Digitale Methoden: Zur Positionierung eines Ansatzes*

Richard Rogers**

Ziel dieses Aufsatzes ist es, ein Grundverständnis von digitalen Methoden und ihrer Anwendung zu vermitteln, insbesondere im Hinblick auf die Nutzung von Software bei dieser Art der Forschung. Dazu wird in einem ersten Schritt argumentiert, warum es bei einer Forschung mittels digitaler Methoden nicht darum geht, Webdaten im Vergleich mit der sozialen Welt jenseits des Webs zu „überprüfen". Vielmehr geht es darum, das Web in seiner eigenen Spezifik für die Forschung nutzbar zu machen. Entsprechend setzen digitale Methoden bei „nativ", d. h. originär digitalen Daten an und unterscheiden sich so von anderen Methoden des Computational Turn in den Sozial- und Geisteswissenschaften. Dies konkretisiert sich in der Forschungspraxis der digitalen Methoden, die anhand der Nutzung des Internet Archive, der Google Websuche, von Wikipedia, Facebook, Twitter und YouTube sowie plattformübergreifenden Studien veranschaulicht wird. Solche Beispiele machen deutlich, dass es bei digitalen Methoden um eine spezifische „Wiederverwendung" von Online-Daten geht, ähnlich wie bei anderen non-responsiven Methoden.

Schlüsselwörter: Digitale Methoden, Online-Forschung, Internet-Forschung, digitale Kommunikation, Forschungssoftware

Digital Methods: On the Positioning of an Approach

The aim of this paper is to provide a basic understanding of digital methods and their application, particularly in regard to the use of software when doing research. To explain this approach, the first section of this paper relates why research using digital methods is not about "checking" web data against the social world, rather, it is about making the particularity of the web functional for general research. Accordingly, digital methods start with "native", that is, originally digital data and objects which differ from other approaches that make up part of the computational turn in the social sciences and humanities. In the second part an approach to digital methods is discussed from a more practical point of view, illustrated through the application of several online services including the Internet Archive, Google Web Search, Wikipedia, Facebook, Twitter, and YouTube, as well as through a description of cross-platform studies. These examples help demonstrate the ways in which digital methods are more about a specific "repurposing" of online data, similar to other non-reactive methods.

Keywords: digital methods, online research, internet research, digital communication, research software

1. Einleitung: Digitale Methoden als Ansatz internetbasierter Forschung

Digitale Methoden sind Verfahren zur Erforschung des Wandels von Gesellschaft und von Kultur mit Hilfe von Online-Daten. Sie nutzen verfügbare digitale Objekte wie

* Bei diesem Aufsatz handelt es sich um eine Übersetzung des Einführungsbeitrags in Richard Rogers (2019): Doing Digital Methods. London: Sage, S. 3–20.

** Prof. Dr. Richard Rogers, Media Studies, University of Amsterdam, Turfdraagsterpad 9, 1012 XT Amsterdam, the Netherlands, https://www.digitalmethods.net, https://orcid.org/0000-0002-9897-6559.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Hyperlinks, Tags, Zeitstempel, Likes, Shares und Retweets und versuchen Erkenntnisse daraus zu generieren, wie diese Objekte durch *deren* „Methoden" verarbeitet werden, die in den entsprechenden Online-Geräten, wie z. B. in der Google Websuche, eingeschrieben sind. Dem liegt eine spezifische Definition von digitalen Objekten zugrunde. Nativ digitale Objekte sind demnach medien- oder plattformspezifische Computerkonstrukte, ob auf der Ebene des Interfaces, einer Ebene unterhalb desselben oder der Ebene der Programmierschnittstelle (API-Parameter), die sich für eine (Um-)Nutzung im Rahmen der Sozial- und Kulturforschung eignen und so ermöglichen. Digitale Methoden sind also mit dem Bemühen verbunden, bestimmte digitale Objekte für eine sozialwissenschaftliche Perspektive nutzbar zu machen. Letztlich geht es um die Frage, was die Grundlage dieser Methoden ist und wie diese online „verankert" werden können – im Sinne des Englischen „grounded".

Als Forschungspraxis sind digitale Methoden Teil des Computational Turn in den Geistes- und Sozialwissenschaften und stehen als solche neben anderen neueren Ansätzen wie Cultural Analytics, Culturomics, Webometrics und Altmetrics. All diese Ansätze unterscheiden sich im Hinblick auf die Art der verwendeten Daten (nativ digitale oder digitalisierte Daten) sowie im Hinblick auf die Methode ihrer Generierung (für das Medium entwickelt oder dorthin migriert). Dabei haben – wie ich im Weiteren zeigen möchte – digitale Methoden ihre Grenzen. Digitale Methoden setzen sich mit den Problemen von Webdaten auseinander, wie etwa mit der Unbeständigkeit von Webdiensten und der Instabilität von Datenströmen durch umkonfigurierte oder eingestellte APIs. Setzt man digitale Methoden ein, muss man sich also auch mit der Qualität von Webdaten befassen sowie mit den Herausforderungen von Längsschnittstudien. Ein aus der Literatur bekanntes Beispiel für solche Herausforderungen war die Löschung von Twitter-Konten und Facebook-Seiten, nachdem Forscher*innen begonnen hatten, die Reichweite der russischen Desinformationskampagne während der US-Präsidentschaftswahlen zu untersuchen (Albright 2017).

Wenn man sich dem Web als Ort zur Untersuchung sozialer und kultureller Phänomene zuwendet, ist man mit einer Reihe von Bedenken konfrontiert. Webdaten gelten als problematisch und haben ein schon historisches Reputationsproblem, insofern als das Web als Medium der Selbstveröffentlichung behandelt und untersucht wurde, aber auch als Medium mit zweifelhaftem Ruf, bevölkert von Pornograph*innen und Verschwörungstheoretiker*innen (Dean 1998). Zur Zeit des „Cyberspace" wurde das „Anything-Goes-Web" analytisch oft als ein separater Bereich, sogar als eine „virtuelle Gesellschaft" (Woolgar 2003) behandelt. Später wurde das Web insbesondere als Raum der Produktion von nutzergenerierten Inhalten erforscht (Jenkins 2006). Heutzutage ist das Web nicht mehr nur ein Raum zur Erforschung von Online-Kulturen. Es ist vielmehr zu einem Untersuchungsort für sehr unterschiedliche kulturelle und soziale Fragen insgesamt geworden, beispielsweise für Untersuchungen über „'concerns of the electorate' from the 'searches they conduct'" oder „the spread of arguments … about political and other issues" (Lazer et al. 2009: 722; Watts 2007). Selbstverständlich bleibt das Web auch ein Ort, um Online-Kulturen zu studieren und Medienforschung zu betreiben. Digitale Methoden sind Ansätze, beides zu untersuchen. Dies ist ein Punkt, auf den ich zurückkomme, wenn ich die Frage behandele, ob man Medienartefakte (wie manipulierte Suchmaschinenergebnisse oder Bots) aus einem Datensatz entfernen sollte, um einen bereinigten Untersuchungsgegenstand zu haben.

Ausgehend von diesen Überlegungen möchte ich im Weiteren ein Grundverständnis von digitalen Methoden und ihrer Anwendung vermitteln, insbesondere im Hinblick auf die Nutzung von Software bei dieser Art der Forschung. Dazu wird im folgenden

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Abschnitt 2 argumentiert, warum es bei einer Forschung mittels digitaler Methoden nicht darum geht, Webdaten anhand eines Vergleichs mit der „weiteren" sozialen Welt zu „überprüfen", sondern warum hier Webdaten in ihrer eigenen Spezifik für die Forschung nutzbar gemacht werden. Hieraus ergibt sich – wie im darauffolgenden Abschnitt 3 deutlich gemacht wird –, dass digitale Methoden bei „nativ", d. h. originär digitalen Daten ansetzen und sich so von anderen Methoden des Computational Turn in den Sozial- und Geisteswissenschaften unterscheiden. Praktisch illustriert wird die Forschungspraxis mit digitalen Methoden im Abschnitt 4 anhand der Nutzung des Internet Archive, der Google Websuche, von Wikipedia, Facebook, Twitter und YouTube bzw. plattformübergreifenden Studien. All dies macht deutlich, dass es bei digitalen Methoden um eine spezifische „Wiederverwendung" von Online-Daten geht.

2. Jenseits der Überprüfung von Webdaten: Zur Online-Verankerung digitaler Methoden

Wie bereits angedeutet, hat das Web die generelle Schwierigkeit, an wissenschaftliche Daten gestellte Standards zu erfüllen (Borgman 2009). Als solche können Webdaten allerdings auch eine zumindest geringfügige Verschiebung der methodologischen Sichtweise bewirken. Wenn Webdaten oft als „messy" und „poor" angesehen werden, worin könnte dann ihr Wert liegen? Oder umgekehrt: Wo und wie können Webdaten fruchtbar gemacht werden? Digitale Methoden versuchen, von den sogenannten Methoden des Mediums zu lernen, also davon, wie Online-Geräte und -Dienste mit Webdaten umgehen (Rogers 2009). Digitale Methoden bedeuteten also zunächst das Studium der Methoden, die in jenen Geräten eingeschrieben sind, die Online-Daten generieren (Rieder 2012). Wie behandeln Suchmaschinen (wie Google) Hyperlinks, Klicks, Zeitstempel und andere digitale Objekte? Wie gehen Plattformen wie Facebook mit Profilinteressen und Nutzerinteraktionen wie Liken, Teilen und Kommentieren sowie dem Liken von Kommentaren um?

