmt. Der Erfolg von *Longform.org* als Sammelportal für großformatigen Journalismus zeigt, dass solche Einrichtungen erfolgreich Publikum und Aufmerksamkeit auf sich ziehen, während traditionelle Informationsmedien damit kämpfen, für ihre eigenen Inhalte effektive Vermittlungspraktiken zu entwickeln.

Es gibt mehrere Richtungen, in die Sammel- und Kuratierungsaktivitäten durch Text- und Daten-Mining-Prozesse sich entwickeln können, besonders

wenn der Datafizierungsprozess der digitalen Medienarchive in Zukunft fort-schreitet. Studien, die auf personenzentriertem Mining (Coll Ardanuy et al. 2016) und auf historischer Geodatenextraktion (Yzaguirre et al. 2016) basieren, zeigen, dass es einen produktiven Raum für neue Arten von nutzerorientierten Sammel- und Kuratierungsdiensten gibt, die auf digitale Medienarchive zurückgreifen, sobald diese datifiziert sind. Diese neuen Arten des Aufbereitens und Sammelns scheinen auf journalistische Medien zugeschnitten zu sein, die neue Verbreitungsmöglichkeiten für ihre archivierte Produktion finden wollen.

Die Beschäftigung mit den Praktiken von Portalen wie *Longform.org* kann uns helfen zu verstehen, wie journalistische Medien die Online-Verbindung mit Archiven organisieren und die Beziehung des Publikums zu Inhalten aus der Vergangenheit neu gestalten können – nicht zuletzt durch Rekontextualisierung, Rekombination, Neubewertung und Verbreitung mit Hilfe digitaler Kommunikationstechnologie und Infrastruktur. Die Daten und Aufbereitungspraktiken solcher Portale sollten als ein Arbeitsbereich des gegenwärtigen Journalismus verstanden werden, der die Bedingungen dafür schafft, dass frühere Perspektiven im Web leichter verfügbar, erfahrbar und programmierbar sind.

Übersetzung: Horst Pöttker, Johannes Rabe

Über die Autoren

Dr. Marco Braghieri ist Wissenschaftlicher Mitarbeiter im Bereich »Social Big Data« in der Abteilung Digitale Geisteswissenschaften (Digital Humanities) am King's College in London. Er hat zu mehreren europäischen Projekten beigetragen, z. B. SoBigData und EHRI (European Holocaust Research Infrastructure). Seine Forschungsarbeit konzentriert sich auf den digitalen Journalismus und seine Überschneidung mit Kommunikationsplattformen. Kontakt: marco.braghieri@kcl.ac.uk

Professor Tobias Blanke ist Lehrstuhlinhaber für Geisteswissenschaften und Künstliche Intelligenz (Artificial Intelligence) an der Universität von Amsterdam. Er ist Professor für Soziale und Kulturelle Informatik und Leiter der Abteilung Digitale Geisteswissenschaften am King's College in London. Seine Forschungsarbeit konzentriert sich auf Große Datenmengen (Big Data), Künstliche Intelligenz (AI) und ihre Einflüsse auf Kultur und Gesellschaft. Kontakt: t.blanke@uva.nl

Dr. Jonathan Gray ist Lehrbeauftragter im Bereich Kritische Infrastrukturen in der Abteilung Digitale Geisteswissenschaften (Digital Humanities) am King's College in London, wo er zur Zeit ein Buch über Datenwelten (Data Worlds) schreibt.

Er ist auch Mitbegründer des Public Data Lab und Wissenschaftlicher Mitarbeiter (Research Associate) im Bereich der Digitale-Methoden-Initiative (Digital Methods Initiative) der Universität von Amsterdam und dem médialab (Sciences Po, Paris). Mehr über seine Arbeit findet sich unter jonathangray.org. Er twittert als @jwyg. Kontakt: jonathan.gray@kcl.ac.uk

