

Verhaltens-, Forschungs- oder Datenschnittstellen? Application Programming Interfaces (APIs) aus diachron und synchron vergleichender Perspektive

Jünger, Jakob

Erstveröffentlichung / Primary Publication

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Jünger, J. (2022). Verhaltens-, Forschungs- oder Datenschnittstellen? Application Programming Interfaces (APIs) aus diachron und synchron vergleichender Perspektive. In C. Schwarzenegger, E. Koenen, C. Pentzold, T. Birkner, & C. Katzenbach (Hrsg.), *Digitale Kommunikation und Kommunikationsgeschichte: Perspektiven, Potentiale, Problemfelder* (S. 157-185). Berlin <https://doi.org/10.48541/dcr.v10.6>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:
<https://creativecommons.org/licenses/by/4.0>

Empfohlene Zitierung: Jünger, J. (2022). Verhaltens-, Forschungs- oder Datenschnittstellen? Application Programming Interfaces (APIs) aus diachron und synchron vergleichender Perspektive. In C. Schwarzenegger, E. Koenen, C. Pentzold, T. Birkner & C. Katzenbach (Hrsg.), *Digitale Kommunikation und Kommunikationsgeschichte: Perspektiven, Potentiale, Problemfelder* (S. 157–185). <https://doi.org/10.48541/dcr.v10.6>

Zusammenfassung: Betrachtet man APIs aus repräsentativer Perspektive, werden darüber erhobene Daten als Spuren von Verhalten bzw. Handeln und damit als Indikatoren für eine zugrunde liegende Wirklichkeit aufgefasst. Aus einer interaktionalen Perspektive sind Daten und Verhalten dagegen miteinander verflochten, die Organisationsweise von APIs kann Verhalten fördern und behindern. Daten werden nicht vorgefunden, sondern interaktiv erzeugt. Schließlich sind APIs in ein Ökosystem aus Plattformbetreibenden, Nutzenden und Drittanbietenden eingebunden, sodass Plattformen aus einer ökologischen Perspektive als kuratierende Datenintermediäre zu betrachten sind. Der Beitrag differenziert diese Perspektiven, indem aus der Analyse von Dokumentationen, Berichterstattung und Literatur die Konsequenzen für wissenschaftliche Forschung abgeleitet werden.

Lizenz: Creative Commons Attribution 4.0 (CC-BY 4.0)

Jakob Jünger

Verhaltens-, Forschungs- oder Datenschnittstellen?

Application Programming Interfaces (APIs) aus diachron und synchron vergleichender Perspektive

1 Einleitung

In den Sozialwissenschaften haben sich Application Programming Interfaces (APIs) als ein Weg der Datenerhebung etabliert. Diese Schnittstellen erlauben es insbesondere auf Social-Media-Plattformen, automatisiert auf strukturierte Daten zuzugreifen, die Auskunft über gesellschaftliche Diskurse, Gemeinschaftsbildung und Nutzungsverhalten geben können. Zahlreiche Studien bauen auf dieser Form der Datenerhebung auf – etwa bei der Analyse von ‚Titterspheres‘ in Ländern wie Australien (Bruns & Moon, 2019), Deutschland (Münch et al., 2021), Japan (Kobayashi et al., 2019), Norwegen (Bruns & Enli, 2018) oder der Türkei (Dogu & Mat, 2019). Dabei wird in solchen Studien immer wieder von spezifischen Herausforderungen bei der Datenerhebung berichtet. Da die Datenzugänge durch die entsprechenden Plattformen kontrolliert werden, stellt sich etwa die Frage, inwiefern damit gewonnene Erkenntnisse valide und nachvollziehbar sind. Zudem wirken Daten auf die Nutzenden zurück. So werden Nutzungsdaten von den Plattformbetreibern ausgewertet, beispielsweise um auf YouTube zu den aufgerufenen

Videos weitere Inhalte zu empfehlen. Auch Drittanbietende nutzen die über APIs bereitgestellten Daten, etwa um Werbekampagnen auszuwerten und über die Sichtbarkeit von Beiträgen Kommunikationsverhalten zu beeinflussen. Damit ist die sozialwissenschaftliche Datenerhebung in ein Ökosystem eingebunden, die in den Datensätzen versammelten Daten verweisen gleichzeitig auf das Verhalten der Nutzenden, auf die Mechanismen der Plattformen und auf die Erhebungsprozesse von Wissenschaftler:innen. Die verschiedenen Einflussfaktoren lassen sich allerdings kaum voneinander trennen, solange die Funktionsweise der Plattformen nicht ausreichend verstanden wird. Für die Interpretation von Befunden ist relevant, inwiefern sich etwa die Zuschreibung von Popularität zu einzelnen Inhalten aus dem Sampling der Forschenden, der strategischen Steuerung durch Plattformen oder den individuellen Präferenzen von Nutzenden ergibt. Sichtbar werden die unterschiedlichen Funktionsweisen, das ist der Ausgangspunkt dieses Beitrags, vor allem im Vergleich verschiedener Plattformen und wenn man sich die Entwicklungsgeschichte von APIs verdeutlicht.

Die Betrachtung von APIs erscheint für ein Verständnis von Onlineplattformen aus zwei Gründen besonders geeignet zu sein. Erstens werden hierüber alle drei Akteurstypen miteinander verbunden, das Verhalten von Nutzenden wird vermittelt über die Plattformbetreibenden sichtbar für Wissenschaftler:innen. Zweitens sind APIs nicht nur für alle Plattformen ähnlich dokumentiert, sondern auch versioniert, sodass sich die Entwicklungsgeschichte nachvollziehen lässt. Im Beitrag wird deshalb ausgelotet, inwiefern eine historisch-vergleichende Perspektive auf APIs Erkenntnisse über die drei miteinander verwobenen, datengenerierenden Prozesse hervorbringt.

2 Made Data, Found Data und Curated Data

Klassischerweise dominiert in sozialwissenschaftlichen Studien eine *repräsentative Perspektive* auf Daten, “focused on the capacity of digital data analyses to represent adequately given social phenomena and patterns” (Marres, 2018, S. 10). Das heißt, Daten werden messtheoretisch als Indikatoren für latente Konstrukte wie beispielsweise Inzivilität oder Rationalität von Diskussionen angesehen. In der methodologisch ausgerichteten Literatur werden dementsprechend vor allem Datenausfälle oder Verzerrungen bzw. Validitätsprobleme bei der Datenerhebung

thematisiert (z. B. Hargittai, 2020; Gerlitz & Rieder, 2013; Ruths & Pfeffer, 2014). Aus dieser Perspektive erscheinen Daten als Hinterlassenschaften von Verhalten, die gesammelt und ausgewertet werden können. Die wissenschaftliche Datenerhebung, -aufbereitung und -analyse von *trace data* oder *tracking data* ist ein Prozess der Datenkonstruktion, der unter der idealisierten Annahme stattfindet, dass Inhalte und Verhalten unabhängig von der wissenschaftlichen Analyse vorgefunden werden (Jensen, 2013; Vis, 2013).

Dagegen setzt sich eine *interaktionale Perspektive* mit der Datafizierung von Verhalten auseinander (z. B. Cukier & Mayer-Schoenberger, 2013; siehe auch Marres, 2018, S. 33). Der Begriff Datafizierung verweist nicht nur darauf, dass Verhalten durch Daten sichtbar gemacht wird, sondern vielmehr darauf, dass Daten eine Grundlage für Verhalten sind: „data can be understood either as representations of reality or as resources for acting in and on reality“ (Jensen, 2014). Besonders offensichtlich wird diese Verwendung von Daten, wenn sie in Empfehlungssysteme eingespielt werden. Das sogenannte kollaborative Filtern führt dann dazu, dass Nutzenden solche Inhalte angezeigt werden, die von anderen Nutzenden mit einer ähnlichen Nutzungshistorie rezipiert wurden (Ricci, Rokach & Shapira, 2015, S. 12). In Bezug auf APIs wird diese Perspektive einer datengetriebenen Verhaltensbeeinflussung auch in den Dokumentationen der Plattformen sichtbar, beispielsweise wenn Facebook eigene APIs als Treiber von Nutzungsverhalten anpreist: „The New Page Experience API allows your app to manage the New Page Experience for Pages to drive public content, engagement, and discourse on your Facebook Page“ (Facebook, 2021).

Nicht nur die Betreibenden von Online-Plattformen basieren Entscheidungen auf Verhaltensdaten, sondern auch die Nutzenden selbst. So dienen die aggregierten Angaben der Views, Comments oder Likes eines Inhalts als Indikator für Popularität (siehe z. B. Haim, Kümpel & Brosius, 2018; Porten-Cheé et al., 2018). Sie können damit als Orientierung dafür dienen, ob sich die Rezeption lohnt („Social Navigation“; Hautzner, Lünich & Rössler, 2012). In der Folge entstehen Daten aus Daten. Ein wesentlicher Unterschied zwischen repräsentativer und interaktionaler Perspektive besteht somit darin, ob Daten als Zeugnisse von Verhalten (*found data*) oder Erzeugnisse einer reflexiven Konstruktion verschiedener Akteure (*made data*) aufgefasst werden (Jensen, 2014, S. 224).