Digitale Methoden versuchen, eine sozialwissenschaftliche Perspektive in die Untersuchung von Online-Diensten einzubringen. „Nowcasting" (ein neumodischer Begriff für Echtzeitprognosen) ist ein gutes Beispiel dafür, wie Suchmaschinenabfragen zur Erforschung des sozialen Wandels eingesetzt werden können (Ginsberg et al. 2009). So wurde etwa über die Erfassung von Ortsangaben und die Intensität von Suchanfragen zum Themenfeld „Grippe" die steigende und fallende Inzidenz der Grippe an bestimmten Orten ermittelt. „Grippe-Orte" zu ermitteln, ist eine kreative Nutzung von Webdaten für die Sozialforschung, die die Palette der „Trend-Forschung" erweitern, die Suchmaschinen bisher unter Namen wie Google Trends, Google Insights for Search, Yahoo Buzz Log, Yahoo Clues, Bing Webmaster Keyword Research, AOL Search Trends, YouTube Keyword Tool, YouTube Trends und das Google AdWords Keyword Tool bieten (Raehsler 2012; US Centers for Disease Control and Prevention 2014). Es handelt sich aber auch um ein Beispiel, bei dem die Datenbasis nicht ausschließlich aus Webdaten besteht, sondern aus (triangulierten) Ergebnissen traditioneller Grippeüberwachungsmethoden, wie sie von den Centers for Disease Control and Prevention in den USA und ihren Äquivalenten in anderen Ländern verwendet werden. Hierbei werden die Daten der Suchmaschinenabfragen mit Daten von Krankenhäusern, Kliniken, Labors, staatlichen Behörden und anderen abgeglichen. Offline wird zum Prüfstein, an dem die Qualität des Online gemessen wird.

Um Webdaten zur Untersuchung sozialer Phänomene einzusetzen, hat der Webometrician Mike Thelwall eben diese Vorgehensweise vorgeschlagen: die Ergebnisse offline zu erden. Angesichts der Unübersichtlichkeit („messiness") von Webdaten sowie

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

der (historischen) Skepsis, die – wie einleitend erwähnt – ihre Verwendung in der Sozialforschung begleitet, entwickelten Thelwall et al. (2005) eine allgemeine Begründung für eine Forschungsstrategie, die einer Offline-Korrelation bedarf:

„One issue is the messiness of Web data and the need for data cleansing heuristics. The uncontrolled Web creates numerous problems in the interpretation of results […] Indeed a sceptical researcher could claim the obstacles are so great that all Web analyses lack value. One response to this is to demonstrate that Web data correlate significantly with some non-Web data in order to prove that the Web data are not wholly random." (Thelwall et al. 2005: 81)

Digitale Methoden werfen also die Frage nach der *Online-Verankerung auf* – der „online groundedness": Wann und unter welchen Bedingungen können Befunde aus Webdaten durch das Web selbst abgesichert werden? Eines der frühen Beispiele, das den eigenständigen Aussagegehalt von Webdaten deutlich macht, ist ein sowohl journalistisches als auch experimentelles. In einem ausführlichen Beitrag in der niederländischen Qualitätszeitung *NRC Handelsblad* fragte der Journalist Joep Dohmen, ob sich die niederländische Kultur angesichts der Morde an dem populistischen Politiker Pim Fortuyn und dem Kulturkritiker Theo van Gogh Mitte der 2000er Jahre sowie angesichts der darauffolgenden Gegenreaktion verhärtet habe (Dohmen 2007). Unter „Verhärtung der Kultur" verstand er die Abnahme von Toleranz gegenüber anderen und die Zunahme von Segmenten sich radikalisierender und extremistischer Individuen in der Gesellschaft. Die für den Beitrag angewandte Methode ist nun interessant im Hinblick darauf, wie Webdaten ein eigener Wert zugesprochen wird. Anstatt sich unter Hooligans zu begeben, Flugblätter und andere gedruckte Produkte zu studieren oder Expert*innen zu befragen, wandte sich der Journalist in seiner Recherche dem Web zu. Er stellte Listen rechtsgerichteter und extremistischer Websites zusammen und verglich die Sprache beider Typen von Websites mit Hilfe der Wayback Machine des Internet Archive im Zeitverlauf. Auf diese Weise wurde deutlich, dass sich die Sprache auf den rechtsgerichteten Seiten im Laufe der Zeit immer mehr der Sprache auf den extremistischen Seiten annäherte. Dieses Vorgehen war zwar journalistisch motiviert, ihm liegt aber ein Denken zugrunde, das nahe an der Praxis der Sozialforschung ist: die Veränderung von Sprache im Web über die Zeit zu erfassen, um sozialen Wandel zu untersuchen. Zusammen mit dem Artikel wurde außerdem der Datensatz veröffentlicht, was für Zeitungen ungewöhnlich ist und das Aufkommen des Datenjournalismus einläutete. Der Journalist wertete die Websites mit einem Close-Reading-Ansatz aus; man könnte sich aber ebenso gut vorstellen, die Quellen mit einem Distant-Reading-Ansatz auszuwerten, wie er typischerweise mit dem Computational Turn und Big-Data-Studien in Verbindung gebracht wird (Moretti 2005; boyd & Crawford 2012).

Ein weiteres datenjournalistisches Projekt, das die Grundidee der digitalen Methoden veranschaulicht, war die Visualisierung von Rezeptanfragen in einem Kartogramm, wie sie in der *New York Times* erschienen ist (Ericson & Cox 2009). Dafür wurden alle Rezepte (auf allrecipes.com), die am Tag vor dem amerikanischen Thanksgiving-Fest abgefragt wurden, geolokalisiert und die Orte ermittelt, von denen die Suchanfragen erfolgten. Für jedes Rezept wurde die Karte nach Häufigkeit der Suchanfragen pro Bundesstaat schattiert (und statistisch normalisiert), wodurch Unterschiede bei den Rezeptanfragen und unter Umständen auch den Essensvorlieben innerhalb der Vereinigten Staaten erkennbar sind. Die Karte zeigt so im weiteren Sinne eine Geographie des Geschmacks.

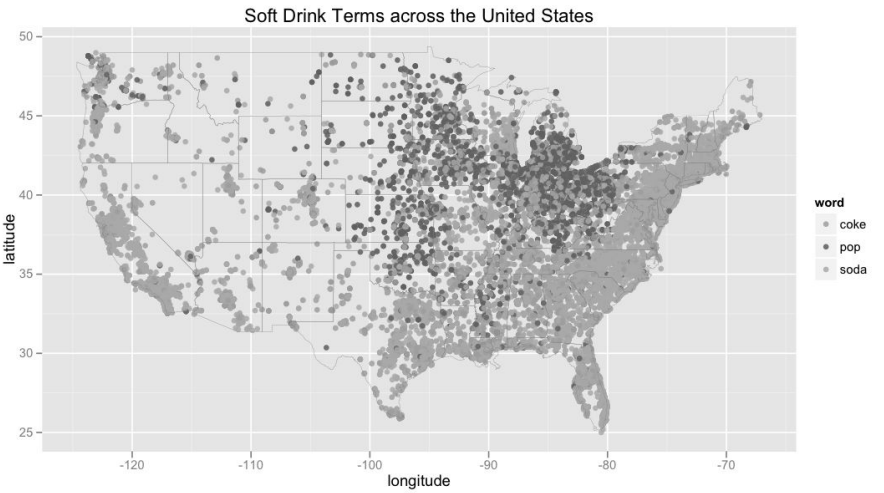
Hier aber stellt sich die Frage, wie diese Art von Ergebnissen nun „verankert" werden kann? Vergleicht man sie offline mit Umfragen oder regionalen Kochbüchern, oder sucht man eher nach Online-Daten, wie z. B. Fotos von Lebensmitteln, die mit Orts- und

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Zeitstempeln versehen sind? Würde dann Flickr oder Instagram eine bessere „Verankerung“ bieten? Wird das Web selbst also zum möglichen Ort der „Verankerung“?

Um eine solche Verankerung im Web zu veranschaulichen, bietet sich das Beispiel einer wissenschaftlichen Untersuchung an, bei der Online-Daten zur Erforschung regionaler Unterschiede verwendet wurden. Ausgangspunkt dieser Studie war die klassische Diskussion über sprachliche Variationen bei der Verwendung der Begriffe „Soda“, „Pop“ und „Coke“ in den USA. Edwin Chen (2012) hat für die Untersuchung dieser regionalen Variationen mit Geotags versehene Tweets mit den Wörtern „Soda“, „Pop“ oder „Coke“ erfasst und so eine Geografie der Wortverwendung auf einer Karte dargestellt (siehe Abbildung 1). Die Ergebnisse wurden mit denen einer älteren Web-Datenerfassungstechnik verglichen, der eine online migrierte Umfragemethode zugrunde lag – eine „virtuelle Methode“, die im nächsten Abschnitt erörtert wird. Für die dafür notwendige Online-Datensammlung wurde eine Webseite erstellt, auf der die Teilnehmer*innen gebeten wurden, ihren bevorzugten Begriff („Soda“, „Pop“, „Coke“ oder andere) auszuwählen und ihren Heimatort einschließlich Bundesland und Postleitzahl anzugeben (siehe Abbildung 2). Die resultierende Karte zeigt eine stärkere regionale Differenzierung als die Twitter-Analyse. Chen bestätigte zwar nicht die früheren Ergebnisse, berichtet aber von „ähnlichen Mustern“, wonach „Pop“ ein Begriff aus dem mittleren Westen ist, „Coke“ eher im Süden und „Soda“ im Nordosten verwendet wird (Chen 2012; Shelton 2011).

