# arbido

2020/2 Abbild und Inszenierung der Gesellschaft

Schubert Irina, Archivinformatikerin bei Docuteam

Tags: Archiv Social Media, Schlaglicht,

## Wie und weshalb wir Twitter-Timelines archivieren

Twitter verfügt über grosses historisches Potenzial. Selbst wenn auf der Plattform sämtliche Tweets bis in die Anfangszeit zurück abrufbar sind, ist die Verfügbarkeit für die längerfristige Zukunft nicht gesichert. Bei docuteam haben wir ein Verfahren entwickelt, Twitter-Timelines zu archivieren.

Fast jeder kennt Twitter, die Microblogging-Plattform, auf der Informationen in Realtime verbreitet werden, Communities und Meinungen gebildet – und ebenso manipuliert werden. Die Publikationsgeschwindigkeit ist hoch, die Reaktionszeit kurz. Das macht Twitter zu einem Brandbeschleuniger für Debatten aller Art. Twitter erlaubt eine Demokratisierung der Berichterstattung, ermöglicht Meinungsfreiheit und dient der Mobilisierung politischer und gesellschaftlicher Bewegungen. Es sorgt aber auch für die Verbreitung von Fake-News, Hasskommentaren und Belästigungen.

Der erste Tweet: "just setting up my twttr" wurde vom Mitbegründer und heutigem CEO Jack Dorsey am 21. März 2006 abgesetzt. "twttr", so hiess der Prototyp damals, wurde eigentlich als Kommunikationsmittel für ein anderes Projekt entwickelt. [1] So zumindest eine populäre Legende. Die Gründer öffneten ihre Plattform bald einer breiten Nutzerschaft – und waren über den Erfolg selbst überrascht. [2] Twitter Inc. wurde ein Jahr später gegründet und entwickelt das Produkt mit mittlerweile rund 5000 Mitarbeitenden noch heute weiter. Bereits 2007 wurden Hashtags eingeführt, um Themen zu strukturieren. Als SMS-kompatibler Dienst entwickelt, waren lange nur 140 Zeichen erlaubt (die restlichen 20 waren für den Usernamen reserviert). Seit 2017 sind Tweets von einer Länge bis 280 Zeichen möglich. Heute (Stand 1. Quartal 2020) zählt der Dienst 166 Mio. sogenannte mDAU («monetizable Daily Active Usage»).[3] Eine Zahl, die seit Jahren stetig wächst. Monatlich werden über 500 Mio. Nutzer\*innen gezählt, die Twitter besuchen, ohne sich einzuloggen.[4] Gewinn macht Twitter Inc. erst seit Ende 2017 und zwar vor allem über Werbung.[5][6] Der Gesamtumsatz im 1. Quartal 2020 liegt bei über 800 Mio. US-Dollar.[7] Pro Sekunde werden aktuell durchschnittlich etwa 6000 Tweets abgesetzt. Das macht rund 500 Mio. am Tag. Die meisten Follower hat aktuell Barack Obama.

Twitter ist auch ein Ort der Selbstinszenierung. Dies gilt für Einzelpersonen ebenso wie für Unternehmen und öffentliche Einrichtungen. Es werden Beziehungen gepflegt und Marketing betrieben. Für das konstante Erstellen und Einspeisen von Inhalten wird viel Geld investiert. Die Kommunikation ist weniger formell als in gedruckten Medien und geht zudem in zwei Richtungen: Posts werden nicht nur gelesen, sondern auch zeitnah kommentiert, «geliked» und «retweetet».

Was bringt unsere Kund\*innen dazu, ihre Twitterkanäle archivieren zu lassen? Neben der empfundenen historischen Relevanz komplettiert das Social-Media-Archiv die klassischen Bestände gedruckter Kommunikationsmittel. Ohne den digitalen Kanal wäre das Bild der Kommunikationstätigkeit unvollständig. Daneben gibt es auch ganz praktische Argumente. Beispielsweise ist es für ein Unternehmen wichtig, den genauen Zeitpunkt und Wortlaut einer öffentlich kommunizierten Mitteilung nachvollziehen und belegen zu können. Sind die Twitterdaten erst einmal archiviert, haben sie eine Prüfsumme. Die Daten können so im Nachhinein nicht unbemerkt verändert werden.

