

*Jean Burgess und Axel Bruns*

*ARC Centre of Excellence for Creative Industries and Innovation*

*Queensland University of Technology*

*Brisbane, Australien*

*je.burgess / a.bruns @ qut.edu.au*

*@jeanburgess / @snurb\_dot\_info*

*http://mappingonlinepublics.net/*

Dr. Jean Burgess ist Deputy Director des ARC Centre of Excellence for Creative Industries & Innovation (CCI) und Associate Professor in Digital Media in der Creative Industries Faculty an der Queensland University of Technology. Sie ist an verschiedenen Forschungsprojekten beteiligt, die computerunterstützte Methoden für die Analyse von Social-Media-Daten in großem Stil anwenden. Zu Ihren Büchern gehören *YouTube: Online Video and Participatory Culture* (Polity Press, 2009), *Studying Mobile Media: Cultural Technologies, Mobile Communication, and the iPhone* (Routledge, 2012) und *A Companion to New Media Dynamics* (Wiley-Blackwell, 2013). Über das letzte Jahrzehnt hat sie mit einer großen Zahl von Regierungs-, Industrie-, und Community-Organisationen gearbeitet, mit Schwerpunkt auf den Anwendungen von sozialen und ko-kreativen Medien bei der Vermehrung von Partizipation, Interessenvertretung und gesellschaftlichem Engagement.

Dr. Axel Bruns is ein Associate Professor in der Creative Industries Faculty an der Queensland University of Technology in Brisbane, Australien, und ein Chief Investigator im ARC Centre of Excellence for Creative Industries and Innovation (<http://cci.edu.au/>). Er ist Autor von *Blogs, Wikipedia, Second Life and Beyond: From Production to Producership* (2008) und *Gatewatching: Collaborative Online News Production* (2005), und Mitherausgeber von *Twitter and Society* (2013), *A Companion to New Media Dynamics* (2012) und *Uses of Blogs* (2006). Bruns Expertise liegt in der Forschung zu User-Led Content Creation, oder Producership, und seine aktuelle Arbeit erforscht insbesondere die Nutzerbeteiligung in Social-Media-Räumen wie Twitter. Sein Forschungsblog ist <http://snurb.info/>, und er tweetet als

---

<sup>1</sup> Eine englischsprachige Version dieses Artikels ist in *M/C Journal* 15.5 (2012) erschienen: <http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/561>.

@snurb\_dot\_info. Siehe auch <http://mappingonlinepublics.net/> für weitere Informationen zu seiner aktuellen Social-Media-Forschung.

## **Social Media und ihre Listen**

Listen sind seit langem ein Ordnungsmechanismus für computervermittelte soziale Interaktion. Auch wenn sie bei weitem nicht der erste derartige Mechanismus waren, boten Blogrolls beispielsweise ihren Bloggern die Chance, eine Liste ihrer Kollegen bereitzustellen; die derzeitige Generation von Social-Media-Umgebungen bietet in ähnlicher Weise eine Listenfunktion für Freunden und Follower. Wo Blogrolls und andere frühere Listen jedoch nutzergeneriert waren, sind die Social-Media-Listen von heute eher von den Plattformen selbst erzeugt, und haben daher einen ebenso großen intrinsischen Wert für Plattform-Anbieter wie für die Nutzer selbst; sowohl Facebook als auch Twitter haben zum Beispiel die Bedeutung der jeweiligen „social graphs“ ihrer Netzwerke (ihre Datenbanken von Nutzerverbindungen) als grundlegende Elemente ihrer noch jungen Geschäftsmodelle hervorgehoben. Dieses stellt dar, was Mejas (2012) als einen „nodocentrism“ beschreibt, der „alle menschliche Interaktion mit den Begriffen der Netzwerk-Dynamik beschreibt (nicht nur irgendeines Netzwerks, sondern eines digitalen Netzwerks mit einer profitgetriebenen Infrastruktur)“ (unsere Übersetzung).

Die kommunikative *Inhalte* von Social-Media-Räumen werden ebenfalls häufig in Form von Listen dargestellt. Bekanntlich werden Blogs in erster Linie durch die umgekehrt chronologische Auflistung ihrer Beiträge definiert (Walker Rettberg 2008), doch gilt dies auch für Social-Media-Plattformen: Twitter, Facebook und andere Social-Media-Plattformen sind von Natur aus um eine endlose, ständig aktualisierte und erweiterte Liste der von den einzelnen Benutzern und ihren Verbindungen hergestellten Beiträge aufgebaut.

Die Idee der Liste bedeutet ein gewisses Maß an Ordnung, und die geordnete Abrufbarkeit von Inhalten, wie sie von der neuesten Generation zentralisiert aufgebauter Social-Media-Plattformen bereitgestellt werden, hat auch zur Entwicklung von umfassenderen und leistungsfähigeren, kommerziellen wie wissenschaftlichen, Ansätzen zur Erforschung sozialer Medien geführt. Am Beispiel von Twitter beschreibt dieser Artikel daher die Herausforderungen solcher „big data“-Forschung, die auf die Inhalte der durch kommerzielle Social-Media-Plattformen bereitgestellten Listen aufbaut.