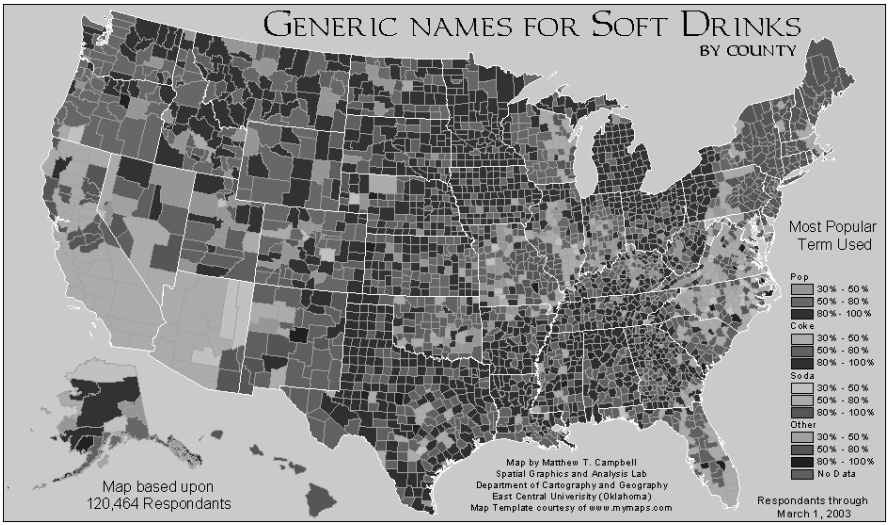
Abbildung 1: US-Karte mit Verteilung der Verwendung von Begriffen für Softdrinks (in geo-getaggten Tweets in Twitter), 2012



Quelle: Chen 2012.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Abbildung 2 US-Karte der selbstberichteten Verwendung von Begriffen für Softdrinks, 2003



Quelle: Campbell 2003.

3. Das Nativ-Digitale und das Digitalisierte: virtuelle und digitale Methoden im Vergleich

Digitale Methoden können von anderen Ansätzen, die sich nach dem Computational Turn in den Sozialwissenschaften und den Digital Humanities entwickelt haben, abgegrenzt werden (siehe Tabelle 1). Ebenso wie andere aktuelle Ansätze zur Untersuchung digitaler Daten verwenden sie Verfahren, die auf Online-Abfragen basieren, sind also durch eine Forschungspraxis gekennzeichnet, die man als „Forschung durch Datenabfrage" („search as research") bezeichnen kann. Sie unterscheiden sich jedoch von anderen Ansätzen dadurch, dass sie sich weitgehend auf nativ-digitale Daten und ebensolche Methoden stützen, im Gegensatz zu digitalisierten Daten und migrierten Methoden.

Tabelle 1: Ansätze zur Untersuchung digitaler Daten

		Methoden	
		digitalisiert	**nativ-digital**
Daten	**digitalisiert**	►Culturomics* ►Cultural Analytics*	
	nativ-digital	►Webometrics ►Altmetrics	►Digitale Methoden

Diese Tabelle vergleicht digitale Methoden mit anderen Verfahren nach dem Computational Turn in den Geistes- und Sozialwissenschaften im Hinblick auf ihre Verwendung nativ-digitaler oder digitalisierter Daten und Methoden. Quelle: Weiterentwickelt auf Basis von Rogers 2014.
* Basieren auf einer Forschung durch Datenabfrage

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Digitale Methoden lassen sich mit zwei verbreiteten Ansätzen der Digital Humanities vergleichen: den Culturomics und den Cultural Analytics (siehe Tabelle 1). Während digitale Methoden mit nativ-digitalen Daten arbeiten, haben Culturomics und Cultural Analytics als Korpus das, was man als digitalisierte Materialien bezeichnen könnte, die dann im Hinblick auf einzelne Lemma (in Culturomics) oder im Hinblick auf formale Materialeigenschaften (in Cultural Analytics) ausgewertet werden. Das Verfahren der Culturomics greift beispielsweise auf Daten von Google Books zurück und ermöglicht so Längsschnittstudien zur Veränderung des Sprachgebrauchs anhand einzelner Wörter oder Phrasen, um daraus breitere kulturelle Trends abzuleiten. Eine solche Forschung kann etwa zeigen, dass die amerikanische Schreibweise allmählich die britische verdrängt, Wendungen wie „celebrity is gained" oder „fame is shorter-lived" sich verbreiten (Michel et al. 2011). Cultural Analytics ist ein Forschungsverfahren, das ebenfalls auf Datenabfragen basiert, allerdings mit einer geringeren Datenverarbeitungstiefe. Hier werden Muster und deren Veränderung nicht anhand von Wörtern untersucht, sondern anhand von formalen Eigenschaften der Medieninhalte, wie dem Farbton, der Helligkeit und der Sättigung von Bildern (siehe u. a. Manovich 2020).

Digitalisierte Daten werden, wie erwähnt, oft als „qualitativ höherwertig" angesehen als Webdaten. Ein Vorteil sowohl von Culturomics als auch Cultural Analytics ist also, dass sie mit ihrer Forschung bei den „guten Daten" ansetzen. Beim Verfahren der Culturomics werden Daten aus einer großen Sammlung historischer Bücher generiert; das entspricht den beteiligten Forscher*innen zufolge einem Studium von Millionen von Büchern oder etwa „4 % of all books ever printed" (Michel et al. 2011: 176). Für die Cultural Analytics ist das bevorzugte Korpus das Gesamtwerk eines Künstlers (wie etwa Mark Rothko) oder der komplette Satz von Titelseiten einer Zeitschrift (wie *Time*). In diesen Fällen gelten die Daten als „gut", weil sie von Anfang an – d. h. unabhängig von ihrer digitalen Form – existieren, lange Zeiträume abdecken und mehr oder weniger vollständig sind. Man kennt für jeden Datensatz den Prozentsatz der fehlenden Daten. Im Web hingegen stammen viele Daten aus der jüngeren Vergangenheit, decken nur einen kurzen Zeitraum ab und sind unvollständig – wobei es meist schwierig ist zu erfassen, was vollständige Daten wären.

Aus den Computational Social Sciences lassen sich ebenfalls zwei Verfahren zum Vergleich mit digitalen Methoden heranziehen: Webometrics und Altmetrics. Beide sind szientometrische oder bibliometrische Ansätze zur Untersuchung von Reputation oder Impact, angewandt auf Webdaten. Als solche übertragen sie die Zitationsanalyse auf das Web, wenn auch auf unterschiedliche Weise. Webometrics untersucht Hyperlinks (ähnlich wie der Issuecrawler, siehe Rogers, 2019: 43-58) und leitet die Reputation oder den Einfluss einer Website aus der Quantität und Qualität der erhaltenen Links ab. Dieses Verfahren verwendet nativ-digitale Objekte (Hyperlinks) und digitalisierte Methoden (Bibliometrie). Altmetrics ist ähnlich und verwendet Social-Media-Metriken (nativ-digitale Aktivitäten wie Retweets), um einem veröffentlichten akademischen Artikel einen „Aufmerksamkeits-Score" zuzuweisen (digitalisierte Methode). Der Score verändert sich in Abhängigkeit von der Menge der Erwähnungen in den Online-Quellen. Erwähnungen in den Nachrichten und in Blog-Beiträgen beispielsweise wiegen mehr als auf Reddit.

Die hierbei anklingenden Schwierigkeiten, Methoden ins Digitale zu verlagern, um Daten online zu erheben, sind Gegenstand eines weiteren sozialwissenschaftlichen Ansatzes jenseits des Computational Turn, der virtuelle Methoden (Hine 2005) genannt wird. Während digitale Methoden versuchen, die Methoden des Mediums zu nutzen, verlagern virtuelle Methoden das sozialwissenschaftliche Instrumentarium ins Web. Ein

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Beispiel dafür sind Online-Umfragen. Die Verlagerung der Methoden ins Internet ist dabei nicht immer einfach möglich. Die Online-, Netz- oder virtuelle Ethnographie ist in der Lage, im Internet Gemeinschaften zu finden, ihnen beizutreten, sie zu beobachten und an ihnen zu partizipieren (Hine 2005). Für andere Techniken suchen virtuelle Methoden noch nach einem Umgang mit einigen Besonderheiten des Web als Untersuchungs- und Datenerhebungsort. Bei Befragungen etwa stellt sich die Frage, wie man die Befragten findet und wie man die Rücklaufquote bestimmt. Hinzu kommt die Herausforderung, beim Sampling die Population von Websites oder Facebook-Seiten zu einem spezifischen Thema zu bestimmen. Die Verlagerung einer Methode ins Internet wirft also Fragen nach der Passung zwischen der Methode und dem Medium auf.

Digitale Methoden hingegen nutzen nicht nur nativ-digitale Daten, sondern auch die Methoden, die für das Medium selbst „nativ" sind. „Nativ" ist hier nicht im ethnographischen oder anthropologischen Sinne gemeint. Vielmehr wird es im Sinne der Informatik verwendet, wo der Begriff den Code bezeichnet, der für einen bestimmten Prozessor oder ein bestimmtes Betriebssystem geschrieben wurde und nicht als Simulation oder Emulation läuft. „Nativ" ist also das, was für das Online-Medium geschrieben wurde, und nicht das, was dorthin migriert wurde.