In Bezug auf APIs wird eine weitere Herausforderung für die Interpretation darauf aufbauender Analysen deutlich: über eine API sichtbare Daten müssen nicht mit der Anzeige auf der Webseite oder in einer App übereinstimmen. Das betrifft etwa die

Sortierung von Kommentaren, die mal chronologisch, mal nach plattforminternen Algorithmen bestimmt wird. Führt man sich verschiedene Nutzungsszenarien vor Augen, verkompliziert sich die Lage weiter. Denkbar ist zwar, dass Nutzende sich beispielsweise die Tweets auf einem Twitterprofil nacheinander durchlesen, vermutlich realistischer ist aber, dass sie häufig nur kurze Benachrichtigungen auf dem Smartphone oder die personalisierten Zusammenstellungen neuer Tweets wahrnehmen. Diese Personalisierung ist über APIs nicht abzubilden, da nur sehr eingeschränkt Rezeptionsindikatoren ausgegeben werden. Es handelt sich bei über APIs erhobenen Daten also lediglich um eine mögliche Repräsentation von Verhalten. Die Perspektive der Nutzenden unterscheidet sich von der Perspektive der Wissenschaftler:innen, ohne dass die genauen Unterschiede bekannt sind.

Für die Betreibenden von Online-Plattformen stellen Daten darüber hinaus eine Geschäftsgrundlage dar. Für den Datenzugriff durch Dritte werden häufig Programmierschnittstellen bereitgestellt, weshalb auch von einer „API economy“ (Doerrfeld et al., 2016) gesprochen wird. Indem sie den Zugang kontrollieren, kuratieren Plattformen die für andere Akteure verfügbaren Daten. Dementsprechend stellen Online-Plattformen nicht einfach neutrale Datenzugänge zur Verfügung, sondern gestalten diesen Zugang als Datenintermediäre (Dorfer, 2016), das heißt vorrangig ausgerichtet an den eigenen ökonomischen Interessen. Ausgehend von der Mediatisierungsforschung kann diese Sicht als *ökologische Perspektive* bezeichnet werden. Der Mediatisierungsbegriff verweist in diesem Zusammenhang darauf, dass die Entstehung, Formatierung und Verwendung von Informationstechnologien sozial vermittelt ist und eben kein rein technischer und erst recht kein objektiver, von Akteuren entkoppelter, Vorgang ist (Altheide, 1994, S. 670; Couldry & Hepp, 2017, S. 15; Livingstone, 2009, S. 5). Diese Sichtweise fordert dazu auf, eine Technologie wie APIs nicht einfach als Schnittstelle zwischen Daten und Analyse, sondern als Mittler mit eigenwilligen Verhaltensweisen aufzufassen.¹ Die Produktion, Aufbereitung und Nutzung von Daten ist in ein Ökosystem eingebettet, in dem das Verhalten von Nutzenden, Betreibenden und aber auch Wissenschaftler:innen vielfältig miteinander verwoben ist.

Alle drei Perspektiven haben ihre Berechtigung, denn Online-Daten sind in dreifacher Hinsicht prozessgeneriert: einmal durch die Nutzenden von Online-

1 In der Akteur-Netzwerk-Theorie wird dafür die Unterscheidung zwischen Zwischengliedern (intermediators) und Mittlern (mediators) eingeführt (siehe Latour, 2017, S. 71).

Diensten, einmal durch die Anbietenden der Datenzugänge und einmal durch die Wissenschaftler:innen im Zuge ihrer Datenerhebung. Wie auch in anderen Fällen prozessgenerierter Daten (Johnson & Turner, 2003) – man denke hier an Gerichtsakten, die als Verwaltungsvorgänge auf die verwalteten Subjekte und aber auch auf die Verwaltung zurückzuführen sind (Bick & Müller, 1984) – wird damit eine kritische Auseinandersetzung mit der Datenentstehung notwendig. Die Einteilung der Perspektiven stellt jeweils unterschiedliche Herausforderungen für wissenschaftliche Analysen heraus: erstens, sind die Daten repräsentativ, zweitens, inwiefern unterliegen die Daten reaktiven Prozessen, und drittens, welche Interessen verfolgen die Plattformen selbst, wenn sie Datenschnittstellen anbieten? Im Folgenden wird in Bezug auf Programmierschnittstellen diskutiert, welche Hinweise auf die Datenentstehung sich aus einem Vergleich von APIs ergeben, wenn man jeweils eine der drei in Tabelle 1 zusammengefassten Perspektiven einnimmt.

3 Application Programming Interfaces

Programmierschnittstellen legen fest, wie zwei verschiedene Programme miteinander interagieren können (Jacobson, Brail & Woods, 2012, S. 5). Online-Plattformen stellen dazu so genannte Endpunkte bereit, das sind URLs wie „<https://api.twitter.com/1.1/followers/list>“, über die auf Daten und Funktionen der Plattform zugegriffen werden kann. Die Entwickler:innen von Apps benutzen diese Endpunkte, um darauf eigene Software aufzubauen und für Endkund:innen bereitzustellen. Die App „Unfollow For Twitter Free“ (HowdyApps, 2020) ermöglicht beispielsweise über den genannten Endpunkt das Auslesen der eigenen Twitter-Follower, um all denjenigen Nutzenden zu ‚entfolgen‘, die einem selbst nicht folgen. Es lassen sich damit aus technischer Sicht grundlegend drei Rollen unterscheiden: API-Betreiber ist in diesem Fall Twitter, API-Konsument ist der Drittanbieter HowdyApps. Hinzu treten als weitere Akteure die Nutzenden der App.

Eine API ist damit nicht nur eine Software oder die Beschreibung einer Software, sondern letztendlich ein Vertrag zwischen Betreiber:in und Konsument:in (Jacobson, Brail & Woods, 2012, S. 4). Dieser Vertrag sichert zu, dass die Schnittstelle in Zukunft auf die immer gleiche Weise funktionieren wird. Denn erst auf dieser Grundlage lohnt sich die Investition von Drittanbietenden. Nicht immer

Tabelle 1: Sozialwissenschaftliche Perspektiven auf APIs²

Perspektive	Repräsentative Perspektive	Interaktionale Perspektive	Ökologische Perspektive
Zentrale Akteure	Wissenschaftler:innen (vermessen Wirklichkeit)	Nutzende (erzeugen Wirklichkeit)	Betreibende (gestalten Wirklichkeit)
Datenentstehung	Found data: Daten resultieren aus Handeln	Made data: Daten sind Grundlage für Handeln	Curated data: Daten als Geschäftsgrundlage
Rolle von APIs	Datenerhebung: Zugriff auf Daten	Datafizierung: Anregung von Aktivitäten	Datenintermediation: Partnerschaften zwischen Akteuren
Wissenschaftliches Hauptproblem	Datenausfälle, Verzerrungen & Validität	Reaktivität & Multiple Kontexte	Organisationsinteressen und Entstehungskontexte
Zeugnisse ²	Zugangsbeschränkungen in den API-Dokumentationen	Nutzungsszenarien der Anbieter	Berichterstattung und Sekundärliteratur

ist der Zeitraum explizit festgelegt, für den eine Funktionsgarantie gegeben wird, doch insbesondere die größeren Unternehmen veröffentlichen hierzu langfristige Entwicklungspläne (für ein Beispiel siehe weiter unten Abbildung 1). Im Gegenzug verpflichten sich die API-Konsument:innen zur Einhaltung der vom Anbieter vorgegebenen Nutzungsbedingungen. Hier wird häufig vorgeschrieben, dass die Drittangebote nicht in Konkurrenz zur Plattform stehen dürfen, sondern diese nur um neue Funktionen erweitern sollen.²

Im Kontext wissenschaftlicher Datenerhebung operieren Wissenschaftler:innen entweder direkt als API-Konsument:innen oder nachgelagert als Endnutzer:innen. Im einfachsten Fall erfolgt der Zugriff über selbst geschriebene Skripte, zum

² Siehe Kapitel 4.

Beispiel mit Python oder R, wobei Wissenschaftler:innen die API-Konsument:innen sind. Einige Programme erlauben die Arbeit mit APIs ohne eigene Programmierung, beispielsweise Facepager (Jünger & Keyling, 2020) oder Rapidminer (RapidMiner, 2020). Weitere Angebote lassen die API hinter einer auf das Wesentliche reduzierten Oberfläche verschwinden (z. B. DMI Tools, 2020). In allen diesen Fällen erfolgt der Zugriff auf eine API in der Regel durch die Anwendungsentwickler:innen, die als API-Konsument:innen im Namen der Wissenschaftler:innen agieren. Insbesondere wenn kommerzielle Dienstleistungen in Anspruch genommen werden, sind also immer mehrere Akteure beteiligt.