## **Twitter-Archive in der Forschung**

Twitter ist eine besonders nützliche Quelle für Social-Media-Daten: mit dem Twitter-API (dem Application Programming Interface, das einen strukturierten Zugang zu Kommunikationsdaten in standardisierten Formaten bietet) ist es Forschern möglich, mit ein

wenig Mühe und ausreichenden technische Ressourcen sehr große Archive öffentlich verbreiteter Tweets zu bestimmten Themen, Interessenbereichen, oder Veranstaltungen aufzubauen. Grundsätzlich liefert das API sehr langen Listen von Hunderten, Tausenden oder Millionen von Tweets und den Metadaten zu diesen Tweets; diese Daten können dann auf verschiedenlichste Weise extrahiert, kombiniert, und visualisiert werden, um die Dynamik der Social-Media-Kommunikation zu verstehen.

Diese Forschung ist häufig um althergebrachte Fragestellungen herum aufgebaut, wird aber in der Regel in einem bislang unbekannt großen *Maßstab* durchgeführt. Die Projekte von Medien- und Kommunikationswissenschaftlern wie Papacharissi und de Fatima Oliveira (2012), Wood und Baughman (2012) oder Lotan *et al.* (2011) – um nur eine Handvoll der letzten Beispiele zu nennen – sind grundlegend auf Twitterdatensätze aufgebaut, die jetzt routinemäßig Millionen von Tweets und zugehörigen Metadaten umfassen, erfaßt nach einer Vielzahl von Kriterien. Was allen diesen Fällen gemein ist, ist jedoch die Notwendigkeit, neue methodische Wege in der Verarbeitung und Analyse derart großer Datensätze zur medienvermittelten sozialen Interaktion zu gehen.

Unsere eigene Arbeit befaßt sich generell mit dem Verständnis der Rolle sozialer Medien in den zeitgenössischen Medienökologie, mit einem Schwerpunkt auf der Beobachtung der Entwicklung und Dynamik interessen- und themenbedingter Öffentlichkeiten. Wir haben zu diesem Zweck große Archive von Twitterdaten erfaßt analysiert, um moderne Krisenkommunikation (Bruns *et al.* 2012), die Rolle sozialer Medien bei Wahlen (Burgess und Bruns 2012), und Formen aktueller Publikumsbeteiligung an Fernseh- und Nachrichtenmedien zu untersuchen (Harrington, Highfield und Bruns 2012).

Mit Hilfe einer modifizierten Version des Open-Source-Twitter-Archivierungstools *yourTwapkeeper* sind wir in der Lage, alle verfügbaren Tweets (und die zugehörigen Metadaten) zu erfassen und zu archivieren, die einen bestimmten Suchbegriff (z.B. „Bundesliga“ oder „Dubstep“), Namen (Guardiola, Bieber, Obama) oder Hashtag (#btw13, #royalwedding, #qldfloods) beinhalten. In ihrer einfachsten Form sind solche Archive Twitter gemeinhin als Text-Dateien (z.B. Komma- oder Tab-getrennt) gespeichert, mit jedem der folgenden Werte in einer eigenen Spalte:

<b>text:</b>	Inhalt des Tweets selbst, in 140 Zeichen oder weniger
<b>to_user_id:</b>	numerische ID des Tweetempfängers (für @replies)
<b>from_user:</b>	Twittername des Tweetabsenders
<b>id:</b>	numerische ID des Tweets selbst

<b>from_user_id:</b>	numerische ID des Tweetabsenders
<b>iso_language_code:</b>	Code (z.B. en, de, fr, ...) der vom Absender benutzten Sprache
<b>source:</b>	vom Absender benutztes Twittertool (z.B. Web, Tweetdeck, ...)
<b>profile_image_url:</b>	URL für das Profilbild des Tweetabsenders
<b>geo_type:</b>	Format der geographischen Koordinaten des Tweetabsenders
<b>geo_coordinates_0:</b>	erstes Element der geographischen Koordinaten
<b>geo_coordinates_1:</b>	zweites Element der geographischen Koordinaten
<b>created_at:</b>	Sendezeit des Tweets in direkt lesbarem Format
<b>time:</b>	Sendezeit des Tweets als Unix Timestamp

Um diese Daten zu verarbeiten, benutzen wir in der Regel eine Reihe unserer eigenen Skripts (geschrieben in der Programmiersprache Gawk), die den Datensatz auf verschiedene Weise bearbeiten oder filtern, und extrahieren eine Reihe zeitlich, qualitativ oder inhaltlich aufschlußreicher Metriken aus den Daten, sodaß es uns möglich wird, Aktivitätsmuster über die Zeit zu erkennen, sowie Inhalte und Themen, wichtige Akteure und die Beziehungen zwischen ihnen zu identifizieren; in einigen Fällen folgen weitere Phasen der Filterung und detaillierten inhaltlichen Analyse von Tweetinhalten. Netzwerkanalyse (der Beziehungen zwischen den Akteuren in einer Diskussion, oder zwischen wichtigen Themenbereichen) wird mit der Open-Source-Anwendung Gephi vorgenommen. Während eine ausführliche methodische Diskussion den Rahmen dieses Artikels sprengen würde, bietet unsere umfassende Projektwebseite Mapping Online Publics ausführliche Details und Beispiele zu unseren Methoden sowie Hilfsmittel zur Datenanalyse und -visualisierung, einschließlich unserer Gawk-Skripts.