Kürzlich wurde ein dritter Typus von digitalen Objekten vorgeschlagen, jenseits des Nativ-Digitalen und des Digitalisierten: das wiedergeborene digitale Objekt als das, was einmal im Medium geboren, archiviert und als archiviertes Objekt in einer digitalen Bibliothek „wiedergeboren" wurde (Brügger 2012). Die Untersuchung von Webarchiven ermöglichte also die Untersuchung nicht nur der nativ-digitalen Materialien, sondern auch der Effekte der Archivierung sowie des Archivs als Institution oder Regime. So beginnt zum Beispiel die Twitter-Sammlung der Library of Congress – wenn sie denn schließlich Forscher*innen zugänglich gemacht wird – mit Jack Dorseys erstem Tweet im März 2006, weist aber gewisse Lücken auf (Nutzerprofile wurden erst ab September 2011 gesammelt), und Forscher*innen werden berücksichtigen müssen, dass sich die Nutzungsbedingungen von Twitter mehrmals geändert haben (Osterberg 2013). Es gibt außerdem bestimmte Twitter-Richtlinien zum Umgang mit den Interessen der Nutzer*innen, die das Archiv befolgen müsste (z. B. keinen Zugang zu gelöschten oder gesperrten Tweets zu gewähren, selbst wenn diese Daten verfügbar sind). Und schließlich ist die Vollständigkeit der Sammlung endlich, denn seit 2018 legt die Library of Congress nur noch spezielle Tweet-Sammlungen an (zur Erstellung von Tweet-Sammlungen für unterschiedliche Forschungszwecke siehe Rogers 2019: 153–178).

Digitale Methoden folgen einer allgemeinen Forschungsstrategie oder einer Reihe von Schritten, die gewisse Affinitäten zu einem Online-Softwareprojekt, Mash-up oder einer Kettenmethodik haben. Zunächst wird eine Bestandsaufnahme der verfügbaren digitalen Objekte gemacht (Hyperlinks, Tags, Retweets, verkürzte URLs, Wikipedia-Edits, anonyme Benutzer-IP-Adressen, Zeitstempel, Likes, Shares, Kommentare etc.). Anschließend wird gefragt, wie die Geräte online mit diesen Objekten umgehen. Wie können wir von Online-Methoden lernen? Hier kommt die Perspektive der Sozialforschung ins Spiel. Wie kann man die Online-Methoden und -Geräte so umfunktionieren, dass man nicht die Online-Kultur oder die virtuelle Gesellschaft untersucht, sondern den Zustand der Kultur und den Wandel der Gesellschaft? An diesem Punkt stellt sich die Frage nach Triangulation und Vergleichsmöglichkeiten. Wie „verankert" man die mit Online-Daten gewonnenen Erkenntnisse? Müssen wir dazu offline gehen, dürfen wir Online- und Offline-Daten und -Methoden kombinieren, oder können die Erkenntnisse im Online selbst „verankert" werden?

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

4. Digitale Methoden als Forschungspraxis

Wie können bestimmte Geräte oder Plattformen (z. B. Internet Archive, Google Websuche, Wikipedia, Facebook, Twitter und andere) für Zwecke der Sozialforschung nutzbar gemacht werden? Zunächst sollte man sich bewusstmachen, dass digitale Methoden oft experimentell und situativ sind, da sie sich mit den Bedingungen des Mediums entwickeln und gelegentlich von einem Gerät zum anderen migrieren. Sie können kurzlebig sein, wenn bestimmte Dienste nicht mehr angeboten werden. Sie können Änderungen zum Opfer fallen, die von einer Plattform vorgenommen werden, wenn z. B. ein Dienst eingestellt, die erweiterte Suche in sozialen Medien entfernt oder wenn eine API geschlossen wird. Bei solchen Änderungen kann die Forschung beeinträchtigt sein oder muss vielleicht sogar eingestellt werden –Längsschnittstudien sind hier besonders betroffen. Forscher*innen, die der Qualität von Webdaten skeptisch gegenüberstehen, werden angesichts der Instabilität der Infrastruktur, die diese Daten bereitstellt, zusätzlich misstrauisch (siehe dazu oben Thelwall et al. 2005). Kritik gilt insbesondere kommerziellen Suchmaschinen und Social-Media-Plattformen. Diese können eine legale Nutzung ihrer Suchmaschinenergebnisse oder anderer ausgegebener Daten zu Forschungszwecken verweigern, da Forscher*innen als eine spezifische Nutzungsgruppe nicht Teil ihres Geschäftsmodells sind.

Im Folgenden werden das Internet Archive, die Google Websuche, Wikipedia, Facebook und Twitter im Hinblick auf die Möglichkeiten betrachtet, die sie für eine Sozialforschung mit Hilfe digitaler Methoden eröffnen. Dabei geht es jeweils um die Frage, welche digitalen Objekte zur Verfügung stehen, wie mit ihnen umgegangen wird und wie man von der Methode des Mediums lernen und dieses für die Sozialforschung nutzbar machen kann.

4.1 Internet Archive

Die Schnittstelle des Internet Archive, die Wayback Machine, hat als Haupteingabe eine einzelne URL. Man erhält durch Eingabe einer URL die unter dieser gespeicherten Seiten bis ins Jahr 1996 zurück und kann sie auch als Einzelseiten ausgeben lassen. Eine zunehmend verbreitete Art und Weise, die Wayback Machine für Forschung zu nutzen, ist, ihren Single-Site-Fokus für die Erstellung von Single-Site-Historien zu verwenden. Hierfür wird ein Screenshot der Oberfläche einer Homepage erstellt und gespeichert, in chronologischer Reihenfolge sortiert und im Stil eines Stop-Motion-Films abgespielt. Typischerweise wird bei einem solchen Vorgehen eine Voiceover-Spur hinzugefügt, in der die so zu sehenden Änderungen interpretiert werden: etwa, wie die Geschichte einer einzelnen Website die Geschichte des Webs erzählen kann, der Vergleich von alten und neuen Medien (wie die Geschichte einer Online-Zeitung) oder die Geschichte einer Institution (wie whitehouse.gov). Eine Single-Site-Geschichte auf diese Weise als Film darzustellen, greift das etablierte Format der Screencast-Dokumentation auf, das insbesondere der Screencast „Heavy Metal Umlaut“ zur Historie des gleichnamigen Wikipedia-Artikels entwickelte, der in gewisser Weise die Geschichte der Editierkultur von Wikipedia erzählt (Udell 2005). Der Wandel einer Plattform und des Crowdsourcing wurde anhand der Geschichte eines einzelnen Wikipedia-Artikels greifbar gemacht. Das erste Beispiel für eine Screencast-Dokumentation einer einzelnen Seite, die aus Screenshots der Wayback Machine des Internet-Archivs erstellt wurde, ist „Google and the Politics of Tabs“ (Rogers & Govcom.org 2008). Durch die Untersuchung der Veränderungen bei den Suchmöglichkeiten, die auf google.com im Laufe der Zeit bevorzugt (und auch wieder verworfen) wurden, wird die Geschichte des Niedergangs der menschlichen

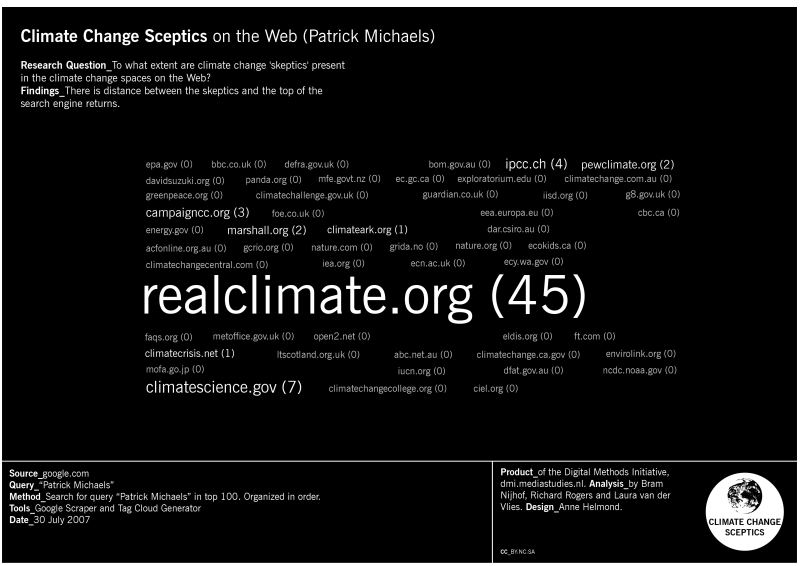
https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Redakteur*innen des Webs (und der Webverzeichnisse) erzählt und die des Aufstiegs des Algorithmus sowie der Funktionsübernahme des Backends von den Bibliothekaren.