Literatur

- Albalad Aiguabella, José Maria (2015): Slow journalism para una nueva audiencia digital. El caso de Longform. org (2010-2015). In: *Revista de Comunicación*, 14, S. 7-25.
- Anderson, Chris (2009): *The longer long tail: how endless choice is creating unlimited demand*. New York, NY: Random House Business.
- Barker, Deane (2016): *Web Content Management: Systems, Features, and Best Practices*. Sebastopol, CA: O'Reilly Media.
- Belt, Don and South, Jeff (2016): Slow Journalism and the Out of Eden Walk. In: *Digital Journalism*, 4(4), S. 547-562.
- Bercovici, Jeff: Longform's New App: More Great Journalism Without The Filter. In: *Forbes*, 17.9.2014. <https://www.forbes.com/sites/jeffbercovici/2014/09/17/longforms-new-app-more-great-journalism-without-the-filter/#552df27b42f6> (1.2.2021)
- Blanke, Tobias; Prescott, Andrew (2016): Dealing with Big Data. In: Griffin, Gabriele and Hayler, Matt (Hrsg.): *Research Methods for Reading Digital Data in the Digital Humanities*, Edinburgh: Edinburgh University Press, S. 184-205.
- Bødker, Henrik (2018): Journalism History and Digital Archives. In: *Digital Journalism*, 6(9), S.1113-1120.
- Boynton, Robert (2013): Notes toward a Supreme Nonfiction: Teaching Literary Reportage in the Twenty-first Century. In: *Literary Journalism Studies*, 5(2), S. 125-131.
- Brock, George (2013): *Out of print: journalism and the business of news in the digital age*. London: Kogan Page Limited.
- Bruns, Axel; Highfield, Tim (2012): Blogs, Twitter, and breaking news: The produsage of citizen journalism. In R.A. Lind (Hrsg.): *Producing theory in a digital world: The intersection of audiences and production in contemporary theory*. New York, London: Peter Lang.
- Chyi, Hsiang Iris; Chadha, Monica (2012): News on New Devices. In: *Journalism Practice*, 6(4), S. 431-449.
- Coll Ardanuy Mariona; Knauth Jürgen; Beliankou Andrei; van den Bos Maarten; Sporleder Caroline (2016): Person-Centric Mining of Historical Newspaper Collections. In: Fuhr Norbert, Kovács László, Risse Thomas, Nejd Wolfgang (Hrsg.)

- Research and Advanced Technology for Digital Libraries*. TPD L 2016, Lecture Notes in Computer Science, vol. 9819. London: Springer, S. 320-335.
- Data Miner (2016): How Data Miner Works. In: *Data Miner* <https://data-miner.io/how-it-works> (3.10.2020)
- Derrida, Jacques (2005): *Paper machine*. Stanford, CA: Stanford University Press.
- Dowling, David; Vogan, Travis (2015): »Can We »Snowfall« This?«. In: *Digital Journalism*, 3(2), S. 209-224.
- Elliott, Chris: The readers' editor on... the richness of digital archives bringing problems for readers and journalists. In: *The Guardian*, 21.10.2012. <https://www.theguardian.com/commentisfree/2012/oct/21/digital-archives-problems-readers-journalists> (9.1.2021)
- Fisher, Mark (2009): *Capitalist realism: is there no alternative?* Alresford: o Books.
- Franklin, Bob (2014): The Future of Journalism. In: *Journalism Studies*, 15 (5), S. 481-499.
- Hartsock, John C. (2000): *A History of American Literary Journalism: The Emergence of a Modern Narrative Form*. Amherst, MA: University of Massachusetts.
- Heyman, Stephen: Google Books: A Complex and Controversial Experiment. In: *The New York Times*, 29.10.2015. https://www.nytimes.com/2015/10/29/arts/inter-national/google-books-a-complex-and-controversial-experiment.html?_r=0 (25.3.2021)
- Hiippala, Tuomo (2017): »The Multimodality of Digital Longform Journalism«. In: *Digital Journalism*, 5(4), S. 420-442.
- Huang, Sonia J. and Wang, Wei-Ching (2014): Application of the long tail economy to the online news market: Examining predictors of market performance. In: *Journal of Media Economics*, 27(3), S. 158-176.
- Jacobson, Susan; Marino, Jaqueline; Gutsche Robert E Jr; Reynolds, Donald W (2018): Should There Be an App for That? An Analysis of Interactive Applications within Longform News Stories. In: *Journal of Magazine Media*, 18(2).
- Jacomy, Mathieu; Venturini, Tommaso; Heymann, Sebastien; Bastian, Mathieu (2014): ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. In: *PLOS ONE*, 9(6), e98679. DOI: <https://doi.org/10.1371/journal.pone.0098679>
- Kachka, Boris: The Future of Reading According to Longform. In: *New York Magazine*, 26.9.2014. <http://nymag.com/daily/intelligencer/2014/09/future-of-reading-according-to-longform.html> (26.3.2021)
- Le Masurier, Megan (2015): What is slow journalism? In: *Journalism practice*, 9(2), S. 138-152.
- Le Masurier, Megan (2016): Slow Journalism: An introduction to a new research paradigm. In: *Journalism Practice*, 10(4), S. 439-447.
- Longform.org (2010): Longform.org. <http://www.longform.org> (24.6.2020)

- Longform.org (2017): Publications. In: Longform.org. <https://longform.org/archive/publications?from=1> (24.5.2020)
- Longform.org: Why We've Removed Longform from the App Store. In: Longform Medium, 26.9.2014. <https://medium.com/@longform/why-weve-removed-longform-from-the-app-store-823d599a34d4> (29.1.2021)
- Longhi, Raquel Ritter; Winques, Kérley (2015): The place of longform in online journalism: Quality versus quantity and a few considerations regarding consumption. In: *Brazilian Journalism Research*, 11(1), S. 104-121.
- Marres, Noortje; Weltevrede, Esther (2013): Scraping the Social? In: *Journal of Cultural Economy*, 6(3), S. 313-335.
- Mayer-Schönberger, Viktor; Cukier, Kenneth (2013): *Big data: a revolution that will transform how we live, work and think*. London: John Murray.
- McQuade, Eric (2015): Aaron Lammer, The Art of Podcasting. In: *The Timbre*, 20 February 2015. <https://web.archive.org/web/20150906075223/http://thetimbre.com/aaron-lammer-art-podcasting-no-5/> (9.12.2020)
- Mitchell, Ryan (2015): *Web Scraping with Python: Collecting Data from the Modern Web*. Sebastopol, CA: O'Reilly Media.
- Mullin, B.: New app from Longform allows freelancers to cultivate audiences. In: *Poynter*, 18.9.2014. <https://www.poynter.org/news/new-app-longform-allows-freelancers-cultivate-audiences> (14.1.2021)
- Nel, François; Westlund, Oscar (2012): The 4C's of Mobile News. In: *Journalism Practice*, 6(5-6), S. 744-753.
- Olson, Jack E. (2003): *Data Quality: The Accuracy Dimension*. Burlington, MA: Morgan Kaufmann Publishers.
- Pybus, Jennifer; Coté Mark; Blanke, Tobias (2015): Hacking the social life of Big Data. In: *Big Data & Society*, 2(2), S. 1-10.
- Rogers, Richard (2013): *Digital Methods*. Cambridge, MA: MIT Press.
- Seaton, Jean (2016): The new Architecture of Communications. In: *Journalism Studies*, 17(7), S. 808-816.
- Shapiro, Michael; Hiatt, Anna; Hoyt, Mike (2015): *Tales From the Great Disruption—Insights and Lessons From Journalism's Technological Transformation*. New York, NY: Big Roundtable Books.
- Smith, Virginia; Connor, Miriam; & Stanton, Isabelle (2015): Going In-Depth: Finding Longform on the Web. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, S. 2109-2118.
- Stensen, Steen; Grøndahl Larsen, Anna M.; Hågvar, Yngve Benestad; Fonn, Brigitte Kjøss (2019): What Does Digital Journalism Studies Look Like? In: *Digital Journalism*, 7(3), S. 320-342.
- Venturini, Tommaso; Bounegru, Liliana; Gray, Jonathan; Rogers, Richard (2018): A reality check(list) for digital methods. In: *New Media & Society*, 20(11), S. 4195-4217.

- Venturini, Tommaso; Jacomy, Mathieu; Bounegru, Liliana; Gray, Jonathan (2017): Visual network exploration for data journalists. In: Scott Eldridge II and Bob Franklin (Hrsg.): *The Routledge Handbook to Developments in Digital Journalism Studies*. Abingdon: Routledge.
- Verborgh, Ruben; De Wilde, Max (2013): *Using OpenRefine*. Birmingham: Packt Publishing.
- Yzaguirre, Amelia; Smit, Mike J.; Warren, Rob (2016): Newspaper archives+ text mining= rich sources of historical geo-spatial data. In: *IOP Conference Series: Earth and Environmental Science*, 34(1), S. 1-8.