In diesem Artikel beziehen wir uns stattdessen auf die technischen, epistemologischen und politischen Herausforderungen solcher Nutzung von großen Twitterarchiven in der medien- und kommunikationswissenschaftlichen Forschung; wir siedeln diese Diskussion im Kontext des von Lev Manovich als „big social data“ beschriebenen Phänomens an. Dabei erkennen wir an, daß unsere empirische Arbeit über Twitter sich auf ein komplexes Forschungsobjekt bezieht, welches selbst geprägt ist durch eine komplexe Reihe menschlicher und nichtmenschlicher Akteure, die in einer dynamischen, ja unsteten Medienökologie (Fuller 2005) operieren, und daß wir Methoden für die Erfassung und Analyse von Daten verwenden, die selbst tief in diese Ökologie eingebettet sind.

## „Big Social Data“

Wie Manovichs Begriff schon aufzeigt, hat die Big-Data-Idee nun vor kurzem auch die Medien- Kommunikations- und Kulturwissenschaften erreicht – deutlich später als die Naturwissenschaften, die traditionell stärker computergestützten Zweige der Sozialwissenschaften, und vielleicht sogar als die erste Welle der „Digital Humanities“ (welche computergestützte Methoden weitgehend auf bereits bestehende, historische „Big Data“-Sammlungen angewendet haben) –, und diese neue Entwicklung wurde zu einem großen Teil durch das dramatische quantitative Wachstum und die scheinbar zunehmende kulturelle Bedeutung von sozialen Medien hervorgerufen: daher sprechen wir nun von „big social data“. Wie Manovich es ausdrückt:

Zum ersten Mal können [den] Vorstellungen, Meinungen, Ideen und Gefühlen hunderter Millionen Menschen folgen. Wir können die Bilder und Videos, die sie erstellen und kommentieren, sehen, die Gespräche, an denen sie beteiligt sind, mitverfolgen, ihre Blogposts und Tweets lesen, ihre Karten navigieren, ihre Musiklisten hören, und ihre Laufbahnen durch den physischen Raum nachverfolgen. (Manovich 2012: 461; unsere Übersetzung)

In den Medien-, Kommunikations- und Kulturwissenschaften ist dieser Moment wegen des wachsenden Umfangs der Social-Media-Beteiligung und wegen der textlichen Spuren, die diese Beteiligung hinterläßt, gekommen – er erlaubt es den Forschern, mit digitalen Werkzeugen und Methoden ausgestattet, „soziale und kulturelle Prozesse und Dynamiken auf neue Art und Weise“ zu erforschen (Manovich 2012: 461; unsere Übersetzung). Allerdings, und für unsere Zwecke in diesem Artikel entscheidend, würden viele dieser wissenschaftlichen Möglichkeiten unverwirklicht bleiben, wenn Open APIs für Social-Software-Plattformen (einschließlich von Social Media) nicht weitreichend verfügbar wären. APIs bieten die technischen Spezifikationen, wie eine Software-Anwendung auf eine andere zugreifen soll, wodurch die Einbettung oder Weiterverbreitung von sozialen Inhalten über Websites möglich gemacht wird (sodaß zum Beispiel die Tweets eines Nutzers auch seiner Facebook-Timeline erscheinen können), oder ermöglichen es Drittanbietern, zusätzliche Anwendungen zu entwickeln, die auf Social-Media-Plattformen aufbauen (wie etwa den Twitter-Nutzer-Ranking-Service Klout), geben aber dabei den Plattformbetreibern auch die Macht, mit Hilfe desselben Codes *de facto* auch derartige Nutzung durch Dritte zu regulieren. Und auch wenn die Plattformanbieter nicht unbedingt auch Forschungsanwendungen im Sinn

haben, stehen die durch APIs bereitgestellten Datenzugriffsmöglichkeiten auch für wissenschaftliche Zwecke zur Verfügung.

Wie Manovich aufzeigt (2012: 464), sind bis vor kurzem fast alle wirklichen „Big Data“-Ansätze zur Social-Media-Forschung von Informatikern unternommen worden. Aber als Teil eines umfassenderen „computational turn“ in den digitalen Geisteswissenschaften (Berry 2012), und wegen der erhöhten Verfügbarkeit von Datenerfassungs- und -analysetools für Nicht-Spezialisten sind nun auch Medien-, Kommunikations- und Kulturwissenschaftler in Begriff, den Abstand zu ihren Kollegen zu verkürzen. Viele der neuen, großangelegten Forschungsvorhaben, die die gesellschaftliche Nutzung und Auswirkung von Social Media untersuchen – einschließlich unserer eigenen Arbeit – und die von verschiedenen führenden Medien-, Kommunikations- und Kulturwissenschaftlern rund um die Welt eingeleitet wurden, haben ihre Arbeit damit begonnen, eine Bestandsaufnahme der Palette verfügbarer Werkzeuge und Methoden zur Social-Media-Datenanalyse zu machen sowie diese oft durch neue Entwicklungen wesentlich zu erweitern. Die durch solche Projekte entwickelte Forschungsinfrastruktur spiegelt daher nun ihre eigenen fachlichen Hintergründe mindestens so sehr wieder wie die grundlegenden Prinzipien der Informatik. Im Gegenzug schaffen solche neuen und oft experimentellen Werkzeuge und Methoden notwendigerweise auch neue epistemologische und methodische Herausforderungen.

### **Das Twitter-API und Twitterarchive**

Das Open API war ein wichtiger Aspekt der Vorstellungen über den Wert des offenen Webs und der „Web 2.0“-Geschäftsmodelle in der Mitte der 2000er Jahre (O'Reilly 2005), welche die offene, plattformübergreifende Verbreitung von Inhalten sowie die Förderung von Innovation am Rande des Marktes durch Drittanbieter-Anwendungsentwicklung herausstellten – und es war in diesem ideologischen Umfeld, daß der Microblogging-Dienst Twitter startete und sich eines raschen Wachstum in seiner Popularität unter Nutzern gleich wie unter Entwicklern erfreute. Wie José van Dijck (2011) überzeugend argumentiert, hat jedoch ein komplexes Zusammenspiel von technischen, wirtschaftlichen und sozialen Dynamiken dazu geführt, daß Twitter eine Transition von einer relativ offenen, *ad hoc* entstehenden und nutzerzentrierten Plattform hin zu einer sehr viel mehr formalisierten Medienfirma durchgemacht hat:

Für Twitter hat der Wandel von einem primär konversationellen Kommunikationstool zu einem globalen, anzeigengestützten Followertool in einer relativ kurzen Zeitspanne stattgefunden. Dieser Wandel resultierte nicht einfach aus der Entscheidung des Eigentümer für ein bestimmtes Geschäftsmodell oder aus der Entscheidung des Unternehmens, seine Hardware-Features zu ändern. Stattdessen war die Verbreitung von Twitter als Tool ein komplexer Prozess, in dem technologische Anpassungen unauflöslich mit Veränderungen in der Nutzerschaft, Transformationen im Inhalt, und Entscheidungen für Geschäftsmodelle verflochten sind. (Van Dijck 2011: 343; unsere Übersetzung)

Die Spezifikationen des Twitter-APIs, sowie die schriftlichen Richtlinien für seine Nutzung durch Entwickler (Twitter, 2012a) sind ein hervorragendes Beispiel für diesen „technologischen Anpassungen“ und Art und Weise, auf die diese direkt mit Twitters Suche nach einem tragfähigen Geschäftsmodell verwoben sind. Diese Veränderungen zeigen, wie die scheinbare semantische Offenheit oder „interpretative Flexibilität“ des Begriffs „Plattform“ es erlaubt, seine Bedeutung im Laufe der Zeit umzuformen, wenn Geschäftsmodelle oder Plattformeigner sich ändern (Gillespie 2010).

Die Freigabe des API wurde erstmals im September 2006 Mal auf dem Twitter-Blog angekündigt (Stone 2006) – nicht lange nach dem Start des Twitterservices, aber nachdem einige beliebte Drittanbieter-Anwendungen (wie ein Mashup von Twitter mit Google Maps, das eine dynamische Anzeige kürzlich geposteter Tweets auf weltweiter Basis erlaubte) bereits entwickelt worden waren. Seitdem hat Twitter vom Erblühen dessen, was das Unternehmen selbst als „Twitter-Ökosystem“ bezeichnet hat (Twitter 2012a; unsere Übersetzung), profitiert, einschließlich von Drittanbietern entwickelter Softwareclients (wie Twitterific und TweetDeck), Anwendungsfälle in wichtigen Institutionen (z.B. großangelegte Social-Media-Visualisierungen der London Riots im *Guardian*) und parasitärer Geschäftsmodelle (einschließlich Social-Media-Metrikendienste wie HootSuite und Klout). Während die Geschichte der Entwicklung von Twitters API-Regeln und den damit verbundenen regulatorischen Instrumente (wie seine Developer Rules of the Road und Terms of Use) viele Drehungen und Wendungen hat, stechen zwei besonders wichtige jüngere Kontroversen um Datenzugriff und -kontrolle hervor. Zum Einen hat das Unternehmen Entwickler und Forscher vom direkten Zugriff auf die „Firehose“ (sehr hohes Volumen), den kompletten Twitter-Feed ausgeschlossen; dies wurde von einer Aktion gegen freie und öffentliche Twitterarchivierungsdienste wie 140Kit und die Web-Version von Twapperkeeper



begleitet (Sample 2011) und fiel mit der Etablierung einer Content-Licensing-Vereinbarung zwischen Twitter und Gnip zusammen, die zu dieser der Firma Gnip ein Monopol für ihr Angebot eines hochvolumigen Zugangs zu Tweets (und den Inhalten anderer Social-Media-Plattformen) auf kommerziellem Preisniveau verschaffte. Eine zweite Welle der Kontroverse ergab sich unter der Entwickler-Community im August 2012 als Ergebnis von Twitters Veröffentlichung seiner neuesten API-Regeln (Sippey 2012), die unter bestimmten Umständen weitere, deutliche Grenzen für Verwendung und Verwendbarkeit des APIs setzten.

Im Wesentlichen ist das Ergebnis dieser Änderungen in den Twitter-API-Regeln, die ohne wirkliche Konsultation mit der Entwickler-Community, die das Twitter-Ökosystem schuf, verkündet wurden, eine erzwungene Neuausrichtung der Entwicklungsaktivitäten: auf der einen Seite versucht Twitter explizit, die Weiterentwicklung API-basierter Tools von Drittanbietern, die „consumer engagement activities“ ermöglichen (wie etwa Endnutzer-Twitterclients), zu beschränken (Sippey 2012), um die Nutzung der von Twitter selbst bereitgestellten und vermarkteten Endnutzertools steigern; auf der anderen Seite zielt das Unternehmen darauf ab, die weitere Entwicklung von „consumer analytics“ und „business analytics“ sowie „business engagement“-Tools zu „ermutigen“.

Implizit in diesen Änderungen ist eine Neupositionierung der Twitter-Nutzer (zunehmend als *Inhaltskonsumenten* statt als aktive Kommunikatoren), aber auch von kommerziellen und akademischen Forschern, die die Nutzung von Twitter untersuchen (als Bereitsteller einer beschränkten Reihe von bestehenden Twitter-„Analytics“-Ansätzen statt als Betreiber einer umfassenderen Untersuchung sowohl der aktuellen Nutzung von Twitter als auch der weiteren Entwicklung solcher Nutzung). Die Änderungen stellen einen Versuch des Unternehmens dar, eine bestimmte, kommerziell wünschenswerte und wertvoll Vision davon, wie Twitter genutzt (und analysiert) werden *sollte*, zu arretieren, und jegliche mögliche Weiterentwicklung jenseits dieser gewünschten Stufe zu verhindern oder zumindest zu verzögern. Obwohl es nicht unwahrscheinlich ist, daß solche Versuche, die Entwicklung gleichsam „einzufrieren“, angesichts der erheblichen, dokumentierten Rolle, die die Twitter-Nutzerbasis historisch in die Erforschung neuer und unvorhergesehener Verwendungen von Twitter gespielt hat (Bruns 2011; Halavais 2013), vergeblich sein werden, untergraben sie die wissenschaftliche Forschungsarbeit, tatsächliche Twitternutzung stichprobenartig zu untersuchen – was bedeutet, daß die Forscher zunehmend gezwungen sind, Zeit und Ressourcen in die Suche nach Umgehungsmöglichkeiten für die durch das Twitter-API auferlegten neuen Beschränkungen zu investieren.

## Technische, politische und epistemologische Fragen

In ihrem kürzlich erschienenen Artikel „Critical Questions for Big Data“ richten danah Boyd und Kate Crawford unsere Aufmerksamkeit auf die Beschränkungen, Politik und Ethik von Big-Data-Ansätzen in den Sozialwissenschaften insgesamt, aber auch berühren in der Erforschung sozialer Medien als einem besonders wichtigen Ort des Social Data Mining. Als Antwort bieten wir die folgenden ergänzenden Punkte, welche sich speziell auf datengestützte Twitter-Forschung beziehen, die sich auf mit Hilfe des Twitter-APIs gesammelten Tweetarchive stützt.

Zum Ersten, und anders als die meisten digitalen Geisteswissenschaften (wo Forscher oft mit einem großen bereits bestehenden Textkorpus beginnen), haben wir im Fall der Twitterforschung keinen Zugang zu einer ursprünglichen Sammlung von Texten – wir können nur auf das zugreifen, was Twitters proprietäres und häufig wechselndes API zur Verfügung stellt.

Die von Twitterforschern verwendeten Werkzeuge bauen auf verschiedene Kombinationen von Teilen des Twitter-APIs – oder, genauer gesagt, der verschiedenen Twitter-APIs (insbesondere die Search- und Streaming-APIs). Wie oben besprochen wird Twitter natürlich bei der Bereitstellung seines APIs nicht durch wissenschaftliche Interessen geleitet, sondern durch den Versuch, einer Reihe von potentiell wertschaffenden Endnutzern zu dienen – insbesondere denjenigen, mit denen Twitter direkte Geschäftsbeziehungen schaffen kann, wie etwa in seiner exklusiven Partnerschaft mit dem amerikanischen Fernsehsender NBC bei der Berichterstattung über die Olympischen Spiele 2012 in London.

Der folgende Abschnitt aus eigener Twitters Entwickler-FAQ hebt die möglichen Konflikte zwischen den Business-Case-Einsatzszenarien, für die die APIs zur Verfügung gestellt werden, und der tatsächlichen Nutzung, die akademische Forscher und andere Dataminers von ihnen machen, hervor:

Twitter-Suche ist optimiert, um Endnutzern relevante Tweets als Reaktion auf direkte, nicht-wiederkehrenden Anfragen wie #Hashtags, URLs, Domains und Keywords bereitzustellen. Das Search-API (das auch Twitters Such-Widget antreibt) ist eine Schnittstelle zu dieser Suchmaschine. Unser Suchdienst ist nicht dazu gedacht, ein erschöpfendes Archiv öffentlicher Tweets zu sein, und nicht alle Tweets werden im Index erfaßt oder zurückgegeben. Einige Ergebnisse werden zur besseren Bekämpfung von Spam und zur Erhöhung der Relevanz verfeinert. Aufgrund von

Kapazitätsengpässen deckt der Index derzeit nur etwa eine Woche an Tweets ab.  
(Twitter 2012b; unsere Übersetzung)

Da externe Forscher keinen Zugang zu den vollen, „rohen“ Daten haben, gegen die die vom API abgerufenen Archive, die wir in unseren späteren Analysen nutzen, verglichen werden könnten, und da unsere Datenzugangsansätze so stark auf Twitters API angewiesen sind – jeder Ansatz mit seinen eigenen technischen Besonderheiten und Grenzen – es ist für uns unmöglich, mit irgendwelcher Sicherheit zu sagen, daß wir ein komplettes Archiv oder wenigstens eine „repräsentative“ Stichprobe (was auch immer „repräsentativ“ in einem datenabhängigen, textualistischen Zusammenhang heißen mag) erfassen.

Mit anderen Worten, die „Listen“ von Tweets, die auf der Grundlage einer Stichwortsuche an uns geliefert werden, sind nicht unbedingt vollständig, und es gibt keine Möglichkeit zu wissen, *wie* unvollständig sie sind. Die Gesamtausbeute selbst der robustesten Erfassungssysteme (die das Streaming-API und nicht nur das Search-API benutzen) hängt ab von einer Reihe von Variablen: von Zugriffsgeschwindigkeitsbeschränkungen (Twitters sogenanntes „API rate limiting“), von den filternden und spambegrenzenden Elementen in Twitters Suchalgorithmus, von Serverausfällen und so weiter; zudem ist es, weil Twitter die Weitergabe von Datensätzen verbietet, auch schwerlich möglich, die eigenen Daten mit denen anderer Forschungsteams zu vergleichen.

Von einer epistemologischen Perspektive resultiert grundlegende Abhängigkeit auch in einer neuen Art von Forschung in den Medien-, Kommunikations- und Kulturwissenschaften: was entsteht, ist eine Form der durch Daten angetriebenen Forschung, die in Richtung eines abduktiven Denkens neigt; damit hebt es Spannungen zwischen den traditionellen Fragestellungen in diskurs- oder textbasierten Disziplinen wie Medien- und Kommunikationswissenschaften, und den Annahmen und Modi der Mustererkennung, hervor, die erforderlich werden, wenn die Arbeit „aus dem Korpus heraus“ stattfinden soll werden, anstatt von außen nach innen (für ein ausführliche Diskussion dieser epistemologischen Fragen in den „Digital Humanities“ allgemein vgl. Dixon 2012).

Schließlich sind auch die Heuristiken unserer Analysen von Twitterdatensätzen vom API vermittelt: die Datenpunkte, die in den Daten fest verankert sind, werden ganz natürlich zu den markantesten, und gestalten dadurch die weiteren Schritte. Zum Beispiel ist es ein gängiges Verfahren in unserer Forschung, die Syntax von Tweets selbst verwenden, um sie als eine der folgenden Arten von Aktivitäten zu kategorisieren:

<b>Originaltweets:</b>	Tweets, die weder @reply noch Retweets sind
<b>Retweets:</b>	Tweets, die ein RT @user... (oder ähnlich) enthalten
<b>Unveränderte Retweets:</b>	Retweets, die mit RT @user... anfangen
<b>Veränderte Retweets:</b>	Retweets, die nicht mit RT @user... anfangen
<b>Wirkliche @replies:</b>	Tweets, die @user enthalten, aber keine Retweets sind
<b>URL-Weiterleitungen:</b>	Tweets, die URLs enthalten

Diese Kategorien werden durch die textlichen und technischen Symbole für bestimmte Arten von Interaktion, die in die Syntax von Twitter selbst eingebaut sind, bestimmt (@replies oder @mentions, Retweets), sowiedurch spezifische Formen von Informationsreferenzen (URLs). Alle von ihnen konzentrieren sich auf (und privilegieren damit tendenziell) eher informationsorientierte Formen der Kommunikation als die ephemären, affektiven oder ambient intimen Verwendungen von Twitter welche leichter durch ethnografische Forschungsansätze beleuchtet werden können: Ansätze, die sich tatsächlich auf einzelne Nutzer, ihre sozialen Kontexte, und den weiteren kulturellen Kontext der Spuren, die sie auf Twitter hinterlassen, konzentrieren können.

### Schlussfolgerungen

In diesem Artikel haben wir einige der soziotechnischen, politischen und wirtschaftlichen Aspekte der strukturierten Twitter-Daten – der Datenlisten –, auf die unsere Forschung aufbaut und die mit Hilfe des Twitter-APIs erfaßt werden können, beschrieben und diskutiert. Wie wir an anderer Stelle argumentiert haben (Bruns und Burgess 2012) – und wie wir hoffentlich in diesem Artikel demonstriert haben – sind die Medien- und Kommunikationswissenschaftler, die tatsächlich an der Nutzung computergestützten Forschungsmethoden beteiligt sind, in einer guten Stellung, um sowohl zu den methodischen Fortschritten, die wir am Anfang dieses Artikels herausgestellt haben, als auch zu den politischen Debatten um die computerbasierten Forschungsmethoden des „big social data“-Moments, auf die wir uns im zweiten Teil dieses Artikels bezogen haben, einen wichtigen Beitrag zu leisten.

Eine dringendes Problem im Bereich der Methodik ist es, die auf die gegenwärtigen Fortschritte aufzubauen, um großangelegte Datamingansätze mit ethnographischen und anderen qualitativen Ansätzen zusammenzubringen, insbesondere auch mit der detaillierten manuellen Textanalyse. Darüberhinaus gibt es für unseren Umgang mit dem „big social

data“-Moment ein drückendes Bedürfnis für die Entwicklung von „Code Literacy“ in den Medien-, Kommunikations- und Kulturwissenschaften. In erster Linie hat eine solche Kompetenz eine wichtig instrumentelle Verwendung: wie Manovich (2012: 472) argumentiert, erfordert die meiste Big-Data-Forschung in den Geisteswissenschaften derzeit kostspielige und zeitaufwendige (und manchmal auch entfremdende) Partnerschaften mit technischen Experten (in der Regel, Informatiker), weil die kostenlosen Tools, die Nichtprogrammierern zugänglich sind, immer einen im Vergleich zu dem, was mit Rohdaten und speziell entwickeltem Code erreicht werden kann, nur einen stark begrenzten Nutzen haben.

Aber Code Literacy ist auch ein Erfordernis rigoroser wissenschaftlicher Forschung im Kontext dessen, was David Berry den „computational turn“ nennt, der für ihn eine „dritte Welle“ der Digital Humanities bedeutet. Berry meint, daß Code und Software zunehmend selbst zu Objekten, nicht nur Werkzeugen, für die Forschung werden könnten:

Ich schlage vor, daß wir einen humanistischen Ansatz für das Thema Computer-Code einführen, indem wir unsere Aufmerksamkeit auf die weiteren Aspekte von Code und Software richten, und sie mit der Materialität dieser wachsenden digitalen Welt verknüpfen. Vor diesem Hintergrund wird die Frage des Codes zunehmend wichtig für das Verständnis in den Digital Humanities, und dient als eine Bedingung der Möglichkeit für die vielen neuen rechnerischen Formen, die unsere Erfahrung der zeitgenössischen Kultur und Gesellschaft vermitteln. (Berry 2012: 17; unsere Übersetzung)

Ein erster Schritt liegt hier in der Entwicklung einer robusteren, belastungsfähigen Kenntnis der konzeptionellen Modelle und methodischen Prioritäten, die durch die Funktion sowohl der Werkzeuge als auch der Quellen, die wir für „big social data“-Forschung nutzen, vorausgesetzt werden. Die Kenntnis davon, wie etwa das Twitter-API die Nutzungskulturen der Plattform vermittelt, sowie ein reflexives Engagement mit seiner vermittelnden Rolle in der datengestützten Twitterforschung, ermöglicht und unterstützt ein viel greifbareres kritisches Verständnis der Politik der Social-Media-Plattformen (Gillespie 2010), die nun so mächtige Akteure in den Medienökologie geworden sind.

## Bibliografie

- Berry, David M. (2012): "Introduction: Understanding Digital Humanities." In: David M. Berry (ed.), *Understanding Digital Humanities*, London: Palgrave Macmillan, pp. 1-20.
- boyd, danah/Crawford, Kate (2012): "Critical Questions for Big Data." In: *Information, Communication & Society* 15/5, pp. 662-79.
- Bruns, Axel (2012): "Ad Hoc Innovation by Users of Social Networks: The Case of Twitter." In: *ZSI Discussion Paper* 16 (<https://www.zsi.at/object/publication/2186>).
- Bruns, Axel/Burgess, Jean (2012): "Notes towards the Scientific Study of Public Communication on Twitter." In: Alexander Tokar *et al.* (eds.), *Science and the Internet*, Düsseldorf: Düsseldorf University Press, pp. 159-169.
- Bruns, Axel/Burgess, Jean/Crawford, Kate/Shaw, Frances (2012): *#qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods*, Brisbane: ARC Centre of Excellence for Creative Industries and Innovation (<http://cci.edu.au/floodsreport.pdf>).
- Burgess, Jean/Bruns, Axel (2012): "(Not) the Twitter Election: The Dynamics of the #ausvotes Conversation in Relation to the Australian Media Ecology." In: *Journalism Practice* 6/3, pp. 384-402.
- Dixon, Dan (2012): "Analysis Tool Or Research Methodology: Is There an Epistemology for Patterns?" In: David M. Berry (ed.), *Understanding Digital Humanities*, London: Palgrave Macmillan, pp. 191-209.
- Fuller, Matthew (2005): *Media Ecologies: Materialist Energies in Art and Technoculture*, Cambridge, Mass.: MIT P.
- Gillespie, Tarleton (2012): "The Politics of 'Platforms'." In: *New Media & Society* 12/3, pp. 347-64.
- Halavais, Alexander (2013): "Structure of Twitter: Social and Technical." In: Katrin Weller *et al.* (eds.), *Twitter and Society*. New York: Peter Lang, pp. 29-41.
- Harrington, Stephen/Highfield, Timothy J./Bruns, Axel (2012): "More than a Backchannel: Twitter and Television." In: José Manuel Noguera (ed.), *Audience Interactivity and Participation*, Brussels: COST Action ISO906 Transforming Audiences, Transforming Societies, Brussels, pp. 13-17 (<http://www.cost-transforming-audiences.eu/system/files/essays-and-interview-essays-18-06-12.pdf>).
- Lotan, Gilad/Graeff, Erhardt/Ananny, Mike/Gaffney, Devin/Pearce, Ian/boyd, danah (2011): "The Arab Spring: The Revolutions Were Tweeted: Information Flows during the 2011

- Tunisian and Egyptian Revolutions.” In: *International Journal of Communication* 5, pp. 1375-1405 (<http://ijoc.org/ojs/index.php/ijoc/article/view/1246/613>).
- Manovich, Lev (2012): “Trending: The Promises and the Challenges of Big Social Data.” In: Matthew K. Gold (ed.), *Debates in the Digital Humanities*, Minneapolis: U of Minnesota P, pp. 460-75.
- Mejias, Ulises A. (2012): “Liberation Technology and the Arab Spring: From Utopia to Atopia and Beyond.” In: *Fibreculture Journal* 20 (<http://twenty.fibreculturejournal.org/2012/06/20/fcj-147-liberation-technology-and-the-arab-spring-from-utopia-to-atopia-and-beyond/>).
- O’Reilly, Tim (2005): “What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software.” In: *O’Reilly Network* 30 Sep. 2005 (<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>).
- Papacharissi, Zizi/de Fatima Oliveira. Maria (2012): “Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt.” In: *Journal of Communication* 62/2, pp. 266-82.
- Sample, Mark (2011): “The End of Twapperkeeper (and What to Do about It).” In *The Chronicle of Higher Education: ProfHacker*, 8 Mar. 2011 (<http://chronicle.com/blogs/profhacker/the-end-of-twapperkeeper-and-what-to-do-about-it/31582>).
- Sippey, Michael (2012): “Changes Coming in Version 1.1 of the Twitter API.” In: *Twitter Developers Blog*, 16 Aug. 2012 (<https://dev.Twitter.com/blog/changes-coming-to-Twitter-api>).
- Stone, Biz (2006): “Introducing the Twitter API.” In: *Twitter Blog*, 20 Sep. 2006 (<http://blog.Twitter.com/2006/09/introducing-Twitter-api.html>).
- Twitter (2012a): “Developer Rules of the Road.” In: *Twitter Developers Website*, 17 May 2012 (<https://dev.Twitter.com/terms/api-terms>).
- Twitter (2012b): “Frequently Asked Questions.” 18 Sep. 2012 (<https://dev.twitter.com/docs/faq>).
- Van Dijck, José (2011): “Tracing Twitter: The Rise of a Microblogging Platform.” In: *International Journal of Media and Cultural Politics* 7/3, pp. 333-48.
- Walker Rettberg, Jill (2008): *Blogging*. Cambridge: Polity.
- Wood, Megan M./Baughman, Linda (2012): “Glee Fandom and Twitter: Something New, or More of the Same Old Thing?” In: *Communication Studies* 63/3, pp. 328-44.