4.2 Google Websuche

Die Google-Websuche ist so vertraut, dass es einiger Distanz bedarf, um ihr Potenzial als sozialwissenschaftliches Forschungswerkzeug jenseits ihres Werts als alltägliches Mittel der Informationssuche zu erkennen. Google verarbeitet digitale Objekte wie Hyperlinks, Klicks und Datumsstempel (zum Erfassen von Aktualität). Es ist eine Ranking- und auch Statusherstellungsmaschine für Quellen pro Stichwort, basierend auf algorithmischen Konstruktionen von Relevanz. Die Relevanz stützt sich zunehmend auf die Klicks der Nutzer und die Aktualität der Seite, statt wie bisher auf die Verlinkung der Seiten. So kann man die Ergebnisse der Suchanfrage „Klimawandel" als eine Liste von Websites, hauptsächlich von Organisationen, interpretieren, die nach Relevanz geordnet sind. Sobald man eine solche Liste der „Top-Quellen" für den Klimawandel hat, kann man jede Quelle nach den Namen von Klimawandelskeptiker*innen abfragen und erfassen, an welcher Position der Suchmaschinenergebnisse diese erscheinen (und mit welcher Häufigkeit). Diese zweistufige Methode hat den Namen „Quellen-Distanz-Messung" („Source Distance"), da sie versucht, die Distanz eines bestimmten Eintrags von der Spitze einer Rankingliste oder eines Unterthemas in einem größeren Themenraum zu erfassen (Rogers 2013b: 32–33, 112–118). Dieses Vorgehen ist das Web-Äquivalent der Untersuchung der Positionierung klassischer Nachrichten (siehe Abbildung 3).

Abbildung 3: Beispiel für eine Quellenwolke



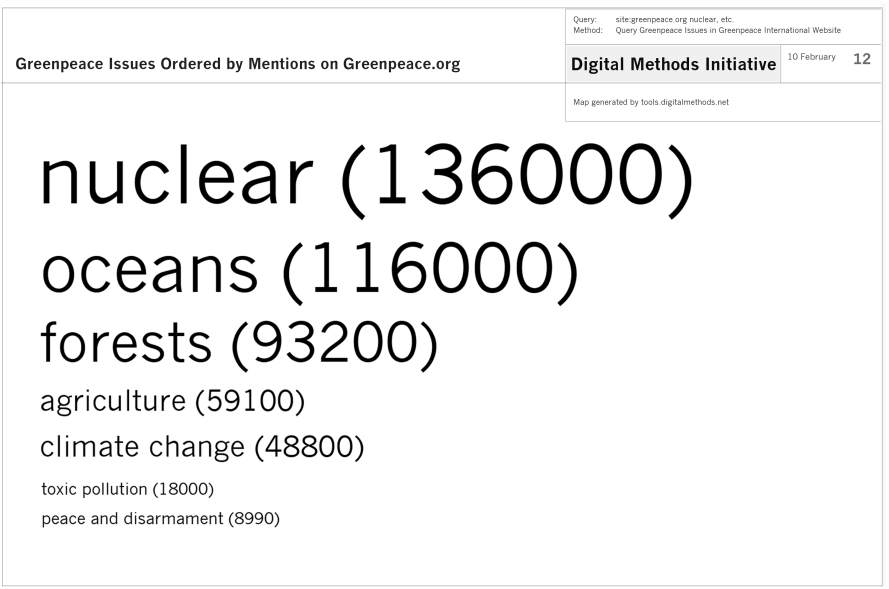
Vorkommen von Klimawandelskeptiker*innen in den Top-Ergebnissen für die Google-Suchanfrage „climate change", Juli 2007. Ausgabe durch das Lippmannian Device, Digital Methods Initiative, Amsterdam.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Google ist nicht nur eine Ranking-Maschine, sondern auch eine massive Indizierungsmaschine, was für die Nutzer*innen bedeutet, dass die Inhalte von Websites nicht nur für einzelne Begriffe, sondern auch für mehrere Begriffe abgefragt werden können, um ein Verständnis dafür zu bekommen, welche Wörter häufiger vorkommen als andere. Man kann eine solche Einzel-Seiten-Indexierung nutzen, um die Anliegen einer Organisation zu untersuchen. Wie sich dies methodisch realisieren lässt, zeigt sich anhand des Beispiels Greenpeace.org. Man startet hier einzelne Suchabfragen zu allen auf der Webseite selbst aufgeführten Kampagnen, um anhand der Anzahl der Erwähnungen zu erfassen, welche eine größere interne Resonanz haben als andere (siehe Abbildung 4). In Anbetracht der Aktualität von Googles Suchergebnissen, würde eine solche Abfrage den aktuellen Stand einer solchen internen Resonanz liefern. Über eine zusätzliche Abfrage von Datumsangaben lassen sich Daten zur Veränderung der Interessen im Zeitverlauf generieren, also dazu, was man „Engagement" nennen könnte.

Man kann auch mehrere Websites nach einzelnen Begriffen oder nach unterschiedlichen Begriffen abfragen. Zum Beispiel könnte man Menschenrechts-Websites nach verschiedenen Begriffen abfragen – wie Kampagnen und Unterthemen –, um ein Verständnis für die Bedeutung jedes Begriffs in der Bandbreite der Organisationen zu bekommen. So wäre es möglich, etwa die Agenden des globalen Human Rights Network zu erforschen. Genau das ist der Zweck des sog. Lippmannian Device (ein weiterer Anwendungsfall des Google Scrapers, der oben beschrieben wurde): Die Software erlaubt es den Benutzer*innen, Quellenwolken (welche Quellen erwähnen welche Themen, siehe Abbildung 3) und Themenwolken (welche Themen werden von den gegebenen Quellen erwähnt, siehe Abbildung 4) zu generieren.

Abbildung 4: Beispiel für eine Themenwolke



Greenpeace-Kampagnen, die auf der Website von Greenpeace.org erwähnt werden, Februar 2012. Ausgabe durch das Lippmannian Device, Digital Methods Initiative, Amsterdam.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

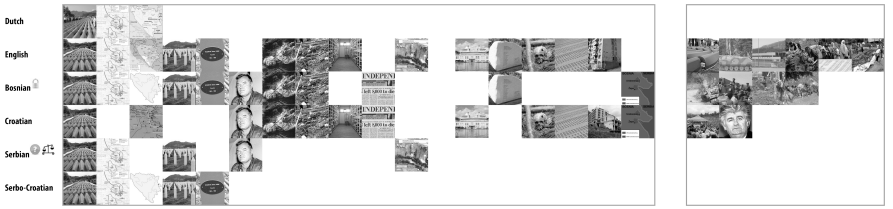
4.3 Wikipedia

Die Online-Enzyklopädie Wikipedia basiert auf einer Reihe von Prinzipien, die ihre Editor*innen befolgen, damit die Artikel „enzyklopädisch" werden und bleiben, nämlich: neutraler Standpunkt, keine Originalrecherche und Überprüfbarkeit der Quellen. Einträge von Wikipedia werden standardmäßig bei inhaltlichen Suchanfragen (in Abgrenzung zu Navigations- und Transaktionsanfragen) unter den Top-Ergebnissen von Google aufgeführt, wodurch die Online-Enzyklopädie für Nutzer*innen eine sehr sichtbare Referenzquelle darstellt. Wie würden Forscher*innen, die digitale Methoden anwenden, hier vorgehen? Betrachtet man die Affordanzen von Wikipedia, so fallen die verschiedenen Sprachversionen auf, wobei jeder Artikel Links zu den anderen Sprachversionen hat („Interwiki-Links"), was Forscher*innen die Sammlung von Artikeln zu einem Thema in verschiedenen Sprachversionen erleichtert. Wenn die Artikel nicht (kürzlich) übersetzt wurden, dann ermöglichen sie eine transkulturelle (oder translinguale) vergleichende Analyse. Was kann verglichen werden? Zu jedem Artikel gibt es eine Reihe von digitalen Objekten wie z. B. anonyme Bearbeitungen mit der IP-Adresse der jeweiligen Bearbeiter*innen, deren Standorte nachgeschlagen werden können, um die Orte der Bearbeitungen zu erfassen. Es gibt auch eine Revisionsgeschichte und eine Diskussionsgeschichte (Talk-Seiten), worüber man die Intensität der Bearbeitungen sowie der Debatten ermitteln kann. Außerdem gibt es den Titel des Artikels, seine Editor*innen (einschließlich Bots), das Inhaltsverzeichnis, Bilder und Referenzen. Alles kann miteinander verglichen werden.

Projekte wie Manypedia und Omnipedia verfügen über automatisierte Mittel zum Vergleich von Wikipedia-Artikeln in verschiedenen Sprachversionen, was in ersterem als LPOV („language points of view") bezeichnet wird. Statt eines Nachschlagewerks wird Wikipedia damit zur Quelle für die Untersuchung kultureller Referenzsysteme oder sogar nationalkultureller Standpunkte. Ein Beispiel dafür ist das „Massaker" von Srebrenica, wie es in der serbischen Version betitelt ist, der „Völkermord" von Srebrenica (bosnisch) und der „Fall" von Srebrenica (niederländisch). Ein Vergleich der Sprachversionen der drei wichtigsten an den Ereignissen vom Juli 1995 beteiligten Parteien zeigt auch unterschiedliche inhaltliche Angaben: So ist beispielsweise von etwa 6000-8000 (serbische Version), 8000 (bosnische Version) oder 7000-8000 (niederländische Version) getöteten Bosnier*innen die Rede (Rogers & Sendijarevic 2012). Der bosnische Eintrag weist daneben deutlich andere Bilder auf, darunter eines vom Grab eines 13-jährigen Jungen, das, da er nicht im kampffähigen Alter war, ein Beweis für einen Genozid sei (siehe Abbildung 5). Die niederländische Version betont die militärische Seite der Ereignisse, und die serbische – nur in letzterer Hinsicht der niederländischen ähnlich – ist die einzige, die einen Abschnitt über die Ereignisse aus Sicht der Republika Srpska enthält, dem Teil von Bosnien und Herzegowina, der unter serbischer Verwaltung steht, in dem sich die Stadt Srebrenica befindet. Die Artikel haben auch keine gemeinsamen Referenzen oder Editor*innen. Die Unterschiede zwischen den Artikeln, von den Unterschieden in den Orten der Bearbeitungen sowie den Aktivitäten der Redakteur*innen ganz zu schweigen, bieten Material für die Untersuchung des kulturellen Gedächtnisses sowie von Kontroversen. Wissenschaftler*innen regen daher dazu an, eigene Artikel zu erstellen, statt auf Übersetzungen von der englischsprachigen Wikipedia zurückzugreifen (Callahan & Herring 2011).