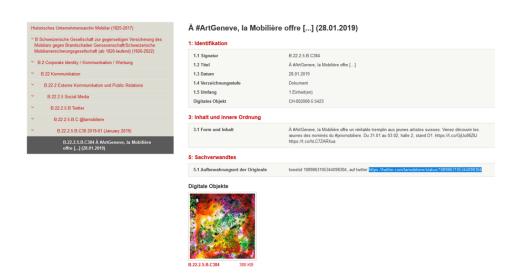
A	8	C	D	E	F	6	H		J	K	L	M	N	0	P	Q R
username	screen_name	id_str	full_text created_at		in reply to state in reply to is quote st		retweet co+favorite co+lang		e .	media_url media_md5 media_filer* media_orig* media_dow* record_o					werecord created at	
2 Die Mobiliar	@diemobiliar	1114155056557895680	@tom_studer Lieber Thomas Studer Viele	2019-04-05 13:17:20	1114137551869	tom_studer	False	0	0 de		None	None	None	None	None	2019-04-06 04:28:
3 Die Mobiliar	@diemobiliar	1115544055537127426	#Jahresergebnis / Mobiliar weiter auf V	2019-04-09 09:16:43	None	None	False	4	12 de		None	None	None	None	None	2019-04-10 04:28:
4 Die Mobiliar	@diemobiliar	1115543468573700096	Für eine erfolgreiche Zukunft investiert d	2019-04-09 09:14:23	1115533585569	diemobiliar	False	0	2 de		None	None	None	None	None	2019-04-10 04:28:
5 Die Mobiliar	@diemobiliar	1115540086085566464	«Die Resultate machen uns stark für morg	2019-04-09 09:00:56	None	None	False	3	20 de		http://pb	s.•b2a77a7	b5•1115540	09 TunR-Qd	None None	2019-04-10 04:28:
6 Die Mobiliar	@diemobiliar	1115533585568751616	Die Mobiliar präsentiert heute den Medie	2019-04-09 08:35:07	None	None	False	1	8 de		http://pb	4192109	5#1115533	59 D3sr4UP	Ws* None	2019-04-10 04:28:
7 Die Mobiliar	@diemobiliar	1115857788717150209	Ob Schaden oder nicht: Unsere Kunden er	2019-04-10 06:03:23	None	None	False	1	9 de		http://pb	s •fd86322	5fb 1115857	79 LPdMM2	7 None	2019-04-11 04:28:
8 Die Mobiliar	@diemobiliar	1116220145255448576	Ob Schaden oder nicht: Unsere Kunden er	2019-04-11 06:03:15	None	None	False	0	7 de		http://pb	s. •c2b803fe	59 1116220	14 tqBdZtN	wa* None	2019-04-12 04:29:
9 Die Mobiliar	@diemobiliar	1117307255525867520	Liebe #GOT Fans, wir wünschen viel Spas#	2019-04-14 06:03:02	None	None	False	10	29 de		http://pb	s. • 0a2f52f7	42 1117307	29 D4F5BrV	VX*None	2019-04-15 04:28:
10 Die Mobiliar	@diemobiliar	1117798200629448707	Lieber Hallensport Warum die Mobiliar P	2019-04-15 14:33:53	None	None	False	0	4 de		None	None	None	None	None	2019-04-16 04:28:
11 Die Mobiliar	@diemobiliar	1117786940630339584	Datendiebstahl, Hackerangriffe oder Schaft	2019-04-15 13:49:08	None	None	True	3	6 de		None	None	None	None	None	2019-04-16 04:28:
12 Die Mobiliar	@diemobiliar	1117773458652454912	Wachgeküsst: Wie die Mobiliar mitgeholf≠	2019-04-15 12:55:34	None	None	False	0	6 de		None	None	None	None	None	2019-04-16 04:28:
13 Die Mobiliar	@diemobiliar	1117665160951365634	Ob Schaden oder nicht: Unsere Kunden er	2019-04-15 05:45:14	None	None	False	1	2 de		http://pb	s. • 7c2e664	15 1117665	10 nrOUrnt	F9# None	2019-04-16 04:28:
14 Die Mobiliar	@diemobiliar	1118165905983377408	Die liebste Versicherung der Schweizer!	2019-04-16 14:55:01	None	None	False	3	8 de		None	None	None	None	None	2019-04-17 04:28:
15 Die Mobiliar	@diemobiliar	1118031320909864960	Ob Schaden oder nicht: Unsere Kunden et	2019-04-16 06:00:13	None	None	False	1	2 de		http://pb	s. 174e5357	25 1118031	32 RzżynoJU	JX None	2019-04-17 04:29:
16 Die Mobiliar	@diemobiliar	1120662553917566976	@BizzarriPatrick Als nicht börsenkotierte	2019-04-23 12:15:48	1119899578512	BizzarriPatri	False	0	0 de		None	None	None	None	None	2019-04-24 04:30:
17 Die Mobiliar	@diemobiliar	1121072337892315136	@Michael Koepfli Keramik, Stein und Eis#	2019-04-24 15:24:08	1120950158259	Michael Ko	False	0	7 de		None	None	None	None	None	2019-04-25 04:29:
18 Die Mobiliar	@diemobiliar	1121434851415089153	@BizzarriPatrick Die Mobiliar Genossensch	2019-04-25 15:24:38	1121343047382	BizzarriPatri	False	0	0 de		None	None	None	None	None	2019-04-26 04:29:
19 Die Mobiliar	@diemobiliar	1122778330380603393	Die neue #Kunst-Ausstellung der Mobiliar*	2019-04-29 08:23:08	None	None	True	1	5 de		None	None	None	None	None	2019-04-30 04:29:
20 Die Mobiliar	@diemobiliar	1124317814830129156	Jean-Philippe Moser übernimmt am 1. Set	2019-05-03 14:20:30	None	None	False	0	5 de		http://pb	s. • 7f72db9l	bar 1124317	83 D5pg9e/	W/ None	2019-05-04 04:28:
21 Die Mobiliar	@diemobiliar	1124246379948277761	Auch die Mobiliar unterstützt die Schwei	2019-05-03 09:36:39	None	None	True	0	3 de		None	None	None	None	None	2019-05-04-04:28:
22 Die Mobiliar	@diemobiliar	1125329528983834624	Unsere Asset Manager setzen beim Anlege	2019-05-06 09:20:41	None	None	False	2	3 de		http://pb	s. • 09cd57f8	5#1125329	52 D535Pd	(W) None	2019-05-07-04:28:
23 Die Mobiliar	@diemobiliar	1125668382479532032	«Wichtig, dass wir uns klarmachen, welch	2019-05-07 07:47:10	None	None	False	4	8 de		None	None	None	None	None	2019-05-08 04:28:
24 Die Mobiliar	Ødiemobiliar	1126087905485705216	Container-Technologie hilft unserer IT bei	2019-05-08 11:34:13	None	None	False	5	15 de		None	None	None	None	None	2019-05-09 04:28:
25 Die Mobiliar	@diemobiliar	1126455639134359555	RT @meteoschweig: Wie funktioniert eige	2019-05-09 11:55:27	None	None	False	4	0 de		None	None	None	None	None	2019-05-10 04:29:
26 Die Mobiliar	@diemobiliar	1126786446071877632	Im Rahmen des Experts Live Café dürfen ♥		None	None	False	7	15 de		None	None	None	None	None	2019-05-11 04:29:
27 Die Mobiliar	@diemobiliar	1128667747381125122	RT @meteoschweiz: Zum Abschluss unset	2019-05-15 14:25:35	None	None	False	8	0 de		None	None	None	None	None	2019-05-16 04:29:
28 Die Mobiliar	@diemobiliar	1128615203720126465	RT @FilmFestLocarno: Ecological Musical F	2019-05-15 10:56:47	None	None	False	5	0 en		None	None	None	None	None	2019-05-16 04:29:
29 Die Mobiliar	Ødiemobiliar	1129011188447272960	@ChrizzAlive Guten Tag Chrizz Wir haben*	2019-05-16 13:10:17	1128743583450	CKaufmant	False	0	0 de		None	None	None	None	None	2019-05-17 04:29:
30 Die Mobiliar	@diemobiliar	1130362329357574144	Zusammen mit den MoBees feiern wir de	2019-05-20 06:39:15	None	None	False	3	7 de		http://pbs.+32eac6adc3-113036232-n42qE49IR3-None			IR3 None	2019-05-21 04:28:	
31 Die Mobiliar	@diemobiliar	1131889989187510272	Wallis und Tessin vertreten: Die Delegier	2019-05-24 11:49:37	None	None	False	1	4 de		None	None	None	None	None	2019-05-25 04:28:
32 Die Mobiliar	@diemobiliar	1132952582719188992	@zyssetli Danke für den Hinweis, das soll*	2019-05-27 10:11:59	1132932954861	zyssetli	False	0	0 de		None	None	None	None	None	2019-05-28 04:29:
33 Die Mobiliar	@diemobiliar	1132937647930974208	Cyberkriminelle kosten die Weltwirtschaff	2019-05-27 09:12:38	None	None	False	2	4 de		None	None	None	None	None	2019-05-28 04:29:
34 Die Mobiliar	@diemobiliar	1135857398973968386	@JupiterKallisto Alles im grünen Bereich -	2019-06-04 10:34:41	1135574818589	JupiterKalls	False	0	0 de		None	None	None	None	None	2019-06-05 04:29:
35 Die Mobiliar	@diemobiliar	1136200509143040001	Die unterschätzte Naturgefahr: Die Hälfte	2019-06-05 09:18:05	None	None	False	7	8 de		None	None	None	None	None	2019-06-06-04:30:
36 Die Mobiliar	@diemobiliar	1138463829719945218	Die Mobiliar hifft mit, das obere Simmenta				False	5	6 de			s. 7a80dacı		82 YalHGAV	(or None	2019-06-12 04:30:
37 Die Mobiliar	@diemobiliar	1138743922698727425	Ob Diversität, gleiche Löhne für Mann und			None	False	0	8 de		None	None	None	None	None	2019-06-13 04:30:
38 Die Mobiliar	@diemobiliar	1139494408510947329	@FabianMolinaNR Lieber Fabian Molina V		1139253311709	EabianMolit	Ealse	0	17 de		None	None	None	None	None	2019-06-15 04:30:
39 Die Mobiliar	@diemobiliar	1141010675059367936	RT @CST aktuelt "Kantone stellen sich hill			None	False	4	0 de		None	None	None	None	None	2019-06-19 04:30:
40 Die Mobiliar	@diemobiliar	1141616198381166593	Für immer mehr dieltale Geschäfte müsse				True	2	2 de		None	None	None	None	None	2019-06-21 04:29:
41 Die Mobiliar	@diemobiliar	1142775434637172736	Die Mobiliar rechnet nach den heftigen G			None	False	4	1 de					43 UwfE1ni		2019-06-24 04:29:
42 Die Mobiliar	@diemobiliar	1143151906111533056	Schadeninspektor Daniel Picci ist seit S		None None False			1	6 de		http://pbs.+1366d9659-114315190-dlNJP2GSSP None					2019-06-25 04:30:
43 Die Mobiliar	@diemobiliar	1143138650777145345	@wiedasleben Hallo Jeff Danke für den H#		1142493295749			0	0 de		None	None	None	None	None	2019-06-25 04:30:
44 Die Mobiliar	@diemobiliar	1145697037800955904	Die Mobiliar wird Mitglied von #Cardossie			None	False	1	6 de		None	None	None	None	None	2019-07-02 04:29:
45 Die Mobiliar	@diemobiliar	1145076150416007169	Erster Platz: Die Mobiliar Kunden sind a			None	False	2	5 de		None	None	None	None	None	2019-07-03-04:29:
45 Dishtabiles	Odlamakiliaa	1146070150416007107	Minimertoin file Swire Entropenance Sund			None	False	- 1	5 do		None	Ness	None	None	None	2010 07 03 04:20

Ein Blick in die Metadaten

Seit dem ersten Tweet sind 14 Jahre vergangen. Da die Plattform ihre Daten bis in die Anfangszeit zurück zugänglich macht (ausser die Kontoinhaber\*in löscht diese selbst), wird Twitter oft auch als Archiv bezeichnet. Dass man darauf mit Blick auf die Langzeitarchivierung nicht zählen sollte, muss hier nicht weiter begründet werden. Was zeigt uns ein Blick in unsere Branche? Das bekannteste Projekt war vermutlich die Archivierung sämtlicher Tweets durch die Library of Congress im Jahr 2010. Es stiess auf grosse Begeisterung, obwohl nur Text archiviert wurde, löste aber auch einige Skepsis in Bezug auf Datenschutz aus. So stehen die archivierten Tweets in ihrer Gesamtheit bis heute unter Verschluss. Das Projekt fand sieben Jahre später sein Ende. Seither archiviert die LoC nur noch ausgewählte Tweets.

Was bedeutet das für Archivar\*innen? Es bedeutet, dass wir uns selbst um die Archivierung kümmern sollten, wenn uns unsere Tweets lieb (das heisst: archivwürdig) sind. Twitter bietet mehrere Programmierschnittstellen an, mit denen sich die Daten automatisiert abgreifen lassen.[8] Daneben gibt es Unternehmen, die eine detaillierte Auswertung von Twitter- und anderen Social-Media-Daten beispielsweise für Marketing- oder politische Zwecke anbieten. Dafür zahlt man aber sehr viel Geld. Und sind die Daten damit wirklich gesichert? Bedingt.

Was machen wir bei docuteam? Wir speichern die Tweets mitsamt Metadaten in eine Datenbank und überführen diese Daten periodisch in ein digitales Langzeitarchiv. Zusätzlich können die heruntergeladenen Daten über eine einfache passwortgeschützte Web-Oberfläche jederzeit angesehen und durchsucht werden. Die letzten 3200 Tweets erfahren täglich ein Update, was die Anzahl Retweets und Likes anbelangt. Kommentare werden nicht archiviert. Neben dem Text des Tweets wird zusätzlich jeweils ein Bild, sofern vorhanden, mitgespeichert. Das Herunterladen von Videos lässt Twitter zurzeit nicht zu. Für das Abgreifen der Daten benutzen wir die Twitter API, für die Archivierung in der Regel unsere eigenen Tools.



Ein Tweet auf der aktuellen Seite von Twitter (links); ein Tweet im Archivkatalog (rechts)

Sobald die Daten ins Langzeitarchiv überführt worden sind, sind sie fixiert. Das heisst, sie erfahren keine Aktualisierung mehr, befinden sich in der sicheren Umgebung eines digitalen Langzeitarchivs und sind über das Archivinformationssystem auffindbar.

Unsere bisherige Erfahrung zeigt, dass unser Ansatz für die Twitterarchivierung besonders bei Unternehmensarchiven auf Interesse stösst. Wir sind aber der Meinung, dass die Archivierung von Twitter- und allgemein Social-Media-Daten grundsätzlich für jedes Archiv eine Überlegung wert ist.

- [1] Reissmann, Ole et al.: «Meilensteine des Weitererzähl-Webs» [online], 2011, https://www.spiegel.de/netzwel... (Zugriff am 19.05.2020)
- [2] Ebd.
- [3] Twitter Inc.: «Quarterly results» [online], 2020, <a href="https://investor.twitterinc.co...">https://investor.twitterinc.co...</a> (Zugriff am 19.05.2020)
- [4] Rao, Deepak: «Testing promoted Tweets on our logged-out experience» [online], 2015, https://blog.twitter.com/en\_us... (Zugriff am 19.05.2020)
- [5] Kollewe, Julia: «Twitter makes first quarterly profit in its history» [online], 2018, https://www.theguardian.com/te... (Zugriff am 19.05.2020)
- [6] Gadkari, Pia: «How does Twitter make money?» [online], 2013, https://www.bbc.com/news/busin... (Zugriff am 19.05.2020)
- [7] Twitter Inc.: «Quarterly results» [online], 2020, <a href="https://investor.twitterinc.co...">https://investor.twitterinc.co...</a> (Zugriff am 19.05.2020)
- [8] Twitter Developer Documentation [online], 2020, https://developer.twitter.com/... (Zugriff am 19.05.2020)



### Irina Schubert

Archivinformatikerin bei Docuteam GmbH.

#### Abstract

#### **Deutsch**

Twitter verfügt über grosses historisches Potenzial. Selbst wenn auf der Plattform sämtliche Tweets bis in die Anfangszeit zurück durchsuch- und abrufbar sind, ist die Verfügbarkeit für die längerfristige Zukunft nicht gesichert. Das wissen wir in unserer Branche nur zu gut. Bei docuteam haben wir deshalb ein Verfahren entwickelt, Twitter-Timelines für Archive, die von uns betreut werden, zu archivieren.

### **Français**

Twitter a un grand potentiel historique. Même si tous les tweets de la plateforme peuvent être recherchés et récupérés depuis les premiers jours, leur disponibilité à long terme n'est pas garantie. Dans notre secteur, nous ne le savons que trop bien. A docuteam, nous avons donc développé une procédure d'archivage des calendriers Twitter pour les archives que nous gérons.