Der Datenzugang über APIs erlaubt genauso wenig wie eine Weboberfläche oder eine App ungefilterten oder direkten Zugang zur Datenbank einer Plattform, sondern erzeugt lediglich Repräsentationen von Datenbankauszügen. Genau genommen stellt das Web selbst bereits eine API dar, bei der ein Browser als API-Konsument über URLs wie „https://twitter.com/DGPuK_DigiKomm“ auf die Daten einer Plattform zugreift. Die Daten werden als HTML-Dokumente an den Browser ausgeliefert und dort mit Formatierungsregeln optisch gestaltet. Dezierte APIs, wie sie für die wissenschaftliche Datenerhebung genutzt werden, liefern dagegen in der Regel JSON-Dokumente aus, in denen die Daten nicht an der optischen Darstellung orientiert, sondern inhaltlich vorstrukturiert sind (siehe dazu Jünger, 2018). Diese Vorstrukturierung verspricht eine effizientere Datenerhebung, erfordert aber eine Auseinandersetzung mit der Funktionsweise der jeweiligen API. Bereits hiermit wird deutlich, dass APIs anders als häufig dargestellt keinen privilegierten Hintereingang (z. B. Courtois & Mechant, 2014, S. 212) bereitstellen. Auch die Metapher des „wilden Westens“ (Puschmann, 2019) für eine zu wenig restringierte Verwendung von APIs durch Wissenschaftler:innen greift zu kurz, da APIs wie alle Vermittlungsmechanismen immer reguliert sind.

4 Dokumentationen, Berichterstattung und Literatur als Zeugnisse der Bedeutung von APIs

Das Ineinandergreifen der drei datengenerierenden Prozesse bei der Analyse von Onlinekommunikation – Erzeugung durch Nutzende, Vermittlung durch Plattformen, Erhebung durch Wissenschaftler:innen – wird sichtbar, wenn man sich die Unterschiede zwischen verschiedenen Plattformen und die

Entwicklungsschritte jeweils einer Plattform vor Augen führt. Dazu wurden für diesen Beitrag einschlägige Dokumente zu den drei Social-Media-Plattformen Facebook, YouTube und Twitter miteinander verglichen. Die APIs dieser drei Anbietenden werden seit vielen Jahren für sozialwissenschaftliche Analysen verwendet, sodass eine eingehende Untersuchung hier besonders wichtig erscheint. Die Analyse umfasst dabei den Zeitraum seit der Gründung von Facebook im Jahr 2004 bis in das Jahr 2018. Verschiedene Datenquellen wurden für den Vergleich herangezogen: Dokumentationen der APIs, Berichterstattungen über APIs und Sekundärliteratur über APIs (siehe Tabelle 2).

Jede dieser Quellensorten macht andere Aspekte sichtbar. Die Sichtung der API-Dokumentationen erlaubt einen Einblick in die technischen und organisatorischen Rahmenbedingungen, denen sich Wissenschaftler:innen vor allem aus einer repräsentativen Perspektive gegenübergestellt sehen. In den Dokumentationen sind die Parameter des Aufrufs, Zugangsbedingungen und das Format der Ergebnisse verbindlich festgehalten (Janetzko, 2017). Auf den Entwicklerseiten der Plattformen ist mindestens die Dokumentation zum aktuellen Stand der API verfügbar. Für die Rekonstruktion des Zeitverlaufs ist jedoch zusätzlich der Vergleich mit älteren Versionen nötig, Facebook verzeichnet etwa für die Jahre 2010 bis 2017 allein dreizehn Versionen (siehe Abbildung 1). Die älteren Dokumentationen sind nicht immer auf der Seite verfügbar, in diesen Fällen wurden sie über archive.org eingesehen. Mitunter ändern sich in den Dokumentationen nur Details, die dennoch aufschlussreich sein können, da sie Zeugnis über organisationsinterne Entscheidungsprozesse ablegen. Um diese Details sichtbar zu machen, wurde WinMerge (2020) eingesetzt – dieses Programm vergleicht zwei Texte und hebt Änderungen hervor. So lässt sich etwa bei Facebook feststellen, dass im Jahr 2018 die Entwicklung von Bots als Anwendungsszenario in die Richtlinien aufgenommen wird (siehe Abbildung 2: WinMerge macht Änderungen zwischen zwei Versionen eines Dokuments nachvollziehbar, hier am Beispiel der Entwicklerrichtlinien von Facebook).

Ergänzend zu den technischen Dokumentationen treten Entwicklerrichtlinien, in denen die Nutzungsbedingungen festgehalten sind. Aus den darin beschriebenen Nutzungsszenarien lassen sich Hinweise gewinnen, welche Interaktionszusammenhänge die Plattformen für Anwendungen, die auf APIs aufbauen, vorsehen.

Während die Dokumentationen im geschichtswissenschaftlichen Sinne als „Überreste“ der untersuchten sozialen Zusammenhänge gelten, werden Ereignisse

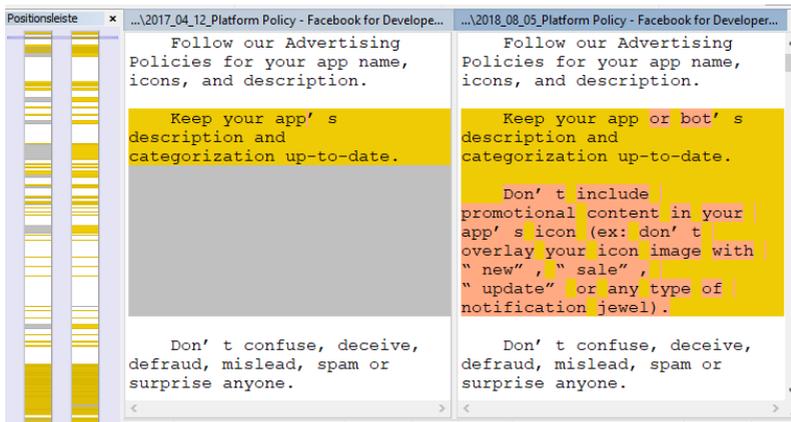
Abbildung 1: Changelog-Archiv für die ersten Jahre der Facebook-API

API Version	Graph API Availability	Marketing API Availability
v2.11	November 7, 2017 - January 30, 2020	November 7, 2017 - Aug 7, 2018
v2.10	July 18, 2017 - November 7, 2019	July 18, 2017 - May 8, 2018
v2.9	April 18, 2017 - July 22, 2019	April 18, 2017 - November 6, 2017
v2.8	October 5, 2016 - April 18, 2019	October 5, 2016 - July 26, 2017
v2.7	July 13, 2016 - October 5, 2018	July 13, 2016 - April 25, 2017
v2.6	April 12, 2016 - July 13, 2018	April 12, 2016 - October 5, 2016
v2.5	October 7, 2015 - April 12, 2018	October 7, 2015 - July 13, 2016
v2.4	July 8, 2015 - October 7, 2017	July 8, 2015 - April 11, 2016
v2.3	March 25, 2015 - July 8, 2017	March 25, 2015 - October 8, 2015
v2.2	October 30, 2014 - March 25, 2017	October 30, 2014 - July 8, 2015
v2.1	August 7, 2014 - October 30, 2016	October 1, 2014 - March 11, 2015
v2.0	October 1, 2014 - August 7, 2016	October 1, 2014 - March 11, 2015
v1.0	April 21, 2010 - April 30, 2015	October 1, 2014 - March 11, 2015

Quelle: <https://developers.facebook.com/docs/graph-api/changelog/archive> (21.03.2020).

im Zusammenhang mit APIs in der journalistischen Berichterstattung und der Sekundärliteratur metakommunikativ eingeordnet. Diese Dokumente können also (im geschichtswissenschaftlichen Sinn) als überlieferte „Tradition“ gelten (Bernheim, 1889, S. 155), die vor allem auch die ökonomischen Zusammenhänge der APIs reflektieren. Die Berichterstattung der zwei größten deutschen IT-Nachrichten anbietenden Heise und Golem wurde deshalb systematisch nach einschlägigen Veröffentlichungen durchsucht. Dazu wurden die Namen der Plattformen mit dem Stichwort API kombiniert (z. B. „facebook api“) und die Teasertexte automatisiert heruntergeladen. Der älteste Eintrag datiert in das Jahr 2005. Die insgesamt 731 Meldungen wurden manuell gesichtet, um alle Ereignisse im Zusammenhang mit Datenzugängen und Programmierschnittstellen zu verschlagworten.

Abbildung 2: WinMerge macht Änderungen zwischen zwei Versionen eines Dokuments nachvollziehbar, hier am Beispiel der Entwicklerrichtlinien von Facebook



Auch die Sekundärliteratur enthält wertvolle Hinweise zur Entwicklung der Schnittstellen. Hier lassen sich wissenschaftliche von außerwissenschaftlichen Quellen unterscheiden, die jeweils über eine Literaturrecherche zum Stichwort „Application Programming Interface“ identifiziert wurden. Die ausgewählten wissenschaftlichen Quellen – hier wurden auch nach 2018 veröffentlichte Quellen berücksichtigt, sofern sie die Bedingungen innerhalb des Untersuchungszeitraums reflektieren – beschäftigen sich vor allem mit den Limitationen automatisierter Datenerhebung und mit den Einschränkungen der APIs, die sich aus der Einführung von Begutachtungsverfahren durch die Online-Plattformen ergeben (siehe Kapitel 5). Außerwissenschaftliche Quellen wurden für die Rekonstruktion der Chronologie von API-Entwicklungen herangezogen. Als Anlaufstelle für aktuelle Entwicklungen gilt beispielsweise ProgrammableWeb, das sich selbst als „the world’s leading source of news and information about Internet-based application programming interfaces“ beschreibt (Programmable Web, 2020). Zudem fließen eigene Erfahrungen bei der Entwicklung und Anwendung von Facepager – eine Software zur automatisierten Datenerhebung über APIs (Jünger & Keyling, 2020) – in die Analyse ein.

Alle Quellen insbesondere im zeitlichen Umfeld dieser Ereignisse wurden gesichtet und verglichen. Damit folgt die Analyse dem Selbstverständnis der

komparativen Methode der qualitativen Analyse (Glaser, 2008). Die Textstellen wurden schließlich zu drei Kategorien verdichtet, die sich auf die drei verschiedenen Perspektiven – repräsentative, interaktionale und ökologische – beziehen.

5 Zugangsverfahren, Nutzungsszenarien und Organisationsweisen

Je nach Perspektive geraten nicht nur unterschiedliche Aspekte, sondern auch unterschiedliche Quellen über APIs stärker in den Blick. Geht es um die Frage, inwiefern sich über APIs reliable und valide Messungen bewerkstelligen lassen, so sind die in den Dokumentationen beschriebenen Zugangsbeschränkungen besonders relevant. Um sich dagegen vor Augen zu führen, welche Rolle APIs für die Interaktion zwischen Nutzenden spielen, sind die von den API-Betreibern angepriesenen Nutzungsszenarien aufschlussreich. Die Einbettung in das Ökosystem wird dagegen vorrangig über Berichterstattung und Sekundärliteratur sichtbar, wenn die Partnerschaften und Konkurrenzbeziehungen zwischen verschiedenen Akteuren thematisiert werden. Im Folgenden werden die untersuchten Dokumente entlang dieser drei Perspektiven zusammengefasst, um die sich jeweils daraus ergebenden Rahmenbedingungen für wissenschaftliche Analysen einzuordnen.

5.1 APIs aus repräsentativer Perspektive

Typischerweise werden APIs im wissenschaftlichen Kontext als Werkzeuge zur Datenerhebung eingesetzt. Aus dieser Sicht geht es darum, die Inhalte der Kommunikation und das Verhalten von Nutzenden strukturiert, standardisiert und effizient zu erfassen.³ Vorausgesetzt werden muss deshalb, dass die über eine Programmierschnittstelle erhobenen Daten ausreichend vollständig oder zumindest repräsentativ für die Aktivitäten auf der Plattform sind. Eine Reihe von Eigenarten spricht zunächst gegen diese Annahme, denn sowohl die Anzahl der

3 Strukturiert: Im Gegensatz dazu stehen etwa Interviews, die zunächst transkribiert und dann kodiert werden müssen. API-Daten enthalten bereits schriftlichen Text und formale Merkmale wie die Anzahl der Likes. Standardisiert: Für alle Untersuchungsobjekte wird das gleiche Erhebungsverfahren angewendet. Effizient: Es werden mehr Daten in kürzerer Zeit erhoben als über eine manuelle Erhebung möglich.

Tabelle 2: Quellen für die Analyse von APIs

Dokumentationen	Ältester berücksichtigter Eintrag
Twitter	20.09.2006
https://developer.twitter.com/ https://blog.twitter.com/developer/ https://twitter.com/twitterapi https://twitter.com/t/help/api https://help.twitter.com/en/rules-and-policies https://developer.twitter.com/en/developer-terms/agreement-and-policy	
YouTube	04.06.2007
http://code.google.com/apis/youtube/reference.html https://developers.google.com/youtube/ https://www.youtube.com/intl/de/yt/dev/api-resources/ https://developers.google.com/youtube/terms/api-services-terms-of-service https://www.youtube.com/t/terms	
Facebook	21.08.2006
https://developers.facebook.com/docs https://developers.facebook.com/blog https://developers.facebook.com/policy/ https://www.facebook.com/apps/site_scraping_tos_terms.php	

Berichterstattung		
Twitter	115 Artikel bei Golem 118 Artikel bei Heise	18.07.2007
YouTube	66 Artikel bei Golem 79 Artikel bei Heise	15.08.2005
Facebook	156 Artikel bei Golem 197 Artikel bei Heise	27.05.2007
Sekundärliteratur		
Wissenschaftlich	Bruns, 2019; Burgess & Bruns, 2012; Courtois & Mechant, 2014; Cukier & Mayer-Schoenberger, 2013; Freelon, 2018; Gerlitz & Rieder, 2013; Gerlitz & Rieder, 2018; Hargittai, 2020; Janetzko, 2017; Jünger, 2018; Puschmann, 2019; Vis, 2013; Ruths & Pfeffer, 2014	2012
Außerwissenschaftlich	Cater, 2013; Doerrfeld, 2016; Jacobson, Brail & Woods, 2012; Lane, 2012; Pro-grammable Web, 2020	2012

Datensätze als auch die Inhalte sind beim Zugriff über APIs in der Regel trunziert, das heißt auf einen bestimmten Umfang beschnitten.

- *Trunkierungen der Zugänge:* Zunächst ist festzustellen, dass einige Inhalte und Funktionen überhaupt nicht über APIs zugänglich sind, weil dafür keine Endpunkte zur Verfügung stehen. So lassen sich auf Twitter die Antworten auf Tweets nicht ohne Weiteres abfragen: „you can not request a list of replies to another user whether public or protected“ (Twitter, 2007). Diese Daten sind lediglich auf der Webseite oder über die Smartphone-App zugänglich. Bei Facebook sind vor allem Zugänge über Suchbegriffe im Lauf der Zeit restriktiver geworden, sodass zum Beispiel über die API nicht mehr gezielt nach Veranstaltungen, Gruppen oder Nutzenden gesucht werden kann (Facebook, 2020a). Ohnehin war ein Zugriff auf die erweiterten Suchfunktionen (Facebook Graph Search, z. B. „Socialists who work nearby Greifswald and like McDonalds“) soweit ersichtlich nie über die API möglich und wurde mittlerweile auch außerhalb der API deutlich abgebaut (Cox, 2019). Einige Daten sind zudem an persönliche Zugänge gebunden, das betrifft zum Beispiel Messenger-Dienste bzw. Privatnachrichten. Ein Zugriff ist nur bei Autorisierung im Einzelfall vorgesehen. Die Anbietenden unterscheiden dafür in der Regel verschiedene Zugriffsbereiche („scopes“), die zunehmend restriktiver gehandhabt und nur nach einer Prüfung freigeschaltet werden (z. B. YouTube, 2019). Darüber hinaus ist davon auszugehen, dass in den Datenbanken der Betreibenden viele weitere Daten ohne Anbindung an eine externe API und auch ohne eine Darstellung in der Weboberfläche liegen. Vor allem nichttaggregierte Daten zu aufgerufenen (views) oder gemeldeten (flagged) Beiträgen, das heißt vor allem Daten über die Rezeption von Beiträgen, sind nicht allgemein zugänglich.
- *Trunkierungen des Datenumfangs:* Die Anzahl der Datensätze ist, abhängig vom Endpunkt der API, nicht immer vollständig. Die Tweets von Nutzenden sind beispielsweise auf die letzten 3.200 beschränkt (Twitter, 2020b). In der kostenlosen Version der Twitter-API ist die Suche zudem auf die letzten sieben Tage beschränkt (Twitter, 2020c). Bei Facebook-Seiten lassen sich für eine Seite pro Jahr maximal 600 Posts erheben, anders als bei Twitter sind aber die Auswahlkriterien nicht ausreichend nachvollziehbar dokumentiert („in einer Rangfolge sortierte veröffentlichte Beiträge“; Facebook, 2020c). Bei

YouTube sind die Suchergebnisse für nicht-eigene Channels auf 500 Treffer beschränkt (YouTube, 2020). In den APIs wirkt eine Vielzahl weiterer Beschränkungen, deren Konsequenzen für die Vollständigkeit von Fall zu Fall unterschiedlich ausfallen. Beispielsweise hängen die abrufbaren Daten von den Privatsphäreinstellungen der Nutzenden und damit von der Beziehung zwischen den Nutzenden der Plattform und den Nutzenden der API ab.

- *Trunkierung der Inhalte:* Eng verbunden mit der Trunkierung der Zugänge sind schließlich die Einschränkungen, welche Datenfelder zugänglich sind. Während Twitter auf einigen Endpunkten vollständige Tweets mit allen dazugehörigen Feldern und eingebetteten Datenfeldern für Nutzerinformationen, Hashtags und URLs zurückgibt („fully hydrated“), hat Facebook das Verfahren mittlerweile so umgestellt, dass jedes gewünschte Feld bereits bei der Abfrage spezifiziert werden muss („fields“-Parameter). YouTube beschreitet schon lange einen Zwischenweg: Hier wird über einen Parameter („parts“) immer ein ganzes Bündel von Feldern gesteuert, sofern die entsprechenden Zugriffsrechte erteilt wurden. Im Fall von Facebook haben die jüngsten Änderungen der API unter anderem dazu geführt, dass die Autor:innennamen der Nutzerkommentare auf öffentlichen Seiten nur noch von den Seiteninhabenden erfasst werden können. Bemerkenswert an dieser Änderung ist, dass sie als so genannte funktionsgefährdende Änderung gekennzeichnet wurde. Diese Änderungen brechen den Vertrag zwischen API-Konsument:innen und API-Betreibenden („breaking changes“), da sie auch ohne langfristige Ankündigung rückwirkend für alle Versionen einer API greifen und so die Anpassung bereits bestehender Anwendungen notwendig machen (Facebook, 2017 & 2020a). Insgesamt greifen bei allen Anbietenden drei Mechanismen ineinander, die zu einer Trunkierung von Inhalt führen: Inhalte können mehr oder weniger vorstrukturiert, mehr oder weniger aggregiert sowie mehr oder weniger anonymisiert sein.

Diese Einschränkungen müssen allerdings nicht zwangsläufig als Verzerrung oder Datenausfall bewertet werden. Denn wenn die erhobenen Daten repräsentativ für die Aktivität von Nutzenden sein sollen, müsste zunächst ein entsprechender Maßstab definiert werden. Das allerdings ist kaum möglich, denn aus Sicht einzelner Nutzenden sind in der konkreten Nutzungssituation nur ausgewählte Inhalte auf dem Bildschirm sichtbar. Benachrichtigungen auf dem

Startbildschirm des Smartphones sehen anders aus als die Ergebnisse einer gezielten Suche über die Weboberfläche. Zudem scheint die Nutzung von Plattformen stark an der Gegenwart ausgerichtet zu sein. Der systematische Zugriff auf alte Facebook-Posts stößt schnell an die Grenzen des Scrollbalkens, sofern Nutzenden diese Möglichkeit überhaupt verwenden. Gleichzeitig wird Aufmerksamkeit durch die Empfehlungssysteme („Algorithmen“) der Betreibenden gelenkt. Über die APIs ist nicht nachvollziehbar, ob und welche Inhalte wie prominent in wessen News Feeds eingespielt wurden. Facebook und andere Anbietende spielen zudem regelmäßig auch ältere Posts ein und zeigen so auf, welche Aktivitäten man selbst vor einem Jahr auf der Plattform vollzogen hat. Indem diese Bedingungen auf das Mitteilungsverhalten von Nutzenden zurückwirken, lässt sich nicht feststellen, wofür die zu einem bestimmten Zeitpunkt für Nutzende auf einer Plattform sichtbaren oder durch Nutzende produzierten Inhalte stehen – Mechanismen der Plattform, Intentionen von Nutzenden oder vielleicht sogar Sinnsetzungen von Wissenschaftler:innen?

Insbesondere der Faktor Zeit stellt die Definition von Untersuchungszeiträumen vor nicht zu lösende Herausforderungen. Die im ersten Moment naheliegende Idee, einen Untersuchungszeitraum oder einen Zeitpunkt (Schnappschuss) abzubilden, muss bereits an der Erhebungstechnik scheitern. Denn während der Datenerhebung verstreicht zwangsläufig Zeit, innerhalb derer Nutzenden auf der Plattform weiter aktiv sind. Deutlich wird dies beispielweise an den Kommentar-, Like- oder Followerzahlen. Erhebt man, sofern über die API möglich, zunächst die aggregierte Anzahl von Kommentaren und anschließend die einzelnen Kommentare, können in der Zwischenzeit Kommentare gelöscht, überarbeitet oder hinzugekommen sein. Insbesondere umfangreiche Erhebungen, wie sie aus einer repräsentativen Perspektive zur Abbildung der Gesamtstruktur nötig werden, laufen schnell in die Begrenzung der Zugriffe pro Zeitfenster („rate limits“) und müssen ggf. über Tage gestreckt werden. Besonders restriktiv ist mit 15 Anfragen je 15 Minuten und einer Maximalzahl von 200 Datensätzen je Anfrage der direkte Zugriff auf die Followerlisten von Twitter-Nutzenden, sodass pro Minute Daten über maximal 200 Follower erhoben werden können (Twitter, 2020d).

Letztendlich bieten APIs also nur eine von vielen möglichen Repräsentationen der Nutzeraktivität. Eine repräsentative Analyse im engeren Sinn ist nicht möglich, denn die Rede von Datenausfällen oder Verzerrungen idealisiert zwangsläufig den datengenerierenden Prozess. Zu diesem Prozess gehören das

Löschen von Inhalten, Einschränkungen der Privatsphäre und die heterogenen Nutzungskontexte von Nutzenden genauso dazu wie die sich stets ändernden technischen und organisatorischen Zugangsbedingungen und die Empfehlungssysteme der Plattformen.

5.2 APIs aus interaktionaler Perspektive

Bei der Sichtung der Dokumentationen wird deutlich, dass APIs die Nutzung der Dienste ermöglichen und vor allem aber auch fördern sollen. APIs haben damit Konsequenzen für die Interaktion zwischen Nutzenden der Plattform. Die Nutzungsszenarien der drei Plattformen umfassen im Wesentlichen drei Kategorien:

- *Inhalte verwalten:* Twitter nennt als Ziele der API unter anderem „publish“, „curate“ und „engage“ (Twitter, 2019). Mit der API lassen sich Inhalte von Twitter in andere Anwendungen einbetten, gleichzeitig aber auch neue Inhalte erstellen und moderieren. Auch YouTube legt von Anfang an den Fokus auf die Verbreitung von Inhalten über Webseiten und Anwendungen von Dritten: „Do you wish to integrate online videos into your application? With the set of YouTube XML APIs, developers can integrate applications directly into YouTube’s growing repository of videos“ (YouTube, 2005). Ein wichtiges Anwendungsfeld ist zudem die Verhinderung von Urheberrechtsverletzungen. Rechteinhaber:innen können über die Content-ID-API Videos hochladen, die dann automatisch mit den Inhalten der Plattform abgeglichen werden. Die Besonderheit bei Facebook wiederum ist, dass Spiele und andere Anwendungen direkt auf der Plattform im Kontext der Facebook-Seite ausgeführt werden. Während die API zu Beginn lediglich Lesezugriff erlaubte, ist mittlerweile selbst das automatisierte Posten mit Bots ein Teil der Nutzungsszenarien. Auch die Einbindung von Buttons zum „Liken“ und „Teilen“ wird über APIs umgesetzt.
- *Marketing und Analyse:* Ein weiteres Feld ist die Analyse des Nutzerverhaltens und die Analyse von Werbekampagnen. Auffallend ist hierbei die direkte Adressierung von Unternehmen mit spezifischen Marketing-APIs (z. B. Facebook, 2015 und YouTube, 2013). Unternehmen können darüber die eigenen Inhalte analysieren und aggregierte Zielgruppenanalysen durchführen.

Hier wird deutlich, dass neben Individuen auch korporative Akteure zu den Nutzenden der Plattformen zu zählen sind. Für wissenschaftliche Fragestellungen erscheinen die spezifischen Marketingzugänge weniger geeignet zu sein, da Wissenschaftler:innen vermutlich eher selten eigene Inhalte analysieren wollen und die angebotsübergreifenden Daten stark aggregiert sind. Wissenschaftliche Datenerhebungen nutzen vielmehr den Nebeneffekt des oben beschriebenen Szenarios aus: um Plattforminhalte wie etwa YouTube-Videos in Webangebote außerhalb der Plattformen einbinden zu können, ist eine Schnittstelle notwendig. Allerdings sind hier in jüngster Zeit Veränderungen zu verzeichnen, im Januar 2020 ist etwa bei Twitter auf der Webseite der Anwendungsfall „Twitter data for academic research“ (Twitter, 2020a) hinzugetreten.

- *Authentifizierung und Infrastruktur*: Alle drei Plattformen agieren als *identity provider*, indem sie anderen Diensten Login-Möglichkeiten bereitstellen. Die als „Facebook Login“ (Facebook, 2013) und ähnlich betitelten Funktionen ersparen es Nutzenden, sich bei vielen einzelnen Webseiten zu registrieren. Durch Dritte betriebene Webangebote – beispielsweise Online-Shops – erhalten vermittels der APIs Zugriff auf das Konto der Nutzenden. Auch für die Plattformen selbst zahlt sich diese Dienstleistung aus; durch die Logins können sie das Nutzungsverhalten außerhalb der eigenen Plattform erfassen. Während dies bei Twitter die im engeren Sinn einzige vom Kerngeschäft unabhängige Infrastrukturdienstleistung ist, bietet Facebook beispielsweise Softwarebibliotheken für Künstliche Intelligenz an. Google (als Betreiber von YouTube) ist ohnehin aktuell einer der größten Infrastrukturdienstleister und bietet eine Vielzahl an APIs beispielsweise zur Spracherkennung oder für das Hosting von Daten an.

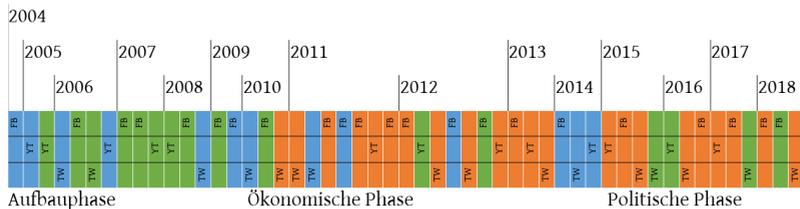
Über die Zeit haben sich die Dienstleistungen der Plattformen und damit auch die Programmierschnittstellen immer weiter ausdifferenziert. Lediglich die Schnittstellen von Twitter sind über lange Zeiträume weitgehend unverändert und stabil geblieben. Für die wissenschaftliche Datenanalyse ist aus diesen Beobachtungen vor allem die Konsequenz zu ziehen, dass Nutzenden- und Plattformverhalten vielfältig miteinander verwoben sind. Programmierschnittstellen tragen ihren Teil dazu bei, Nutzende zu Kommunikation und Interaktion anzuregen, indem sie von Unternehmen zur Moderation von Inhalten, für Marketingkampagnen und

zur Vereinfachung beim Anlegen von Nutzerkonten eingesetzt werden. Über APIs lassen sich etwa Inhalte von Nachrichtenseiten automatisiert in Social-Media-Kanäle einspielen oder es können in die Onlineplattformen eingebettete Chat-Bots realisiert werden. APIs erscheinen aus dieser Perspektive weniger als Mittel der Datenerhebung und -analyse, vielmehr werden sie zur Gestaltung von Kommunikation und damit als ein Mittel der Datafizierung von Verhalten eingesetzt.

5.3 APIs aus ökologischer Perspektive

Die zeitliche Entwicklung der APIs verdeutlicht, wie sich um Programmierschnittstellen ein Ökosystem aus Nutzenden, Plattformen und Drittanbietenden herausgebildet hat. Bereits kurz nach ihrer Gründung etablierten Facebook, YouTube und Twitter Schnittstellen und profitieren in der Zeit des „Web 2.0“ (O'Reilly, 2005) davon, dass andere Webangebote ihre Dienste einbinden und so deren Bekanntheit steigern. Während zu Beginn vor allem Erweiterungen der Schnittstellen zu verzeichnen sind, kommt es ungefähr um das Jahr 2010 vermehrt zu Einschränkungen (siehe Abbildung 3: Abfolge von Schlüsselereignissen der Entwicklung von APIs bei Facebook (FB), YouTube (YT) und Twitter (TW)). Nachdem unzählige Drittanbietende eigene Dienste auf den Schnittstellen aufgebaut hatten, erobern sich die Anbietenden ihren Einfluss zurück (ausführlich siehe Jünger, 2021). Besonders deutlich wird diese Vorgehensweise bei Twitter. Bereits im Jahr 2008 akquiriert Twitter den Dienst Summize und integriert dessen Suchfunktionalität in die API (Arrington, 2008; Ihlenfeld, 2008). Im gleichen Jahr wird Tweetie, die erste Twitter-App für das iPhone, aufgekauft und später unter eigener Marke wiederveröffentlicht (Barczok, 2010). Damit einher gehen Begrenzungen der maximalen Zugriffe auf die API, sodass die Entwicklung und Verbreitung weiterer Clients durch Drittanbietende effektiv beschnitten wird. Neben Clients wurden viele Funktionalitäten wie URL-Kürzung (TinyURL) oder Analysedienste (Gnip) zunächst von Drittanbietenden entwickelt und dann integriert oder nachgebaut und verdrängt. Auch über die Nutzungsbedingungen der Dienste wird mittlerweile Konkurrenz direkt oder indirekt untersagt, wenn es zum Beispiel bei Facebook heißt: „Don't use the Ads API if you're an ad network or data broker“ (Facebook, 2019).

Abbildung 3: Abfolge von Schlüsselereignissen der Entwicklung von APIs bei Facebook (FB), YouTube (YT) und Twitter (TW)



Blau: Gründungen, Akquisitionen, Produkte; Grün: Erweiterungen der APIs; Orange: Einschränkungen der APIs. Quelle: eigene Dokumentenanalyse.

Auch die Plattformanbietenden stehen untereinander in komplexen Beziehungen. So blockierte Twitter zeitweise im Juli 2012 für Instagram den API-Zugang, nachdem es von Facebook übernommen wurde (Hedemann, 2012). Instagram wiederum blockierte die Anzeige seiner Bilder im Twitter-Feed. Wenig später implementierte Twitter in die eigene App die von Instagram bekannten Bildfilter, was auf heise.de mit der Überschrift „Twitter baut Instagram nach“ kommentiert wurde (Kuri, 2012). Doch nicht nur Konkurrenz prägt das Ökosystem, sondern auch Partnerschaften und Allianzen: „Finally, Twitter signed partnership deals with Google, Yahoo, and Microsoft to include tweets in the results of their respective search products” (van der Mersch, 2016). Auch bei Facebook wird der Zugang zur Public Feed API nur privilegierten Partner:innen gewährt: „Access to the Public Feed API is restricted to a limited set of media publishers and usage requires prior approval by Facebook. You cannot apply to use the API at this time” (Facebook, 2014).

Auf die Phase der ökonomischen Eroberung folgt eine Phase, in der bis heute die politische Bedeutung der Plattformen diskutiert wird. Besonders disruptiv waren hier die Vorgänge um Cambridge Analytica. Das Unternehmen hatte auf Grundlage von Daten, die insbesondere über die API von Facebook erhoben wurden, im US-amerikanischen Wahlkampf psychometrische Wahlkampfunterstützung angeboten. Erste Artikel darüber wurden bereits 2015 im Guardian veröffentlicht (Davies, 2015). Darauf reagierte Facebook schnell mit Zugriffsbeschränkungen, sodass etwa der Zugriff auf Daten von Freunden und Gruppen

nicht mehr möglich war, eine ausführliche Berichterstattung blieb allerdings zunächst aus. Erst drei Jahre später, im Jahr 2018, gelang es den Medien erneut, Cambridge Analytica in den Fokus der Berichterstattung und der Politik zu rücken. Dieses Mal war der politische Druck größer und die Konsequenzen auf Seiten der Plattformen fielen schärfer aus (Cadwalladr & Graham-Harrison, 2018; Rosenberg, Confessore & Cadwalladr, 2018). Alle drei Plattformen haben mittlerweile präventive App Reviews eingeführt, sodass bereits vor dem ersten Zugriff auf APIs ein Begutachtungsprozess durchlaufen werden muss.

Um die Bedeutung von APIs als Mittel der Datenintermediation zu erfassen, ist ein Vergleich mit Webscraping hilfreich. Als Webscraping wird eine Technik bezeichnet, bei der Daten direkt aus den Webseiten der Plattformen extrahiert, anstatt über dezidierte APIs erhoben werden. Allerdings wird diese Vorgehensweise von den Anbietenden geächtet. Das ist vor allem vor dem Hintergrund der offenen Architektur des Web bemerkenswert. Die automatische Erschließung von Webseiten durch Suchmaschinen kann als ein wesentlicher Bestandteil des Web begriffen werden, ohne den die aktuelle Bedeutung des Internet kaum vorstellbar wäre. Webseitenbetreibende können jedoch in der Datei robots.txt Regeln formulieren, deren Einhaltung sie von anderen Organisationen und deren Bots erwarten, wenn auch technisch kaum durchsetzen können. Facebook schließt Crawling darüber komplett aus: „Notice: Crawling Facebook is prohibited unless you have express written permission“ (Facebook, 2020b). Es folgt eine Liste von Bots, mit deren Urhebern anscheinend Vereinbarungen getroffen wurden (Applebot, baiduspider, Bingbot etc.). Die Liste endet damit, dass alle anderen Bots ausgeschlossen werden: „User-agent: * Disallow: /“ (Facebook, 2020b).

APIs können vor diesem Hintergrund als ein Mittel der Einflussicherung verstanden werden, da Zugänge individuell kontrollierbar sind. Auch Wissenschaftler:innen sind in diesem Feld nur (wenig privilegierte) Akteure. Die Folgen von Cambridge Analytica sind in der Wissenschaft als „APIcalypse“ (Bruns, 2019) bezeichnet worden, weil der wissenschaftliche Zugang vor allem bei Facebook nahezu verunmöglicht wurde. Gleichzeitig wurden von den Plattformen verschiedene kuratierte Datenzugänge für wissenschaftliche Zwecke eingerichtet. Besonders prominent ist die von einem wissenschaftlichen Board geleitete Initiative Social Science One. Dass es hier dennoch nicht uneingeschränkt um unabhängige Forschung geht, wird am Modus der Datenzugänge deutlich: Für verschiedene Datensätze (z. B. geteilte URLs) werden Ausschreibungen veröffent-

licht, auf die sich Wissenschaftler:innen bewerben können. Der inhaltliche Fokus ist zwar breit, aber durchaus festgelegt, etwa für den Zugang zu Facebook-Daten: „effect of social media on elections and democracy“ (Social Science One, 2020). Ausschreibungen sind in der Wissenschaft durchaus üblich, in jedem Fall ist aber ein Steuerungseffekt anzunehmen. Damit begeben sich Plattformen in die Rolle von Datenintermediatoren für Wissenschaftler:innen. Das daraus hervorgehende soziale System darf deshalb nicht nur als Forschungsmittel verstanden, sondern muss auch als Forschungsgegenstand untersucht werden.

6 Fazit – Die Mediatisierung der Wissenschaft

Ausgangspunkt des Beitrags war die Unterscheidung von drei Perspektiven, aus denen Programmierschnittstellen (APIs) von Online-Plattformen betrachtet werden können. Während APIs aus repräsentativer Perspektive Nutzerverhalten abbilden und aus interaktionaler Perspektive Verhalten ermöglichen, sind sie aus ökologischer Perspektive ein Mittel der Einflussicherung. In den für wissenschaftliche Zwecke über APIs erhobenen Daten spiegeln sich dementsprechend drei miteinander verwobene Prozesse wieder, die Aktivitäten der Nutzenden, die Vermittlung der Plattformen und die Erhebungsbemühungen von Wissenschaftler:innen. Daran anschließend wurde die Frage gestellt, inwiefern eine historisch-vergleichende Perspektive auf APIs Erkenntnisse über diese drei miteinander verwobenen, datengenerierenden Prozesse hervorbringt.

Die Beschäftigung mit den Unterschieden und Gemeinsamkeiten zwischen verschiedenen Programmierschnittstellen zeigt auf, dass vollständige oder repräsentative Datenerhebungen aufgrund der Zugangsmechanismen der Plattformen kaum realisierbar sind. Das liegt zum einen daran, dass wissenschaftliche Analysen bislang nicht zu den zentralen Nutzungsszenarien der Onlineplattformen gehören, ist aber auch dadurch verkompliziert, dass die Definition von Grundgesamtheiten an Grenzen stößt. Dagegen wird bei einer Betrachtung im Zeitverlauf besonders deutlich, dass APIs Teil einer dynamischen sozialen Wirklichkeit sind – sie bilden nicht ab, sondern sind Schnittstellen sowohl für die Kommunikation zwischen Nutzenden als auch für strategische Partnerschaften zwischen Unternehmen.

Die besondere Bedeutung der Vermittlungsebene ist in der Kommunikationswissenschaft in der Tradition der Mediatisierungsforschung schon lange erkannt

worden: “the format of the ITs [information technologies; JJ] [...] structure the connections of people to one another and to the activity” (Altheide, 1994, S. 670). Auch in Bezug auf Programmierschnittstellen bedeutet dies eine unauflösbare Verwobenheit von Nutzenden und Plattformen. Nun gibt es keinen Grund davon auszugehen, dass Wissenschaftler:innen im Vergleich zu anderen Akteuren in einer erkenntnistheoretisch oder handlungspraktisch überlegenen Position stehen. Dementsprechend ist davon auszugehen, dass sich der Entstehungsprozess der Daten in den Daten widerspiegelt: „the message reflects the process by which it was constituted“ (Altheide, 1994, S. 665). Das hat starke erkenntnistheoretische Konsequenzen, denn es betrifft die Frage, was eigentlich bei der Verwendung von APIs erforscht wird. Je nach Perspektive wird die Antwort anders ausfallen: Verhalten auf Plattformen, Verhalten von Plattformen und gleichzeitig auch das Verhalten von Wissenschaftler:innen. Man muss also davon ausgehen, dass nicht nur die Nutzung von Plattformen einer Mediatisierung unterliegt, vielmehr ist spätestens bei der Datenerhebung über APIs auch die Wissenschaft mediatisiert.

Jun.-Prof. Dr. Jakob Jünger ist Juniorprofessor für Kommunikationswissenschaft mit dem Schwerpunkt Digital Media and Computational Methods an der Universität Münster, jakob.juenger@uni-muenster.de

Quellenverzeichnis

- Altheide, D. L. (1994). An Ecology of Communication: Toward a Mapping of the Effective Environment. *The Sociological Quarterly*, 35(4), 665–683. <https://doi.org/10.1111/j.1533-8525.1994.tb00422.x>
- Arrington, M. (15.7.2008). Interview With Evan Williams: Summize Acquisition, API Issues And Their Revenue Model. *TechCrunch*. Abgerufen von <https://techcrunch.com/2008/07/15/interview-with-ewan-william-summize-acquisition-api-issues-and-their-revenue-model/>
- Barczok, A. (10.4.2010). Twitter will kostenlosen iPhone-Client anbieten. *Heise*. Abgerufen von <https://www.heise.de/newsticker/meldung/Twitter-will-kostenlosen-iPhone-Client-anbieten-974768.html>
- Bernheim, E. (1889). *Lehrbuch der historischen Methode: Mit Nachweis der wichtigsten Quellen und Hilfsmittel zum Studium der Geschichte*. Leipzig: Duncker & Humblot.

- Bick, W., & Müller, P. J. (1984). Sozialwissenschaftliche Datenkunde für prozeßproduzierte Daten: Entstehungsbedingungen und Indikatorenqualität. In W. Bick, R. Mann, & P. J. Müller (Hrsg.), *Sozialforschung und Verwaltungsdaten* (S. 123–159). Stuttgart: Klett-Cotta.
- Bruns, A. (2019). After the ‘APICalypse’: social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11), 1544–1566. <https://doi.org/10.1080/1369118X.2019.1637447>
- Bruns, A., & Enli, G. (2018). The Norwegian Twittersphere. *Nordicom Review*, 39(1), 129–148. <https://doi.org/10.2478/nor-2018-0006>
- Bruns, A., & Moon, B. (2019). One Day in the Life of a National Twittersphere. *Nordicom Review*, 40(s1), 11–30. <https://doi.org/10.2478/nor-2019-0011>
- Burgess, J., & Bruns, A. (2012). Twitter Archives and the Challenges of “Big Social Data” for Media and Communication Research. *Journal of Media and Culture*, 15(5). <http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/561>
- Cadwalladr, C. & Graham-Harrison, E. (17.3.2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. Abgerufen von <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Cater, M. (2016). A Brief History of API-Based Web Applications. Abgerufen von <https://smartbear.de/blog/test-and-monitor/a-brief-history-of-api-based-web-applications/>
- Couldry, N., & Hepp, A. (2017). *The mediated construction of reality*. Cambridge: Polity.
- Courtois, C., & Mechant, P. (2014). An evaluation of the potential of Web 2.0 APIs for social research. In G. Patriarche, H. Bilandzic, J. L. Jensen, & J. Jurišić (Hrsg.), *Audience research methodologies: Between innovation and consolidation* (pp. 212–224). New York, London: Routledge.
- Cox, J. (10.6.2019). Facebook Quietly Changes Search Tool Used by Investigators, Abused By Companies. *Vice*. Abgerufen von https://www.vice.com/en_us/article/zmpgmx/facebook-stops-graph-search
- Cukier, K., & Mayer-Schoenberger, V. (2013). The Rise of Big Data: How It’s Changing the Way We Think About the World. *Foreign Affairs*, 92(3), 28–40.

- Davies, H. (11.12.2015). Ted Cruz using firm that harvested data on millions of unwitting Facebook users. *The Guardian*. Abgerufen von <https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data>
- DMI Tools (2020). <https://wiki.digitalmethods.net/Dmi/ToolDatabase> (21.03.2020).
- Dogu, B., & Mat, H. O. (2019). Who Sets the Agenda? Polarization and Issue Ownership in Turkey's Political Twittersphere. *International Journal of Communication*, 13, 229–250.
- Doerrfeld, B., Sandoval, K., Wood, C., Lauret, A., & Anthony, A. (2016). The API economy: Disruption and the Business of APIs: Nordic APIs. <https://nordicapis.com/api-ebooks/the-api-economy/>
- Dorfer, L. (2016). Datenzentrische Geschäftsmodelle als neuer Geschäftsmodelltypus in der Electronic-Business-Forschung: Konzeptionelle Bezugspunkte, Klassifikation und Geschäftsmodellarchitektur. *Schmalenbachs Zeitschrift Für Betriebswirtschaftliche Forschung*, 68(3), 307–369. <https://doi.org/10.1007/s41471-016-0014-9>
- Facebook (2013). Facebook Login Overview. *Internet Archive*. <https://web.archive.org/web/20130624010259/https://developers.facebook.com/docs/facebook-login/overview/>
- Facebook (2014). Public Feed API. *Internet Archive*. [https://web.archive.org/web/20140406172840/https://developers.facebook.com/docs/public_feed/\[archive.org\]](https://web.archive.org/web/20140406172840/https://developers.facebook.com/docs/public_feed/[archive.org])
- Facebook (2015) Marketing API. *Internet Archive*. <https://web.archive.org/web/20150316002553/http://developers.facebook.com/docs/marketing-apis>
- Facebook (2017). Version 2.11. Funktionsgefährdende Änderungen nach 90 Tagen. <https://developers.facebook.com/docs/graph-api/changelog/version2.11#gapi-90-pages>
- Facebook (2019). Facebook Platform Policy. *Internet Archive*. [https://web.archive.org/web/20190801212325/https://developers.facebook.com/policy/\[archive.org\]](https://web.archive.org/web/20190801212325/https://developers.facebook.com/policy/[archive.org])
- Facebook (2020a). Breaking Changes. <https://developers.facebook.com/docs/graph-api/changelog/breaking-changes>
- Facebook (2020b). Robots.txt. <https://facebook.com/robots.txt> (20.03.2020).
- Facebook (2020c). Page Feed. <https://developers.facebook.com/docs/graph-api/reference/v6.0/page/feed>

- Facebook (2021). Pages API. <https://developers.facebook.com/docs/pages/20.1.2021>.
- Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>
- Gerlitz, C., & Rieder, B. (2013). Mining one percent of Twitter: collections, baselines, sampling. *M/C Journal*, 2(16), 1–18.
- Gerlitz, C., & Rieder, B. (2018). Tweets Are Not Created Equal: Investigating Twitter's Client Ecosystem. *International Journal of Communication*, 12, 528–547.
- Glaser, B. G. (2008). The Constant Comparative Method of Qualitative Analysis. *Grounded Theory Review*, 7(3). Abgerufen von <http://groundedtheoryreview.com/2008/11/29/the-constant-comparative-method-of-qualitative-analysis-1>
- Haim, M., Kümpel, A. S., & Brosius, H.-B. (2018). Popularity cues in online media: A review of conceptualizations, operationalizations, and general effects. *Studies in Communication | Media*, 7(2), 186–207. <https://doi.org/10.5771/2192-4007-2018-2-58>
- Hargittai, E. (2020). Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review*, 38(1), 10–24. <https://doi.org/10.1177/0894439318788322>
- Hautzner, L., Lünich, M., & Rössler, P. (2012). *Social Navigation: Neue Orientierungsmuster bei der Mediennutzung im Internet*. Baden-Baden: Nomos.
- Hedemann, F. (27.07.2012). Twitter auf Isolationskurs: Nach LinkedIn folgt jetzt Instagram. *t3n*. Abgerufen von <https://t3n.de/news/twitter-isolationskurs-linkedin-404538/>.
- HowdyApps (2020). Unfollow For Twitter Free [Software]. <https://play.google.com/store/apps/details?id=follow.unfollow.tool> (22.01.2021)
- Ihlenfeld, J. (16.7.2008). Twitter kauft Summize. *Golem*. Abgerufen von <https://www.golem.de/0807/61084.html>
- Jacobson, D., Brail, G., & Woods, D. (2012). *APIs: A strategy guide: Creating channels with application programming interfaces*. Beijing: O'Reilly.
- Janetzko, D. (2017). The role of APIs in data sampling from Social Media. In L. Sloan & A. Quan-Haase (Hrsg.), *The SAGE handbook of social media research methods* (S. 146–160). Los Angeles: SAGE reference. <https://doi.org/10.4135/9781473983847.n10>

- Jensen, K. B. (2013). How to do things with data: Meta-data, meta-media, and meta-communication. *First Monday*, 18(10). <https://doi.org/10.5210/fm.v18i10.4870>
- Jensen, K. B. (2014). Audiences, Audiences Everywhere: Measured, Interpreted, and Imagined. In G. Patriarche, H. Bilandzic, J. L. Jensen, & J. Jurišić (Hrsg.), *Audience research methodologies: Between innovation and consolidation* (S. 227–239). New York: Routledge.
- Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Hrsg.), *Handbook of mixed methods in social & behavioral research* (S. 297–319). Thousand Oaks: SAGE.
- Jünger, J. (2018). Mapping the field of automated data collection on the web. Data types, collection approaches and their research logic. In C. Stützer, M. Welker & M. Egger (Hrsg.), *Computational social science in the Age of big data. Concepts, methodologies, tools, and applications* (S. 104–130). Köln: Halem.
- Jünger, J., & Keyling, T. (2020). Facepager. An application for automated data retrieval on the web [software]. <https://github.com/strohne/Facepager/>
- Jünger, J (2021; im Erscheinen). A brief history of APIs. How social media providers shape the opportunities and limitations of online research. In: U. Engel, A. Quan-Haase, S. Liu & L. Lyberg (Hrsg.). *Handbook of Computational Social Science*. London: Routledge. <https://doi.org/10.4324/9781003025245-3>
- Kobayashi, T., Ogawa, Y., Suzuki, T., & Yamamoto, H. (2019). News audience fragmentation in the Japanese Twittersphere. *Asian Journal of Communication*, 29(3), 274–290. <https://doi.org/10.1080/01292986.2018.1458326>
- Kuri, J. (2012). Twitter baut Instagram nach. *heise online*. Abgerufen von <https://www.heise.de/newsticker/meldung/Twitter-baut-Instagram-nach-1766114.html>
- Lane, K. (2016). API Evangelist History of APIs. Abgerufen von <https://history.apievangelist.com/>
- Latour, B. (2017). *Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie*. (4. Aufl.). Frankfurt am Main: Suhrkamp.
- Livingstone, S. (2009). On the Mediation of Everything. ICA Presidential Address 2008. *Journal of Communication*, 59(1), 1–18. <https://doi.org/10.1111/j.1460-2466.2008.01401.x>
- Marres, N. (2018). *Digital sociology: The reinvention of social research*. Cambridge: Polity.

- Münch, F. V., Thies, B., Puschmann, C., & Bruns, A. (2021). Walking Through Twitter: Sampling a Language-Based Follow Network of Influential Twitter Accounts. *Social Media + Society*, 7(1), 1–17. <https://doi.org/10.1177/2056305120984475>
- O'Reilly, T. (2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>
- Porten-Cheé, P., Haßler, J., Jost, P., Eilders, C., & Maurer, M. (2018). Popularity cues in online media: Theoretical and methodological perspectives. *Studies in Communication / Media*, 7(2), 208–230. <https://doi.org/10.5771/2192-4007-2018-2-80>
- Programmable Web (2020). About ProgrammableWeb. <https://www.programmableweb.com/about> (4.3.2020).
- Puschmann, C. (2019). An end to the wild west of social media research: a response to Axel Bruns. *Information, Communication & Society*, 22(11), 1582–1589. <https://doi.org/10.1080/1369118X.2019.1646300>
- RapidMiner (2020). RapidMiner Studio. Visual workflow designer for the entire analytics team [software]. <https://rapidminer.com/products/studio/> (20.03.2020)
- Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender Systems: Introduction and Challenges. In F. Ricci, L. Rokach, & B. Shapira (Hrsg.), *Recommender systems handbook* (2. Auflage, S. 1–34). New York: Springer.
- Rosenberg, M., Confessore, N., & Cadwalladr, C. (17.3.2018). How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>
- Ruths, D., & Pfeffer, J. (2014). Social sciences. Social media for large studies of behavior. *Science*, 346(6213), 1063–1064. <https://doi.org/10.1126/science.346.6213.1063>
- Social Science One (2020). CrowdTangle Data at Facebook. <https://socialscience.one/rfp-crowdtangle>
- Twitter (2007). Twitter Development Talk. *Internet Archive*. <https://web.archive.org/web/20071011034812/http://groups.google.com:80/group/twitter-development-talk/web/api-documentation>

- Twitter (2019). Developers. Tap into what's happening. Publish and analyze Tweets, optimize ads, and create unique customer experiences. *Internet Archive*. <https://web.archive.org/web/20190801135712/https://developer.twitter.com/>
- Twitter (2020a). Twitter data for academic research. *Internet Archive*. <https://web.archive.org/web/20200109004518/https://developer.twitter.com/en/use-cases/academic-researchers>
- Twitter (2020b). Get Tweet timelines. https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline
- Twitter (2020c). Search Tweets. <https://developer.twitter.com/en/docs/tweets/search/overview>
- Twitter (2020d). Follow, search, and get users. <https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-list>
- van der Mersch, V. (2016). Twitter's 10 Year Struggle with Developer Relations. <https://nordicapis.com/twitter-10-year-struggle-with-developer-relations/>
- Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10). <https://doi.org/10.5210/fm.v18i10.4878>
- WinMerge (2020). Ein Open-Source-Programm zum Unterscheiden und Zusammenführen von Dateien [software]. <https://winmerge.org/> (21.030.2020).
- YouTube (2005). Developers. *Internet Archive*. https://web.archive.org/web/20050813012849/https://www.youtube.com/developers_intro.php
- YouTube (2013). YouTube Analytics API. *Internet Archive*. <https://web.archive.org/web/20130115212038/https://developers.google.com/youtube/analytics/v1/>
- YouTube (2019). OAuth Application Verification FAQ. *Internet Archive*. <https://web.archive.org/web/20190211221256/https://support.google.com/cloud/answer/9110914>
- YouTube (2020). Search: list. <https://developers.google.com/youtube/v3/docs/search/list>