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access -

Abbildung 5: Wikipedia als Quelle zur Erforschung kultureller Sichtweisen



Vergleich der Bilder der Artikel zu „Srebrenica" in der niederländischen, englischen, bosnischen, kroatischen, serbischen und serbokroatischen Wikipedia-Sprachversion, 20.12.2010. Ausgabe durch das Cross-Lingual Image Analysis Tool, Digital Methods Initiative, Amsterdam.

4.4 Facebook

Die am häufigsten untersuchten digitalen Objekte auf Facebook waren lange Verbindungen (Freunde) und Vorlieben (Profile) (Lewis et al. 2008a, Lewis et al. 2008b). Über die API von Facebook und mit Hilfe der Software Netvizz konnte man zum Beispiel Forschung zu egozentrierten Netzwerken betreiben, indem man die verfügbaren Daten einer Person und ihrer Freunde nutzte. Als digitale Objekte bieten die Profile von Facebook die Möglichkeit, etwas zu untersuchen, was ich als Postdemografie („postdemographics") bezeichne: die Medienpräferenzen und den Geschmack von Gruppen von Social-Media-Nutzer*innen. In einer experimentellen Arbeit mittels der erweiterten Suche von MySpace wurden Kompatibilitätsvergleiche zwischen den Interessen der Freunde von John McCain und Barack Obama angestellt, und zwar vor den US-Präsidentschaftswahlen 2008, bei denen sich die beiden gegenüberstanden (siehe Abbildung 6). Eine solche Forschung nutzt die Profile, um den sog. Kulturkampf zu untersuchen und das Ausmaß der Polarisierung zwischen „roten" (republikanischen) und „blauen" (demokratischen) Anhänger*innen zu erfassen.

Nach dem „Ethik-Turn" in der Social-Media-Forschung, wohl ausgelöst durch die De-Anonymisierung von Student*innen des Harvard College, die Gegenstand der oben diskutierten Geschmacks- und Bindungsforschung waren, durch Michael Zimmer (2010), änderte Facebook seine API und untersagte die Erforschung von Freunden und deren Profilen (Rieder 2015). Über die API waren danach nur noch Facebook-Seiten (und -Gruppen) für Forscher*innen zugänglich (später wurden auch geschlossene Gruppen von der Datenanalyse ausgeschlossen, selbst wenn man ihnen beigetreten war). Damit wurden auf Facebook die Seite und die offene Gruppe die für die Analyse operationalisierbaren digitalen Objekte sowie das, was Nutzer*innen dort tun können: liken, reagieren, teilen und kommentieren; auch können Seiten andere Seiten liken. Mit einer der digitalen Methoden, die mit mit Netvizz entwickelt und gebaut wurden, kann ein „Netzwerk von gelikten Seiten" („inter-liked page network") mit einem oder zwei Verlinkungsschritten („degrees of separation") erzeugt werden. Ein solches Netzwerk kann dann mit dem Netzwerk-Story-Telling-Ansatz und/oder einer Auswertung der „beliebtesten Inhalte" („most engaged-with content") analysiert werden (siehe hierzu Rogers 2019).

Forscher*innen können zudem eine Reihe von Seiten zusammenstellen, die sich auf ein bestimmtes Thema, z. B. Rechtsextremismus, beziehen, um die Inhalte zu erfassen, auf die Nutzer*innen am häufigsten reagieren. Bei Facebook hat man Zugriff auf den Datensatz zum „Beitrags-Engagement" pro Seite, auch im Längsschnitt. Man kann er-

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Abbildung 6: Aggregierte Profile der Interessen der Top-100-„Freunde" von Barack Obama und John McCain, MySpace.com, September 2008

A MYSPACE RELATED SERVICE!

elfriendo BETA

FAQ | ABOUT

PROFILE BASED ON OBAMA		COMPATIBILITY	PROFILE BASED ON MCCAIN	
General	, barack obama, reading, music, writing, history, politics, movies, traveling, friends, bob marley, peace, family guy, chicago, books, democrats, photography, harry potter, as a u, running, the producers, jon stewart, the daily show, art, senator, macs, napoleon dynamite, conan o'brien, sleeping, guinness, italy, the beatles, psychology, cats, dancing, shopping, summer, poetry, democracy, voting, gay rights, obama, mythology, coffee, borat, concerts, george orwell, david sedaris, icons, buddhism, ingrid chavez, pink floyd, piano	17%	General	yes, sometimes, edwin mccain, friends, reading, music, swimming, concerts, piercings, shopping, com, movies, metal, get paid to take surveys!, create your own!, (8), death metal, art, what?, soemtimes, laguna beach, jj mccain, john mccain, the beach, fugazi, freya, from autumn to ashes, full blown chaos, from first to last
Music	radiohead, belle and sebastian, the beatles, the roots, pearl jam, sufjan stevens, bob dylan, swervedriver, the smiths, cat power, nina simone, amy winehouse	0%	Music	the shins, com, the bravery, coldplay, evanescence, guster, blink 182, dido, green day, taking back sunday, the used
Movies	little miss sunshine, american beauty, goodfellas, alice in wonderland, amelie, a clockwork orange, archangel, secretary, memento, magnolia, lost in translation, rushmore	0%	Movies	love actually, old school, garden state, kill bill, pirates of the caribbean, napoleon dynamite, shawshank redemption, big fish, princess bride
Television	the office, arrested development, weeds, the daily show, lost, heroes	16%	Television	family guy, project runway, top chef, america's next top model, csi, desperate housewives, lost
Books	atlas shrugged, books, alice in wonderland, gone with the wind, harry potter, 1984, america: the book, catch-22, josef mengele, gravity's rainbow, the baroque cycle (quicksilver, being written before obama decied on a political career it offers an honest introspective look that few other politicians could ever have offered:	0%	Books	to kill a mockingbird, com, me talk pretty one day, "sneaking into the flying circus, it's just enough to get the blood pumping!
Heroes	my mom, , johnny cash	0%	Heroes	haha, kat von d, john lennon, barry goldwater

Analyse und Ausgabe durch Elfriendo.com, Govcom.org Foundation und Digital Methods, Initiative, Amsterdam.

fassen, welche Inhalte (und welche Inhaltstypen) eine Reaktion („engagement") hervorgerufen haben (und welche Arten von Reaktion). Was spricht jemanden an „Stop Islamization of the World" an, um auf „Gefällt mir", „Kommentieren" und „Teilen" zu klicken (siehe Abbildung 7)?

Man kann auch einer Facebook-Gruppe beitreten und dadurch Zugang zu den Daten erhalten. Forscher*innen sollten dann ein „Forschungsprofil" erstellen, durch das sie sich als solche zu erkennen geben und in Form eines Facebook-Posts, eines Blogeintrags oder einer Webseite auf das jeweilige Projekt hinweisen, um so Thema, Forschungsfragen, Methoden und erwartete Ergebnisse transparent zu machen. Die Erforschung von sensiblen oder im Verborgenen stattfindenden Aktivitäten kann allerdings andere Arten des Vorgehens erfordern. Ein bemerkenswertes Beispiel ist die Idee, Anzeigen in rechtsgerichteten Gruppen zu schalten und deren Mitglieder einzuladen, an einem Projekt teilzunehmen, um so sowohl transparent zu sein als auch eine Einverständniserklärung für die Forschung zu erhalten (siehe Bartlett et al. 2011). Umgekehrt kann eine Rechtfertigung für Online-Beobachtungen ohne Selbstidentifizierung in ihrem Wert für das öffentliche Interesse bestehen.

4.5 Twitter

Bei der anfänglichen Untersuchung von Tweets wurden diese als „belanglos" oder „mit Weiterleitungswert" kategorisiert, was schließlich auch von den Twitter-Nutzer*innen als RT (Retweets) kodifiziert wurde, also als jene Tweets, die so interessant sind, dass sie erneut getwittert wurden (Rogers 2013a). Zu den Retweets gesellten sich weitere digitale Objekte, die von den Nutzer*innen gestaltet wurden, insbesondere der Hashtag, der Inhalte nach Themen gruppiert, etwa zu einer Veranstaltung. Für die Forschung be-

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Abbildung 7: Reaktionen auf bestimmte Inhalte („most engaged-with“)

Nach Anzahl der Likes, Kommentare und Shares, auf der Facebook-Seite, Stop Islamization of the World, Januar 2013.

inhalten per Hashtag retweetete Tweets die Möglichkeit, bedeutende Tweets des Tages zu erfassen, wie z. B. die zu den Iran-Wahlen und ihren Nachwirkungen im Juni 2009.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Wie kann ein Stream solcher Daten nutzbar gemacht werden? Ein Ansatz der digitalen Methoden bestand darin, die zeitlich rückwärtsgewandte Reihenfolge von Twitter umzukehren und die wichtigsten Retweets pro Hashtag in chronologischer Reihenfolge zu platzieren, um so die Geschichte eines Twitter-Ereignisses darzustellen (siehe Abbildung 8). Eine Kernfrage blieb allerdings das Verhältnis zwischen dem, was vor Ort, und was in den sozialen Medien passiert. Auf diese Problematik hatte Evgeny Morozov in Bezug auf die Wahl im Iran 2009 hingewiesen. Dabei zitierte er den Leiter der neuen Medien von Al Jazeera mit der Aussage, dass während der Wahlkrise im Iran sechs Twitter-Nutzer*innen von Teheran aus twitterten – die restlichen Tweets kamen aus anderen Regionen der Welt (Morozov 2012: 15).

Abbildung 8: Die drei wichtigsten Retweets pro Tag mit dem Hashtag #iranelection, 10.-30.062009

In chronologischer Reihenfolge wird die Geschichte der Iran-Wahlkrise von Twitter erzählt. Datensammlung bei rettiwt.net. Digital Methods Initiative, Amsterdam, 2009.

Das Unternehmen Twitter erkannte zunehmend, wie Mitbegründer Jack Dorsey sagte, dass Twitter „gut bei Naturkatastrophen, von Menschen verursachten Katastrophen, Veranstaltungen, Konferenzen, Präsidentschaftswahlen" (Dorsey zitiert nach Sarno 2009) sei. Im Jahr 2009 änderte Twitter deshalb seinen Slogan von „What are you doing?" zu „What's happening?", was eine Transformation von Twitter als Tool zum Folgen von Freunden (in der Hoffnung auf die Erfahrung von unmittelbarer oder mittelbarer Intimität) hin zu einem Nachrichtenmedium zum Folgen von Ereignissen (insbe-

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

sondere Wahlen und Katastrophen) zeigt. Hierdurch wurde Twitter zu einem Datenbestand, der nicht nur einen kommerziellen, sondern auch einen historischen Wert hat, was auch daran deutlich wird, dass die amerikanische Library of Congress Twitter als digitales Archivierungsprojekt sieht. Routinen zum Aufbau von Tweet-Sammlungen und zur Ausgabe als Ereignis-Chronologien für die „Fernanalyse von Ereignissen" gehören als eine spezifische digitale Methode zu den wissenschaftlichen Verwendungsweisen von Twitterdaten (siehe Rogers 2019: 153–178).

Man kann auch eine Tweet-Sammlung zu einem Thema – einen „issue space" — erstellen, wie z. B. zu Weltgesundheit und Entwicklungspolitik oder zu Menschenrechten. Einmal abgegrenzt, kann ein solcher Raum von Akteuren, die Kampagnen, Veranstaltungsankündigungen, Ressourcen, Story-Links und andere Formate zu einem Thema posten, mit einer Vielzahl von auch kritischen Verfahren untersucht werden. Beispiele sind Engagement-Metriken, die versuchen, die dominanten Stimmen, Anliegen, Engagements, Standpunkte und Orientierungen zu erforschen. Sie können in einem doppelten Sinne als kritisch gelten. Erstens bieten sie eine Alternative zu den „Eitelkeitsmetriken" oder reinen Follower-Zahlen, die oft durch Bezahlung erhöht werden, um sich besser zu präsentieren. Zweitens zeigen sie, ob im jeweiligen Twitter-Raum bestimmte Teilnehmer*innen und Themen marginalisiert werden, und untersuchen gleichzeitig kritisch die dominanten Akteure im Hinblick auf Thementrends und andere Formen des Engagements.

4.6 YouTube

Die erste archivierte YouTube-Seite zeigt eine Dating-Plattform mit dem Slogan „broadcast yourself". Schnell wurde YouTube aber mit Amateur-Inhalten oder allgemeiner mit nutzergenerierten Inhalten in Verbindung gebracht, und der Begriff der „partizipativen Kultur" (Jenkins, Ito, & boyd 2016) kam auf. Nachdem zunächst der „cult of the amateur" (Keen 2007) und der Wert von Videos wie „Charlie bit my finger" kritisiert wurde, beklagten später andere die Kommerzialisierung der Plattform, die sich darin zeigte, dass unter den meistgesehenen Videos kommerzielle Inhalte (z. B. Musik) die von den Nutzer*innen erstellten Videos verdrängten. Jüngere YouTube-Studien konzentrieren sich auf die Aktivitäten, die es erfordert, um ein YouTuber oder Influencer zu werden. Solche Dynamiken lassen sich bei einer Vielzahl von Themenfeldern auf YouTube ausmachen, einschließlich extremistischer. Die Verfahren der digitalen Methoden eröffnen hier die Möglichkeit, die Funktionsweise von Empfehlungssystemen und Geräte- oder Ranking-Kulturen kritisch zu analysieren. Ist es der Amateur, das kommerzielle Video oder der Influencer, der als „up next" und in den Suchergebnissen empfohlen wird? Einen Ansatz schwedischer Forscher*innen (Johansson et al., 2019) zur Untersuchung von Spotify aufgreifend lässt sich an dem „Tear Down" von YouTube der Output seiner verschiedenen Empfehlungssysteme kritisch untersuchen und erfassen, wie gerankt wird und wer davon profitiert (Rogers 2019: 249–259).

4.7 Jenseits von Studien zu nur einer Plattform

Studien zu nur einer Plattform sind größtenteils als API-basierte und somit als toolgetriebene Forschung entstanden. Anstelle der Forscher*innen diktiert hier das Social-Media-Unternehmen die verfügbaren Daten und die Bedingungen ihrer Zugänglichkeit. Wissenschaftler*innen, die sich mit Fragen von „Transmedialität"- und „Crossmedialität" befassen, beklagen häufig einen solchen Fokus auf die einzelne Plattform – sowohl konzeptionell als auch empirisch. Dabei wird argumentiert, dass sich die Historie eines

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Ereignisses oder eines Themenraums (einschließlich einzelner Kampagnen und deren Wirksamkeit) typischerweise über mehrere Plattformen entfaltet und Akteure mehr als nur eine Plattform für die Kommunikation ihrer Themen nutzen. Es gibt aber auch hier einen Ansatzpunkt im Rahmen der digitalen Methoden, nämlich Link-Analyse-Software (wie Issuecrawler oder Hyphe), die Einblicke in die Plattformen (und Websites) eröffnet, die für ein Thema relevant sind. Dabei liefern solche Linkanalysen auch Hinweise darauf, welche Plattformen für eine weitergehende Untersuchung hilfreich sein könnten. Jede Plattform kann isoliert untersucht werden, aber plattformübergreifende Analysen profitieren von einem Vergleich der Datenpunkte einer jeden Plattform (z. B. der Likes auf Facebook vs. Twitter, Hashtags auf Twitter vs. Instagram, Weblinks auf der einen vs. der anderen Plattform). Eine solche Analyse lenkt den Blick auf die Unterschiede einzelner Plattformen (z. B. Hashtag-Inflation auf Instagram im Vergleich zu Twitter). Die Durchführung von plattformübergreifenden Analysen ist somit auch die Untersuchung von spezifischen „Plattformnutzungskulturen" oder „Plattformsprachen".

5. Fazit: Das „Nutzbar-Machen" als Kern der digitalen Methoden

Digitale Methoden haben ihren Ausgangspunkt bei der Beobachtung der ontologischen Differenz von Objekten, die im Medium entstehen, und solchen, die in das Medium migriert sind. Diese Beobachtung wird anschließend auf die Methoden ausgeweitet. Es gibt Methoden, die man als „für das Medium nativ" („of the medium") bezeichnen kann, und solche, die auf das Medium übertragen wurden. Auch wenn diese Unterscheidung nicht absolut (oder absolutistisch) ist, so regt sie doch zum Nachdenken darüber an, wie man sich dem Medium zu Forschungszwecken annähert. Anstatt die Flüchtigkeit der Daten und die Instabilität des Mediums zu beklagen und daraus zu schließen, dass Webdaten keine „guten Daten" seien, kann man von den in die Medien „eingebauten" Methoden nämlich auch lernen und sie für die Forschung nutzbar machen.

Dieses „Nutzbar-Machen" („Repurposing") von digitalen Methoden hat Parallelen zu bzw. geteilte Perspektiven mit Ansätzen wie denen des „reverse engineering" und der „unobtrusive measures". Digitale Methoden machen sich die neue Medienpraxis des offenen Webs zunutze.

Beim Reverse Engineering entwickelt man ein Verständnis für ein System (und spezifiziert es, um es zu imitieren oder zu emulieren), „without the original drawings" (Chikofsky & Cross 1990: 14) zu kennen. In ähnlicher Weise lernt man mit den digitalen Methoden – ohne die Innensicht der sprichwörtlichen Black Box zu kennen – vom Medium und setzt dieses Wissen neu ein. Wie empfehlen die Suchmaschinen und Plattformen Inhalte, und was könnte man von solchen Funktionsweisen lernen? Anregungen für ein solches methodisches Lernen kann beispielsweise die Fachpresse bieten (inklusive der Literatur zur Suchmaschinenoptimierung (SEO)) oder man lernt vom Speichern einzelner Suchergebnisse. Bei Plattformen lässt sich bei der Untersuchung ihrer API und (wechselnder) Datenfelder ansetzen. Das „Reversing" ist ebenfalls Teil anderer Verfahren, die zum Einsatz kommen. Ein Beispiel ist die „Reverse-Look-up"-Software bei der Identifikation mehrerer Websites, die mit einer einzigen Google-Analytics- oder AdSense-ID verbunden sind. Die „umgekehrte Bildersuche" ist ein weiteres Beispiel, bei dem rekonstruiert wird, welche Websites ein bestimmtes Bildfaksimile oder eine Annäherung an dasselbe enthalten.

Der Repurposing-Ansatz hat auch eine gewisse Ähnlichkeit mit non-responsiven Methoden und deren Untersuchung von „Residuen". Mit einem solchen Ansatz können Daten auf „nicht-reaktive Weise" erhoben werden, indem etwa Umfragen und Befragungen zugunsten einer Konzentration auf „Spuren" vermieden werden (Webb et al.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

1966). Dies lässt sich auf das Digitale übertragen. Was lässt sich aus den Spuren, die Nutzer*innen online hinterlassen, beobachten und lernen? Zum Beispiel sind „Shares" in den sozialen Medien nicht nur Mittel zur Platzierung von Posts und zum „Boosten" (wie wir von den Methoden der Plattformen lernen), sondern sie sind zugleich auch Indikatoren für die Inhalte, die Gruppen aktivieren. Vielleicht passt der Begriff „Spuren" heute nicht mehr, denn man forscht oft mit protokollierten (statt hinterlassenen) Aktivitäten, also auf der Basis dessen, was man einmal „registrierende Interaktivität" („registrational interactivity") genannt hat.

Digitale Methoden sollten als ein webbasiertes Vorgehen betrachtet werden, „putting things on other things", wie der Netzkünstler Heath Bunting einst die digitalen Medien beschrieb. Auch Webkartographen beschreiben die frühen Mash-Ups (oder „Web Application Hybrids") (Woodruff 2011) als ein solches Vorgehen. Im „Repurposing" klingen zudem andere Re-Wörter an, wie „Remixing" (in Lawrence Lessigs (2004) Verständnis von kreativen Outputs). Praktisch gesprochen macht man eine Bestandsaufnahme einer digitalen Kulturlandschaft anhand der vorliegenden Objekte, man fragt, wie sie von Online-Geräten verarbeitet werden, und überlegt, wie diese Bestandsaufnahme und das Lernen von Geräten für eine produktive Forschung genutzt werden könnten. Hyperlinks verbinden Webseiten, können aber auch Hinweise auf den Reputationswert sein. Retweets sind Gesten des Teilens von Inhalten, können aber auch als Bewertungen von Inhalten betrachtet werden. Und Likes sind sowohl soziale Bindungen als auch Indikatoren für Reaktionen. Identische Google-Analytics- und AdSense-IDs zeigen gemeinsame Besitzverhältnisse an, können aber, wenn sie in einem Mapping erfasst werden, auch Einflussnetzwerke oder Strategien von Mediengruppen sichtbar machen. Insgesamt geht es bei digitalen Methoden also darum, Online-Objekte und -Methoden zu rekombinieren und nutzbar zu machen.

Literatur

Albright, J. (2017). Itemized posts and historical engagement. Six now-closed Facebook pages. *Data set, Tableau Public*, *5*.

Bartlett, J., Birdwell, J., & Littler, M. (2011). The new face of digital populism. https://demosuk.wpengine.com/files/Demos_OSIPOP_Book-web_03.pdf?1320601634 [08.02.2021].

Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, *3*(4), 1–30.

boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878.

Brügger, N. (2012). When the present web is later than the past: Web historiography, digital history and internet studies. *Historical Social Research*, *37*(4), 102–117.

Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the American Society for Information Science and Technology*, *62*(10), 1899–1915.

Campbell, M. T. (2003). Generic names for soft drinks by county. The Pop vs. Soda page. http://www.popvssoda.com/countystats/total-county.html [08.02.2021].

Chen, E. (2012). 'Soda vs. pop', blog posting, Edwin Chen's Blog., 6. Juli 2012. Retrieved from http://blog.echen.me/2012/07/06/soda-vs-pop-with-twitter [08.02.2021].

Chikofsky, E. J., & Cross, J. H. (1990). Reverse engineering and design recovery: A taxonomy. *IEEE Software*, *7*(1), 13–17.

Dean, J. (1998). *Aliens in America: Conspiracy cultures from outerspace to cyberspace*. Ithaca: Cornell University Press.

Dohmen, J. (2007). Opkomst en ondergang van extreemrechtse sites. *NRC Handelsblad*, 25. August 2007.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Ericson, M., & Cox, A. (2009). What's cooking on Thanksgiving. *New York Times*, 26. November 2009.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*, 1012–1014. https://www.nature.com/articles/nature07634?report=reader&fbclid=IwAR2Hi-EJK2FvjytCu6QQzY0vTsL5JN8Dg6pIBGxreI6Oj_Zv6GgdH9n6Uos [08.02.2021].

Hine, C. (Ed.) (2005). *Virtual methods: Issues in social research on the internet*. Oxford: Berg.

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York: New York University Press.

Jenkins, H., Ito, M., & boyd, d. (2016). *Participatory culture in a networked era*. Cambridge, Malden: Polity.

Johansson, A., Eriksson, M., Vonderau, P., Snickars, P., & Fleischer, R. (2019). Spotify teardown. Cambridge, MA: MIT Press.

Keen, A. (2007). *The cult of the amateur*. New York: Doubleday.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., … Gutmann, M. (2009). Life in the network: The coming age of computational social science. *Science*, *323*(5915), 721–723. https://europepmc.org/articles/pmc2745217 [08.02.2021].

Lessig, L. (2004). Free culture. New York: Penguin.

Lewis, K., Kaufman, J., & Christakis, N. (2008a). The taste for privacy: An analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, *14*(1), 79–100.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008b). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, *30*(4), 330–342.

Manovich, L. (2020). *Cultural analytics*. Cambridge MA, London: MIT Press.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Team, T. G. B., … Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176–182.

Moretti, F. (2005). *Graphs, maps, trees: Abstract models for a literary history.* London: Verso.

Morozov, E. (2012). *The net delusion: How not to liberate the world*. Harmondsworth: Penguin.

Osterberg, G. (2013). Update on the Twitter archive at the Library of Congress. *Library of Congress blog*, 26. Dezember 2013. https://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/ [08.02.2021].

Raehsler, L. (2012). What people search for. *Search Engine Watch*, 18. April 2012. https://www.searchenginewatch.com/2013/03/02/what-people-search-for-most-popular-keywords/ [08.02.2021].

Rieder, B. (2012). What is in PageRank? A historical and conceptual investigation of a recursive status index. *Computational Culture*, *2*, September 2012. http://computationalculture.net/what_is_in_pagerank [08.02.2021].

Rieder, B. (2015). The end of Netvizz (?). *The Politics of Systems Blog*, *23 January*. http://thepoliticsofsystems.net/2015/01/the-end-of-netvizz/ [08.02.2021].

Rogers, R. (2009). *The end of the virtual: Digital methods. Inaugural Lecture*. Amsterdam: Amsterdam University Press.

Rogers, R. (2013a). *Debanalizing Twitter: The transformation of an object of study*. Proceedings from Proceedings of WebSci13, New York.

Rogers, R. (2013b). *Digital methods*. Cambridge, MA: MIT Press.

Rogers, R. (2014). Political research in the digital age. *International Public Policy Review*, *8*(1), 73–87.

Rogers, R. (2019). *Doing digital methods*. London: Sage.

Rogers, R., & Govcom.org. (2008). 'Google and the politics of tabs', screencast documentary. https://movies.digitalmethods.net/google.html [08.02.2021].

Rogers, R., & Sendijarevic, E. (2012). *Neutral or national point of view? A comparison of Srebrenica articles across Wikipedia's language versions*. Proceedings from Paper presented at Wikipedia Academy 2012, Berlin.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Sarno, D. (2009). Twitter creator Jack Dorsey illuminates the site's founding document. Part II. *Los Angeles Times*, *19 February*. https://latimesblogs.latimes.com/technology/2009/02/jack-dorsey-on.html [08.02.2021].

Shelton, T. (2011). The (expanded) pop vs. soda debate. *Floating Sheep blog*, 3. October 2011. http://www.floatingsheep.org/2011/10/expanded-pop-vs-soda-debate.html [08.02.2021].

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual review of information science and technology*, *39*(1), 81–135. http://www.researchgate.net/profile/Lennart_Bjoerneborn/publication/220141962_Webometrics/links/552013e80cf2a2d9e14334e1.pdf [08.02.2021].

Udell, J. (2005). 'Heavy metal umlaut', screencast documentary. http://jonudell.net/udell/2005-01-22-heavy-metal-umlaut-the-movie.html [08.02.2021].

US Centers for Disease Control and Prevention (2014). CDC announces winners of the ‚Pedict the influence season challenge', press release, 18. Juni 2014.

Watts, D. J. (2007). A twenty-first century science. *Nature*, *445*, 489-489. https://www.nature.com/articles/445489a [08.02.2021].

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). Unobtrusive measures: Non-reactive research in the social sciences. Skokie, Ill.; Rand McNally & Co.

Woodruff, A. (2011). Web cartography, or putting things on top of other things. *andy-woodruff.com*, *9. Juni*. http://andywoodruff.com/blog/web-cartography-or-putting-things-on-top-of-other-things/ [08.02.2021].

Woolgar, S. (Ed.) (2003). *Virtual society: Technology, cyberbole, reality*. New York: Oxford University Press.

Zimmer, M. (2010). 'But the data is already public': On the ethics of research in Facebook. *Ethics and Information Technology*, *12*(4), 313–325.

https://doi.org/10.5771/1615-634X-2021-1-25 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY