
Teil II: **Ethische und rechtliche Aspekte**

Janine Luth, Konstanze Marx, Christian Pentzold

Ethische und rechtliche Aspekte der Analyse von digitalen Diskursen

1 Einleitung: Digitale Diskursanalyse in normativ-regulativer Perspektive

Die Beschäftigung mit digitalen Diskursen operiert in vielgestaltigen medialen Sphären und greift auf Material aus sehr diversen kommunikativen Konstellationen zu. Die für sie relevanten Korpora generieren sich aus Diskursfragmenten, die sich im Grad ihrer Öffentlichkeit und Zugänglichkeit, ihres professionellen oder partizipativen Entstehungszusammenhangs, ihrer politischen Relevanz oder sozialen Akzeptabilität unterscheiden. Eine solche Ausgangslage wirft unweigerlich rechtliche und ethische Aspekte der Analyse digitaler Diskurse auf, die in diesem Kapitel diskutiert werden. Dazu werden zunächst die grundlegenden juristischen Aspekte und ethischen Axiome identifiziert. Im Anschluss werden forschungspraktische Handlungsoptionen und Entscheidungswege für ausgewählte digitale vernetzte Kommunikationsformen skizziert. Das Kapitel schließt mit Empfehlungen für eine rechtlich und ethisch reflektierte Datenerhebung.

In den vorangegangenen Kapiteln wurde deutlich, dass die Erforschung digitaler Gegenstände mit entsprechend dafür entwickelten neuen oder angepassten bestehenden Methoden mit ethischen und rechtlichen Fragen unweigerlich verknüpft ist. Die Thematisierung dieser Fragen ist insofern von besonderer Bedeutung, als dass Diskursen bei der Vermittlung zwischen rechtlichen und ethischen Fragen eine Schlüsselrolle zukommt. Diskurse sind die Aushandlungsinstanz für die Verzahnung von Recht und Ethik. Sichtbar- und Zugänglichkeit des digitalen Diskurses (siehe zu den Eigenschaften zusammenfassend etwa Meier-Vieracker in diesem Band) sind Bedingungen für diese Aushandlungsprozesse.

In diesem Kapitel stehen wir nun aber vor der Aufgabe, die Untersuchung dieses digitalen Diskurses wiederum unter rechtlichen und ethischen Aspekten in den Blick zu nehmen, blenden also die auf der Objektebene konstatierte Funktion für die Aushandlungen von Recht und Ethik zugunsten einer Betrachtung auf der Metaebene aus analytischen Gründen aus.

Rechtliche Aspekte kommen z. B. immer dann ins Spiel, wenn Bild- und Schriftzeichen (oder ggf. auch Tondokumente) als Belege angeführt werden sollen, ohne dass die urhebenden Personen um ihr Einverständnis gebeten werden können. Soweit dies über das wissenschaftliche Zitat zunächst noch recht unkompliziert abgedeckt zu sein scheint, stellt sich dann aber auch die Frage, ob es ethisch vertretbar

ist, Personen in einen wissenschaftlichen Rezeptionskreislauf zu setzen, den sie weder gewünscht noch erwartet haben. Weitere Fragen stehen an, wenn es um den Korpusaufbau mit digitalen oder digitalisierten Daten geht. Welche Daten dürfen zu einem solchen Korpus zusammengeführt werden? Wo, für wen und unter welchen Bedingungen sollte es zugänglich gemacht werden? Ethische und rechtliche Fragestellungen beziehen sich also auch auf die Bereiche Datenspeicherung, Datennutzbarkeit und Datentransparenz.

Im Blick auf dieses Problemfeld hat das Kapitel zum Ziel, ethische und rechtliche Fragen zu bündeln und zu diskutieren, anhand digitaler vernetzter Kommunikationsformen beispielhaft darzustellen und Vorschläge zu unterbreiten, wie mit spezifischen ethisch-rechtlichen Problemstellungen konkret umgegangen werden kann.

Die Beschäftigung mit den rechtlichen und ethischen Herausforderungen der Analyse digitaler Diskurse steht vor einem Dilemma: Auf der einen Seite gibt es das berechnete Interesse, so konkret und aktuell wie möglich Antworten auf die Frage zu finden, wie juristisch konform und ethisch integer agiert werden kann. Auf der anderen Seite macht es die Dynamik sowohl der technologischen Entwicklung als auch von Rechtsprechung bzw. Rechtsetzung unmöglich, längerfristig passgenaue Handreichungen oder Entscheidungsvorlagen anzubieten. Allzu schnell kann eine einmal gefundene Praxis veraltet sein und eine Neubewertung der Situation und ihres normativen wie regulativen Kontexts erforderlich machen.

Die Diskussion dieses Kapitels navigiert folglich zwischen Konkretion hinsichtlich daten- und plattformenindizierter Spezifika und Abstraktion im Sinne allgemein anwendbarer Heuristiken. Dabei sind die Veränderbarkeit und Kontextgebundenheit rechtlicher und ethischer Bewertungen und Entscheidungen natürlich anzuerkennen. Diese Bewertungen und Entscheidungen stehen in der Spannung von axiomatischen Forderungen und deren stets prekären – weil diskutier- und revidierbaren – Einlösung. Die entsprechenden normativen, als universell angesehenen Prinzipien umfassen allgemeine Menschenrechte, den Schutz der Menschenwürde, das Recht auf (informationelle) Selbstbestimmung sowie die Sicherheit und Schutzwürdigkeit der Person (Ess 2009 und Hudson & Bruckman 2004). Sie bilden das ethisch-rechtliche Grundvokabular, welches stets unter den vorliegenden Bedingungen einer Studie forschungspraktisch ausbuchstabiert werden muss. Dazu bestehen institutionelle Vorkehrungen und Mechanismen, etwa in Form von Ethikkommissionen und *institutional review boards* (Buchanan & Ess 2009 und Eynon, Fry & Schroeder 2008). Sie orientieren sich in ihrer Arbeit neben akademischen Statuten und Empfehlungen wissenschaftlicher Kommissionen an Rechtsordnungen wie dem Persönlichkeitsrecht, dem Urheberrecht und Datenschutzrechten. In dem Ansinnen, Forschungsvorhaben gemäß der angesetzten Standards und Ver-

fahren zu prüfen und so deren juristische Konformität und ethische Plausibilität abzusichern, laufen diese Institutionen Gefahr, entweder zu überdeterminiert oder zu unterdeterminiert zu sein. Das heißt, die Vorgaben und Festlegungen können entweder zu starr ausfallen und damit die Offenheit des Forschungsprozesses über Gebühr beschneiden oder sie sind zu wenig bestimmt und liefern keine Orientierung, was genau getan und was nicht getan werden kann. Konkrete Orientierung liefern hier methodische Gütekriterien, mittels derer forschungsethische Überlegungen adressiert werden. Diese betreffen etwa die Objektivität und Integrität des wissenschaftlichen Prozesses und seiner Ergebnisse sowie die Berücksichtigung der informierten Einwilligung und Freiwilligkeit, der Nicht-Schädigung v. a. durch Zusicherung und Einhaltung von Anonymität und Vertraulichkeit (Hopf 2005).

2 Rechtliche Problemstellungen digitaler Diskursanalysen

Intuitiv ist bei der wissenschaftlichen Beschäftigung mit Datensammlungen bereits klar, dass Aspekte des Speicherns, des Weitergebens und vor allem auch des Veröffentlichens rechtliche Implikationen haben. Dennoch erscheinen rechtliche Aspekte bei der Erforschung von Daten aus Social-Media-Kommunikation (Nutzer*innen schreiben, ohne sich bewusst zu sein, dass sie wissenschaftliches Forschungsobjekt werden können, in einer öffentlichen Sphäre) sowie von digitalisierten Daten, die über einen längeren Zeitraum vorgehalten werden sollen, nicht erst seit dem Data-Mining-Paragrafen (UrhG § 60d) in Deutschland in gewisser Weise wie ein Störfaktor. Man könnte also überlegen, ob die Forschung zunächst durchgeführt und eine rechtliche Absicherung im Nachhinein herbeigeführt werden kann. Dies wäre ein gangbarer, aber rechtlich und ethisch fragwürdiger Weg, der den Forschungsaspekt in den Vordergrund rückt. Es ist ein Weg, der Gatto (2014: 64) zufolge allerdings lange überwiegend praktiziert wurde mit der Einschränkung, die gesammelten Daten zwar zu erforschen, aber nicht zu verbreiten.

Allerdings ist für viele Projekte mittlerweile Voraussetzung, dass die Datenspeicherung und Nachnutzbarkeit rechtlich abgesichert steht (Knuchel & Luth 2018). Folgende Fragen sind dabei zu klären: Was ist öffentlich? Wann müssen Daten anonymisiert werden? Inwiefern ist die Anonymisierung auch eine unzulässige Veränderung des Datums, etwa bei Bildern? Wer sind die Urheber*innen der Daten? Wer hat die Rechte an den Daten? Welche rechtlichen Konsequenzen können auf die Wissenschaftler*innen zukommen, wenn sie die Daten erheben, ohne die Urheber*innen in Kenntnis zu setzen? Wann genau sind die Urheber*innen

in Kenntnis zu setzen und wie? Wofür können Wissenschaftler*innen haftbar gemacht werden? Welche Rechtsansprüche müssen erfüllt sein? Wo ist die jeweils zutreffende Rechtsgrundlage zu finden?

Knuchel und Luth (2018) schlagen vor, diese allgemeinen Fragen bei der digitalen Korpusanalyse in drei Leitfragen zu gliedern und sich diese vor Forschungsbeginn zu beantworten: 1.) Mit welchen Daten darf gearbeitet werden? 2.) Wie können Daten nachhaltig bereitgehalten werden? 3.) Soll überhaupt mit diesen Daten gearbeitet werden (oder stehen ethische Gründe dagegen und lassen sich andere, weniger sensible Daten finden)?

Die Basis, auf der rechtliche Fragen bei der Erhebung, Bearbeitung und Archivierung digitaler Daten relevant werden, sind die Urheberschaft und das Nutzungsrecht. Ein Text – ob gedruckt oder digital im Netz – hat eine*n oder mehrere Urheber*innen. Diese Urheber*innen haben ein unveräußerliches Urheberrecht an ihren Werken und können an diesen ein Nutzungsrecht einräumen. Das Urheberrecht hat zunächst eine natürliche Person inne. Das Nutzungsrecht wiederum könnte auch bereits auf eine juristische Person übergegangen sein, beispielsweise an ein Unternehmen. Es wäre demnach zu klären, ob zum Zeitpunkt der Datenerhebung das Nutzungsrecht bereits auf eine andere natürliche oder juristische Person übergegangen ist. Unter Umständen besteht die Möglichkeit, mit dem*der Rechteinhaber*in einen Vertrag über die Nutzung der Daten zu schließen. Hierbei lohnt sich ein Blick in das Data Mining Gesetz aus § 60d UrhG und der daraus erwachsenen Diskussion, ggf. ist die Form der Datenerhebung und Datennutzung bereits darüber gedeckt. Weiterer Spielraum könnte sich für Wissenschaftler*innen aus Art. 5 Abs. 3 GG, der Forschungsfreiheit ergeben (Rolfes & Wendel 2018). Schließlich wird auch die Frage der Nachnutzbarkeit juristisch verhandelt: Unter welchen Voraussetzungen dürfen Datenbanken, bestehend aus digitalen Daten, veröffentlicht und für andere Wissenschaftler*innen nutzbar gemacht werden? Letzteres wird in den meisten Fällen im Sinne der Nachhaltigkeit sogar als Anforderung an Forschungsprojekte formuliert.

Es handelt sich hier nicht um grundsätzlich neue Fragen, auch in der Arbeit mit traditionellen Printmedien oder belletristischer Literatur sind sie natürlich relevant. Für digitale Daten stellen sie sich insofern neu, als dass bei Social-Media-Plattformen, wie Instagram, Twitter, Facebook, Foren und der Wikipedia, die Nutzungsrechte vielfach nicht mehr bei der einzelnen schreibenden Person liegen, sondern bei den Unternehmen. So beschreiben de la Durantaye und Raue, dass in der analogen Welt die Kernfragen „oder Zugangsfragen im Urheberrecht“ – wer darf ein Werk unter welchen Umständen nutzen? – „in der analogen Welt [...] selten problematisiert“ wurden (2020: 83). Es muss auch bedacht werden, dass digitale Texte Ergebnisse kollaborativer Arbeitsprozesse sein können. Ein klassisches

Beispiel hierfür sind Wikipedia-Artikel, die von mehreren Autor*innen geschrieben und immer wieder, u. U. auch von neu dazukommenden Schreiber*innen, modifiziert werden.

Eine wichtige Bezugsgröße für die hier gestellten Fragen ist im deutschen Recht das Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (§ 60d UrhG – Text und Data Mining für Zwecke der wissenschaftlichen Forschung), welches am 1. März 2018 in Kraft trat und zum 7. Juni 2021 in einer Neufassung in Kraft getreten ist (de la Durantaye & Raue 2020: 85). Das Gesetz ist im Zusammenhang mit § 44b UrhG-E zu sehen, welches bestimmt, wann das Mining zu kommerziellen Zwecken erlaubt ist (vgl. Heesen & Jüngels 2021: 45). Der Gesetzgeber hat sich entschieden, die kommerzielle und die wissenschaftliche Nutzung auseinanderzuziehen, was sowohl aus rechtlicher Perspektive als auch aus Sicht der wissenschaftlich Betroffenen sinnvoll erscheint (vgl. Heesen & Jüngels 2021: 45). Der Gesetzestext aus § 60d UrhG, der für die digitale Korpusanalyse relevant ist, lautet wie folgt:

- (1) Vervielfältigungen für Text und Data Mining (§ 44b Absatz 1 und 2 Satz 1) sind für Zwecke der wissenschaftlichen Forschung nach Maßgabe der nachfolgenden Bestimmungen zulässig.
- (2) Zu Vervielfältigungen berechtigt sind Forschungsorganisationen. Forschungsorganisationen sind Hochschulen, Forschungsinstitute oder sonstige Einrichtungen, die wissenschaftliche Forschung betreiben, sofern sie
 1. nicht kommerzielle Zwecke verfolgen,
 2. sämtliche Gewinne in die wissenschaftliche Forschung reinvestieren oder
 3. im Rahmen eines staatlich anerkannten Auftrags im öffentlichen Interesse tätig sind.

Nicht nach Satz 1 berechtigt sind Forschungsorganisationen, die mit einem privaten Unternehmen zusammenarbeiten, das einen bestimmenden Einfluss auf die Forschungsorganisation und einen bevorzugten Zugang zu den Ergebnissen der wissenschaftlichen Forschung hat.

- (3) Zu Vervielfältigungen berechtigt sind ferner
 1. Bibliotheken und Museen, sofern sie öffentlich zugänglich sind, sowie Archive und Einrichtungen im Bereich des Film- oder Tonerbes (Kulturerbe-Einrichtungen),
 2. einzelne Forscher, sofern sie nicht kommerzielle Zwecke verfolgen.
- (4) Berechtigte nach den Absätzen 2 und 3, die nicht kommerzielle Zwecke verfolgen, dürfen Vervielfältigungen nach Absatz 1 folgenden Personen öffentlich zugänglich machen:

1. einem bestimmt abgegrenzten Kreis von Personen für deren gemeinsame wissenschaftliche Forschung sowie
 2. einzelnen Dritten zur Überprüfung der Qualität wissenschaftlicher Forschung.
 3. Sobald die gemeinsame wissenschaftliche Forschung oder die Überprüfung der Qualität wissenschaftlicher Forschung abgeschlossen ist, ist die öffentliche Zugänglichmachung zu beenden.
- (5) Berechtigte nach den Absätzen 2 und 3 Nummer 1 dürfen Vervielfältigungen nach Absatz 1 mit angemessenen Sicherheitsvorkehrungen gegen unbefugte Benutzung aufbewahren, solange sie für Zwecke der wissenschaftlichen Forschung oder zur Überprüfung wissenschaftlicher Erkenntnisse erforderlich sind.
- (6) Rechtsinhaber sind befugt, erforderliche Maßnahmen zu ergreifen, um zu verhindern, dass die Sicherheit und Integrität ihrer Netze und Datenbanken durch Vervielfältigungen nach Absatz 1 gefährdet werden.

De la Durantaye und Raue (2020: 85) folgern, dass nach dem § 60d UrhG für die automatisierte Auswertung einer Vielzahl von Texten das Ursprungsmaterial automatisiert und systematisch vervielfältigt und das dabei entstehende Korpus öffentlich (unter bestimmten Voraussetzungen) zugänglich gemacht werden dürfen. Mit der Neufassung des § 60d UrhG erlaubt der Gesetzgeber auch die Aufbewahrung durch die Wissenschaftler*innen, solange wie die Daten benötigt werden. „Bisher müssen die Forscherinnen und Forscher alle Vervielfältigungen und das Korpus nach Abschluss der Forschungsarbeiten löschen. Zulässig ist jedoch eine Übermittlung an die Gedächtnisorganisationen aus §§ 60e und 60f UrhG, die die Korpora dann dauerhaft aufbewahren dürfen. Eine solche dauerhafte Aufbewahrung soll es nach dem Regierungsentwurf nicht geben. Die zur Forschung Berechtigten (mit Ausnahme der Einzelforscherinnen und -forscher) sollen in Zukunft ihre Korpora selbst aufbewahren dürfen. Allerdings gilt das nur solange, wie es zum Zweck der wissenschaftlichen Forschung oder der Überprüfung der Ergebnisse notwendig ist“ (Heesen & Jüngels 2021: 48).

Heesen und Jüngels (2021: 50) kritisieren jedoch, dass der deutsche Gesetzgeber hinter den Möglichkeiten zurückbleibt, die sich durch das europäische Recht eröffnen (§ 60d UrhG ist eine Umsetzung der Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates): „Es wird vertreten, dass ‚eine Überprüfung wissenschaftlicher Erkenntnisse‘ dauerhaft erforderlich sei. Ob sich so im Wege einer wörtlichen und teleologischen Auslegung argumentieren ließe, dass § 60d V UrhG-E eine dauerhafte Aufbewahrung gestattet, ist zweifelhaft.“ Die Aufbewahrenden sähen sich also weiterhin einer Rechtsunsicherheit gegenüber, so die beiden Autor*innen.

Neben dem Wunsch nach Sicherheit im Arbeiten mit den digitalen Daten stehen auch ethisch-rechtlich wichtige Frage nach Anonymisierungsvorgängen im Raum. Mit dieser Form hat man es zu tun, wenn Usernamen auftreten. Einen digital veröffentlichten Post zu anonymisieren, ist ein zweischneidiges Schwert: Zwar schützt dies die Person, weil Name und Profilbild nicht in einer wissenschaftlichen Publikation sichtbar sind, jedoch ist die Veröffentlichung des Posts dann unter Umständen nicht mehr durch das Zitationsrecht gedeckt, da in das Wort-Bild-Zitat eingegriffen wurde. Rechtlich wäre demnach zu empfehlen, keine Anpassungen in einem solchen Kommunikat vorzunehmen (vgl. auch Münch in Vorbereitung). Da darüber hinaus auch immer mehr Drittmittelgeber fordern, dass die Korpora nach Projektablauf langfristig archiviert werden, um eine weitere Nutzung auch durch Dritte sicherzustellen, wird ein Datenmanagement erforderlich und sollte ebenfalls rechtzeitig bei der Projektplanung bedacht werden (vgl. Apel, Föhles & Kratzke 2017).

Der rechtliche Rahmen scheint somit durch das Data Mining-Gesetz gute Möglichkeiten zur Forschung zu eröffnen, die jedoch in Teilen noch immer nicht die Rechtssicherheit bieten, die von den Forscher*innen gewünscht wird. Zudem scheinen die Nutzungsbedingungen der hier untersuchten Plattformen dieser grundsätzlichen Offenheit teilweise entgegenzustehen.

3 Ethische Problemstellungen digitaler Diskursanalysen

Durch die rechtlichen Voraussetzungen wird der Rahmen für die Erforschung digitaler Diskurse nicht vollständig abgesteckt. Notwendig ist zudem eine ethische Einordnung unserer Forschungshandlungen gerade dort, wo zwischenmenschliche Kommunikation in den Vordergrund rückt, die nicht explizit zum Zwecke der Forschung entsteht. Nachfolgend gehen wir auf die daraus resultierenden Implikationen ein.

Digitale Diskurse eröffnen Fragen nach den Bedingungen, Berechtigungen und Grenzen, um Korpora zusammenzustellen, zu sichern, auszuwerten und zu teilen. Der Umgang mit digital zugänglichen Diskursfragmenten muss reflektieren, dass diese in ihrer Periodizität und in ihrem Grad an Zerstreutheit variieren, sodass sich keine verlässlichen Grundgesamtheiten bestimmen lassen. Immer bleiben potentielle diskursive Fragmente der zeitlich und räumlich entgrenzten „mass self-communication“ (Castells 2009: 70) unberücksichtigt. Die Sichtbarkeit und Publizität von Kommunikaten gründen in sozialen Dynamiken und technologischen, algorithmisch basierten Prozessen. Verlassen sich Diskursanalysen

auf diese Selektionsmechanismen, laufen sie Gefahr, gewisse Kommunikate aufgrund ihrer Präsenz, ihrer Quelle oder der Sprechenden unreflektiert zu privilegieren, während sie zugleich dazu beitragen, randständige Diskurspositionen oder -akteur*innen weiter zu marginalisieren. Der Grad an Diffusion und Vernetzung bzw. Zentralität, also Rückführbarkeit auf Leitmedien, hat auch Konsequenzen für die Überprüfbarkeit und Belegbarkeit von Aussagen. Zwar muss es nicht die Aufgabe sein, die Wahrhaftigkeit oder Korrektheit der in einem Diskurs geäußerten Positionen zu überprüfen – gerade der Eigensinn der diskursiven Zirkulation von Denksystemen und Aussagen steht vielfach im Fokus. Jedoch kann sich die Beschäftigung mit digitalen Diskursen angesichts der Konjunktur von sogenannten *Fake News* im Sinne von Desinformationen (bzw. ihrer diskursiven Neuthematisierung), medialer Hetzrede und postfaktischen Tatsachenbehauptungen nicht gänzlich dieser Herausforderung entziehen.

Ethisch relevant werden außerdem die variable Sichtbarkeit und Persistenz von Dokumenten, weshalb synchron und diachron stets nur empirische Ausschnitte aus dem möglichen Gesamtdiskurs präsent sind und es unklar bleibt, inwiefern konkrete Korpora die virtuellen Korpora repräsentieren. Realisierte Elemente eines Diskurses können aus verschiedenen Gründen nicht zugänglich sein, etwa durch geschützte Bereiche, Bezahlschranken oder De-Publikation. Sich zur Lösung dieser Probleme auf Suchmaschinen und derzeit zentrale Kommunikationsplattformen wie Facebook, Twitter und Instagram zu verlassen, hieße wiederum, deren rechtliche und technologische Entscheidungen zu akzeptieren, welcher Teil eines Diskurses sichtbar und welcher gesperrt sein soll. Damit verbunden sind ethische Fragen, inwiefern Äußerungen ohne Wissen und Zustimmung ihrer Sprechenden genutzt werden dürfen, nur weil sie unter Umständen öffentlich zugänglich sind (boyd & Crawford 2012: 673). Technische und rechtliche Hürden treten zudem eventuell bezüglich der Verarbeitung, Weitergabe oder Speicherung von Diskursfragmenten in Kraft. Wie wir oben ausgeführt haben, gelten gerade für die populären Plattformen neben der DSGVO häufig in den AGBs festgehaltene Beschränkungen für Datenbesitz und Datenweitergabe, die Forschungsvorhaben vor die Aufgabe stellen, ethisch-normative Erwartungen an Offenheit und Nachvollziehbarkeit der Analyse mit vertraglichen Bindungen auszubalancieren und ggf. auch dafür zu bezahlen. Diese Umstände und Bedingungen von digitalen Diskursen haben Auswirkungen auf die Herstellung von Öffentlichkeit sowie den Umgang mit persönlichen Informationen, welche auch eine (forschungs-)ethische Komponente in sich tragen.

(a) Öffentlichkeit(en) digitaler Diskurse

Digitale Diskurse können prinzipiell in jeder dazu verfügbaren Kommunikationsform hervorgebracht werden. Ihre situativen Merkmale und die dabei gebrauchten technisch-medialen Optionen bedingen nicht nur die kommunikativen Möglichkei-

ten und sozialen Beziehungsgefüge. Zugleich beeinflussen die verfügbaren Repertoires an Kodes (wie Schrift, Bild oder Bewegtbild), Zeichenmodalitäten (visuelle, auditiv usw.), Kommunikationsrichtungen (uni- oder bidirektional), Funktionsweisen (der Speicherung, Übermittlung oder auch Modulation) sowie die Zeitlichkeit (unmittelbar synchron bis stark asynchron) die Zahl und Relation der Kommunikationspartner*innen (one-to-one, one-to-some oder auch one-to-many) sowie die sozialen Rahmen auch die ethische Bewertung von digitalen Diskursen. Gerade im Blick auf digitale Diskurse vervielfachen sich somit nicht nur die Möglichkeiten, Botschaften herzustellen und zu teilen, sondern auch die ethischen Problemstellungen werden komplexer.

Besonders augenfällig wird diese Einsicht angesichts der vielgestaltigen Öffentlichkeiten in digitalen Diskursen (Bruns 2019 und Papacharissi 2010). In massenmedialen Diskursen war zumindest davon auszugehen, dass diese vornehmlich von professionellen Institutionen wie Verlagen und Medienanstalten und dort privilegiert zu Wort kommenden Kommentator*innen öffentlich angestoßen wurden und ihren Widerhall in disperser und nicht-öffentlicher Anschlusskommunikation des Publikums fanden (Neidhard 1994). Diese schon immer zu idealistische Trennung von Medienöffentlichkeit, Themen- und Encounter-Öffentlichkeit ist angesichts der Vielzahl semi-öffentlicher und sich dynamisch wandelnder Sphären und Arenen nicht mehr aufrechtzuerhalten (Eisenegger et al. 2021 und Klaus & Drüeke 2017).

Äußerungen in digitalen Diskursen haben ein potenzielles Publikum, für das eine Nachricht prinzipiell erreichbar wäre. Davon zu unterscheiden ist das schwankende, aber messbare empirische Publikum, also diejenigen, die eine Seite besuchen, einen Link klicken oder einen Kommentar hinterlassen. Von solchen potenziellen und empirischen Publika sind indessen wiederum das intendierte Publikum als die vorgestellte Leser- und Autor*innenschaft einer Botschaft sowie das mit bestimmten Aussagen adressierte Publikum zu unterscheiden. Öffentlichkeit wird in diesem Sinn zur abgestuften *publicness* von Botschaften in professionell hergestellten wie persönlichen Öffentlichkeiten. Deren Wahrnehmungschancen sind im Prinzip zwar gleich hoch, ihre empirische Rezeptionswahrscheinlichkeit sowie die intendierten und adressierten Publika variieren indessen und sind geprägt durch ungleiche Chancen, sich äußern zu können und Aufmerksamkeit zu generieren (Bennett & Pfetsch 2018).

Zusätzlich ethisch relevant wird das mehrschichtige Verhältnis von Öffentlichkeit bzw. Privatheit dadurch, dass nicht nur einzelne Botschaften, sondern auch die verwendeten Kommunikationsformen aus Sicht der Nutzenden, der Betreibenden wie der Forschenden mit einem variablen ‚sense of audience‘ einher gehen. Mit Nissenbaums (2011: 2) Konzept von Privatheit als „contextual integrity“ zufolge unterscheiden sich Plattformen als auch einzelne darauf verfügbar gemachte Seiten und kommunikative Situationen entsprechend der jeweils zugeschriebenen

Kommunikationsnormen, der erwartbar aktiven Akteur*innen und der relevanten Informationen. Diese Zuschreibungen sind konventionalisiert, nur bleibt es schwierig, daraus pauschale Maximen abzuleiten, was in digitalen Diskursen als öffentlich und was als privat zu gelten hat (Pentzold 2015 und 2017).

Ethisch ist überdies zu berücksichtigen, dass Sprecher*innen in einem Diskurs ihre Beiträge nicht nur an verschiedenen große, sondern auch an verschiedenen zusammengesetzte Publika adressieren. Dabei kann nicht ohne Weiteres davon ausgegangen werden, dass sie hierbei auch eine (fachwissenschaftliche) Öffentlichkeit im Blick haben, wie sie Diskursanalysen häufig herstellen (Sveningsson-Elm 2009). So stellen Boellstorff et al. (2012: 135) folgenden Unterschied fest: „not only what is public versus private from an etic perspective, but also what the people we study emically perceive as public or private“.

(b) Anonymität in digitalen Diskursen

Ethisch ebenso von Bedeutung sind die in digitalen Diskursen möglichen Formen der Selbstdarstellung. Um über sich selbst Identität herzustellen, bestimmte Teildentitäten zu inszenieren und andere Akteur*innen im Netz zu identifizieren, eröffnen Kommunikationsformen verschiedene Optionen und Ressourcen (vgl. dazu auch den Beitrag von Vogel in diesem Band). Identitätsarbeit kann dabei etwa durch die Wahl eines Nicknamens oder einer E-Mail-Adresse, das Ausfüllen eines Profils, die Kommunikationsweise in einer sozialen Netzwerkplattform, die Nutzung eines Bildes bzw. einer Selbstbeschreibung im Chat oder die Gestaltung eines Charakters in einem Online-Spiel erfolgen (Baym 2010 und Fraas, Meier & Pentzold 2012: 74–76).

Jedoch sind die systemgenerierten und mitnutzerproduzierten Identitätsmarker nicht völlig kontrollierbar und können somit als Mittel zur Selbst- und Fremdbeschreibung dienen. Erstere werden von den technischen Anwendungen selbst produziert und können als Identitätsmerkmal gedeutet werden. So registriert ein Chat-Kanal die Dauer des Einloggens sowie die Zahl und den Zeitpunkt von Beiträgen, ein E-Mail-Programm gibt Auskunft darüber, wann eine Nachricht verschickt wurde, oder Facebook verzeichnet die Aktualisierung eines Profils. Zweitere umfassen alle Informationen, die von anderen über eine Person hinterlassen werden. Dies kann eine Käuferbewertung auf Ebay oder ein Kommentar zu einem Video auf YouTube sein. In diesem Sinn ermöglichen und beschränken die sozialen und technischen Bedingungen die Inszenierungs- und Identifizierungsweisen in digitalen Diskursen. Die Selbstdarstellung findet ferner in unterschiedlich anonymen Konstellationen und unter variierenden Authentizitätsnormen statt.

Gerade an dieser Stelle treten ethische, insbesondere forschungsethische Überlegungen in den Vordergrund. Neben der Frage, wie viel Raum in Kommunikationsformen dem spielerischen Ausleben und der Vervielfältigung von

Identitätsaspekten gegeben werden soll (ein Thema, das vor allem in frühen Arbeiten behandelt wurde; vgl. Bruckman 2002), geht es forschungsethisch darum, die Anonymität der Beteiligten abzusichern und Vertraulichkeit zuzusagen (Buchanan & Ess 2009, Eynon, Fry & Schroeder 2008 und Hopf 2005).

(c) Forschungsethische Problemfelder

Ausgehend von der variablen Öffentlichkeit von Diskursaussagen und der Identifizierbarkeit der Sprechenden eröffnen sich zwei forschungsethische Problemfelder: *Öffentlichkeit vs. Anonymität* und *Anonymität vs. informierte Zustimmung* (Pentzold 2015 und 2017).

Das erste Problemfeld *Öffentlichkeit vs. Anonymität* fokussiert die Konsequenzen der (bedingten) Öffentlichkeit von selbst-, fremd- oder systemgenerierten Informationen für das Gewährleisten von Anonymität. Während in Diskursanalysen üblicherweise diese Überlegungen aufgrund der bereits bestehenden Öffentlichkeit des erfassten Materials beiseitegeschoben wurden, stellen sie sich aufgrund des vielgestaltigen Verhältnisses von öffentlicher und privater Kommunikation in digitalen Diskursen. Eine forschungsethische Anforderung stellt etwa Anonymität dar. Hiermit beziehen wir uns auf Nicht-Identifizierbarkeit, also den Umstand, dass Personen untereinander ihre Identität (ihren Namen, ihr Aussehen, ihre Fähigkeiten und Ähnliches) nicht bekannt geben und diese ohne realweltlichen Eindruck auch nicht ohne Weiteres feststellbar ist. Das ist insbesondere der Fall, wenn es sich nicht um professionell erzeugte Botschaften handelt (bspw. durch Politiker*innen oder Journalist*innen). Anonymität bleibt dann nicht nur eine Möglichkeit, erwünschte oder bisher unterdrückte Selbstaspekte in Diskursen zu thematisieren bzw. auszuleben (etwa im Role Play in Online Games). Es stellt sich darüber hinaus die Frage, inwiefern die durch Forscher*innen hergestellte Anonymität nicht einem Autor*innenrecht gegenübersteht.

Hinzu kommen weitere ethische Abwägungen, etwa hinsichtlich der Brisanz von Metadaten, also systemgenerierten Informationen zu IP-Adresse, Standort, Zeitpunkt oder Browserverlauf, die mögliche Anhaltspunkte zur De-Anonymisierung und Identifikation von Personen liefern. In diesem Zusammenhang sprechen Beaulieu und Estalella (2012: 29) von einer zumindest potentiellen „traceability“ von wörtlichen Zitaten oder Namen, das heißt der „possibility of locating digital data on the Internet using search engines or any other mechanism enabled by digital platforms (log files, user profiles, etc.)“ (Beaulieu & Estalella 2012: 32). Dieser Umstand verkompliziert den Umgang mit Belegstellen und Diskursausschnitten, wenn Äußerungen mühelos auf Sprechende zurückgeführt werden können und somit einer „deductive disclosure“, wie Boellstorff et al. (2012: 137) schreiben, dienen können. Es entsteht die für Forschende ethisch relevante Frage, ob wörtlich zitierte Belege, die die Grundlage linguistischer Analysen bilden, im Fließtext

aufgeführt werden können oder welche Alternativen der Darstellung eines Untersuchungsgegenstandes es überhaupt geben könnte.

Das zweite Problemfeld *Anonymität vs. informierte Zustimmung* erfasst die Schwierigkeiten, welche die Öffentlichkeit und Anonymität bzw. Pseudonymität des Interagierens online für das Absichern von informierter Zustimmung mit sich bringen. Auf der einen Seite steht die forschungsethische Forderung nach *Anonymität der einbezogenen Personen*, auf der anderen Seite das ebenso geforderte *Einholen informierter Zustimmung* von anonym, pseudonym (mit Nutzernamen) oder mit bürgerlichem Klarnamen agierenden Autor*innen. Grundsätzlich ist es für Diskursanalysen problematisch, informierte Zustimmung einzuholen und dabei das Vorgehen und die Erwartungen der Forschenden mit den Ansprüchen der in die Untersuchung einbezogenen Personen abzugleichen. In der Praxis umfasst die Rekonstruktion von diskursiven Strukturen und diskursiver Praxis das Sammeln von Dokumenten und das retrospektive Erfassen dokumentierter Aktionen. Entsprechend können die Sprechenden häufig nur nachträglich, nicht aber im Diskursgeschehen, angesprochen werden. Indes setzen sich traditionelle Diskursanalysen selten damit auseinander, Zustimmung für die Erhebung und Analyse publizierten Materials einzuholen. Eher zum Thema wird dies in der Gesprächsanalyse, bei der die Aufzeichnung von Alltagskommunikation mit der Zustimmung der Beteiligten und der Tilgung von personenbezogenen Angaben vereinbar sein sollte (Deppermann & Hartung 2012 und Gülich & Mondada 2012). Digitale Diskurse vereinen nun Elemente von beiden Materialbereichen (vgl. dazu auch den Beitrag von Meier-Vieracker sowie Gredel & Mell in diesem Band): Sie untersuchen manifeste Wissensordnungen und ihre multimodale Realisierung, zugleich erweitern sie den Fokus auf authentische informelle Konversationen und Diskussionen jenseits massenmedialer Arenen und ihrer professionellen Sprechenden. Zusammengefasst bedeutet das, dass forschungsethische Entscheidungen in beiden Problemfeldern in mindestens viererlei Hinsicht zu treffen sind (Pentzold 2015):

1. Darf ein Dokument gesammelt werden?
2. Darf Material zitiert werden?
3. Kann die informierte Zustimmung der involvierten Akteur*innen eingeholt werden?
4. Muss informierte Zustimmung eingeholt werden?

Die Antworten auf diese Frage sind selten eindeutig. Vielmehr ist es nötig, forschungsethische Abwägungen zu treffen, die das Thema und die Brisanz des Diskurses, die Äußerungssituation und die genutzte Kommunikationsform ebenso berücksichtigen wie die bekannten Personenmerkmale (z. B. ob es sich um Minderjährige handelt)

und die Möglichkeit, mit ihnen in Kontakt zu kommen. Konsequenterweise liegt das Spektrum an Optionen zwischen zwingend zu beachtenden Vorgaben für Datensammlung und Datenauswertung und optionalen Vorkehrungen.

Im Blick auf die zu treffenden forschungsethischen Entscheidungen ist demnach zu prüfen, inwiefern ein Diskursfragment aus in einem weit öffentlichen, bedingt öffentlichen, bedingt privaten und auch stark privaten Rahmen entnommen wird. Ebenso ins Gewicht fällt die namentliche bzw. anonyme Kennung der Akteur*innen, also ob sie namentlich-persönlich, pseudonym-identifizierbar, pseudonym-nicht identifizierbar und (vermeintlich) anonym agieren. Ein zusätzlicher Faktor ist der Grad an Interaktion zwischen Diskursakteur*innen und Forschenden, die synchron wechselseitig, asynchron wechselseitig, einseitig und nicht vorhanden sein kann. Eine Entscheidung mag ebenso andere Faktoren wie die potenzielle Vulnerabilität der einbezogenen Subjekte oder die Sensibilität der jeweiligen Aussage beachten (Pentzold 2015).

4 Fünf Schritte zur Adressierung forschungsethischer und rechtlicher Rahmenbedingungen

Um die forschungsethischen Vorgaben und rechtlichen Rahmenbedingungen fallweise umzusetzen bzw. korrekt zu adressieren, bietet sich eine allgemeine Heuristik der gestaffelten Konkretisierung und kommunikativen Validierung an (Ess 2009 und McKee & Porter 2008). Sie umfasst fünf Schritte.

Erstens steht an, die fallweisen Umstände und damit die mediale und kommunikative Konstellation zu bestimmen, aus deren Kontext eine diskursanalytische Studie Daten bezieht bzw. die mit den in einer Studie gebrauchten Daten identifizierbar wird. Auch bei der Nutzung bestehender Korpora ist zu eruieren, welche originären bzw. fortbestehenden oder sich gewandelten Diskurssituationen den Daten zugrunde liegen und damit für eine rechtliche und ethische Betrachtung berücksichtigt werden müssen.

Zweitens geht es darum, die in einem Fall relevanten Axiome und Regularien zu identifizieren. Dabei wird nicht die Geltung oder Verbindlichkeit von Gesetzen, Rechtsnormen oder ethischen Prinzipien in Frage gestellt, sondern andersherum geprüft, welche Rechte bzw. moralischen Grundsätze durch ein Vorhaben in besonderem Maße tangiert werden. Dies können zum Beispiel Fragen des intellektuellen Eigentums oder des Rechts am eigenen Bild sein, es kann auch Fragen der Anonymität oder der erforderlichen informierten Zustimmung betreffen.

Drittens sollte nach *best practice*-Beispielen oder Musterlösungen gesucht werden, an denen das zu bearbeitende Vorgehen geschult werden kann. Hier können auch die verschiedentlich herausgegebenen Leitfäden zum Tragen kommen, so etwa die aktuellen Ethical Guidelines der Association of Internet Researchers (AoIR 2019) bzw. die Handreichungen zum Umgang mit qualitativ, also zumeist in Text- und Bildform, bzw. quantitativ, also häufig numerisch, vorliegenden Datenbeständen (GESIS 2021 und QualiService 2021). Auch eine Orientierung an vorbildhaften Präzedenzstudien kann helfen, die relevanten Rechtsbestimmungen und ethischen Normen auf der einen und die Gegebenheiten des empirischen Feldes auf der anderen Seite überein zu bringen und somit ein akzeptables Vorgehen zu entwerfen.

Viertens ist vorgesehen, die rechtlich und ethisch reflektierten Prozedere in Datenerhebung bzw. -generierung, Datenaufbereitung, Datenauswertung, Datenpräsentation und Datensicherung sowie -verfügarmachung kommunikativ zu validieren. Dazu dient die Konsultation mit Bezugsgruppen und gegebenenfalls den beforschten Akteurskreisen bzw. einzelnen Personen selbst. Bezugsgruppen sind neben internen Foren von Fachkolleg*innen und Kooperationspartner*innen und externen Foren, bestehend aus Ethikkommissionen, Geldgebern und anderen Disziplinen, auch weitere mögliche Akteure. Dies können politische Akteure sein oder zum Beispiel Journalist*innen, die sich für eine Studie interessieren und Fragen zu ihrem rechtlichen oder ethischen Status formulieren. Abbildung 1 veranschaulicht diese Bezüge zwischen Forschenden, forschungsethisch relevanten Bezugsgruppen und Kontextbedingungen.

Das heuristische Verfahren wird fünftens abgeschlossen mit der Festsetzung des Vorgehens, dessen Umsetzung und Darstellung. Dies muss kein starr linearer Prozess sein. Das heuristische, schrittweise Eruiere der rechtlichen und ethischen Umstände und der forschungspraktischen Umsetzung impliziert vielmehr ein unabgeschlossenes, rekursives Verfahren. Jede noch so umsichtige Prüfung kann vor dem Problem stehen, dass einmal getroffene Entscheidungen revidiert werden müssen – sei es, weil Aspekte von vornherein unberücksichtigt geblieben sind, sei es, weil sich die empirische Situation oder das Umfeld der Studie gewandelt haben.

Die ersten drei Schritte in dem skizzierten Prozedere sollen im Folgenden für prinzipiell zugängliche Social Media-Formate (wir schließen Blogs hier ein) und Messengerdienste sowie die Wikipedia im Sinne einer angewandten, auf fachspezifische Studien ausgerichteten Ethik (vgl. Giaxoglou 2017: 230) exemplarisch vollzogen und synoptisch dokumentiert werden.

Die bekanntesten Social Media-Plattformen sind neben Facebook und Instagram WhatsApp, TikTok, Snapchat, Twitter, Pinterest, aber auch Xing und LinkedIn. Wir unterscheiden hier, wie oben bereits hergeleitet, nach dem Grad der Öffentlichkeit und gruppieren daher die meistgenutzten Plattformen Facebook und

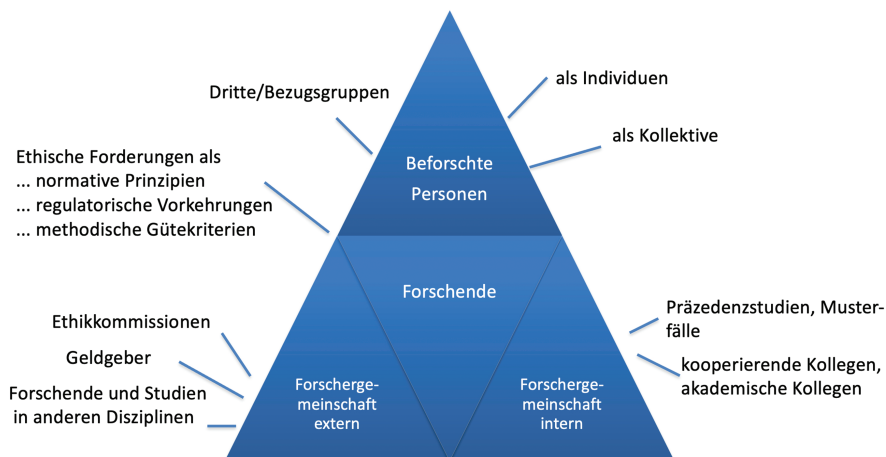


Abb. 1: Forschungsethische Bezugsgruppen. Quelle: eigene Darstellung in Anlehnung an McKee & Porter 2008: 17.

Instagram sowie TikTok, Twitter und Pinterest zu einem Fallbeispiel Social Media. Über eine gedankliche Brücke vom Mikroblogging, wie es auf Twitter geschieht, hin zum Bloggen werden in diesen Punkt auch ethisch relevante Fragen integriert, die sich für Daten auf Blogs ergeben. WhatsApp oder Snapchat sind Messengerformate, die wir für eine zweite Fallbeispielbetrachtung zusammenfassen.

Die kostenpflichtigen berufsorientierten Plattformen Xing („Facebook für den Beruf“¹) und LinkedIn stehen aufgrund ihrer erhöhten Zugänglichkeitsbarriere nicht im Fokus der hier angestellten Überlegungen. Für diese beiden Angebote lassen sich jedoch die Überlegungen zum hier diskutierten ersten Fallbeispiel Social Media adaptieren.

4.1 Fallbeispielgruppe 1: Social Media

Fallweise und mediale Konstellation

Die oben vorgenommene Gruppierung der Plattformen Facebook, Instagram, TikTok, Twitter und Pinterest basiert auf dem Grad der Öffentlichkeit, die wir als niedrigschwellige Zugänglichkeit spezifizieren wollen. Facebook, Instagram und TikTok können nach einer Registrierung ohne App webbasiert rezipiert werden. Mit Siever

¹ <https://www.giga.de/tipp/was-ist-xing-einfach-erklart/>.

(2015: 22) betrachten wir die Registrierung nicht als Zugänglichkeitsbeschränkung, weil diese allen offensteht. Für Twitter und Pinterest (hier mit dem Umweg über beliebige Pin-URLs oder Unterseiten²) ist keine Registrierung für die Rezeption der Inhalte notwendig und die Rezeption ist ebenfalls ohne App webbasiert möglich.

Alle Dienste bieten die Möglichkeit, unter Klarnamen oder Pseudonym Texte, Bilder und audiovisuelle Daten zu posten. Die Veröffentlichung von Aussehen und Stimme integrierend sehen wir uns also ggf. einem sehr breiten Spektrum persönlicher Informationen gegenüber. Darüber hinaus sind die Plattformen auf hohe Nutzer*innenaktivität in Frequenz und Reichweite angelegt, die jeweils wieder mehr Sichtbarkeit und Resonanz generieren. Dabei stammen die Inhalte vielfach aus dem privaten oder auch dem beruflichen Kontext, also aus einem sehr persönlichen Bereich. Dazu gehören Angaben über Wohnort, Freundeskreise, bestehende oder vergangene Liebesbeziehungen, Familienkonstellationen, Beschäftigungsverhältnisse, persönliche Verluste, Krankheiten oder Todesfälle, dazu gehören natürlich auch Meinungen zu unterschiedlichen (auch politischen und gesellschaftlichen) Sachverhalten. Die Vernetzung von Nutzer*innen untereinander gibt Aufschluss über soziale Konstellationen, die z. T. auch analog bestehen und über die weitere Informationen akquiriert werden können, die zu Persönlichkeitsclustern verdichtet werden können. Über die aktiv eingegebenen Informationen (in der Gestaltung der Profilseiten oder in Texten) hinaus werden Metadaten, die über spezifische Viewer ausgelesen werden können, in veröffentlichten Bildern und Videos zugänglich. Das können neben kameraspezifischen Einstellungen z. B. auch Positionsdaten sein (Welchering 2020: 22–24). Blogs sind uneingeschränkt zugänglich und bieten für die Schreiber*innen einen breiteren Gestaltungsspielraum mit Blick auf die Texte, eingebundenen Bilder oder Videos. Die Möglichkeit, Twitterfeeds auf dem Blog einzubinden, wird von vielen genutzt.

Identifikation relevanter Axiome und Regularien

Häufig wird gerade im Zusammenhang mit Sozialen Medien die Mutmaßung geäußert, dass es sich um Daten handle, die ohnehin öffentlich sind, man sie folglich zweifelsohne auch für wissenschaftliche Analysen verwenden könne. Den Nutzer*innen sei schließlich bewusst, in welchen medialen Kontexten sie interagieren. Darüber hinaus willigen sie mit ihrer Registrierung, die Bedingung für eine produktive Partizipation ist, in die Nutzungsbedingungen der jeweiligen Plattformen ein. Dieser Vorgang wird aus Sicht der Plattformbetreiber*innen

² https://praxistipps.chip.de/pinterest-ohne-anmeldung-nutzen-so-gehts_38220.

als Akt des Rechtstransfers an sie behandelt, in den jeweiligen Allgemeinen Geschäftsbedingungen ist das spezifiziert. So heißt es beispielsweise bei Facebook: „Insbesondere wenn du Inhalte, die durch geistige Eigentumsrechte geschützt sind (wie Fotos oder Videos), auf oder in Verbindung mit unseren Produkten teilst, postest oder hochlädst, gewährst du uns eine nicht-exklusive, übertragbare, unterlizenzierbare und weltweite Lizenz, deine Inhalte (gemäß deinen Privatsphäre- und App- Einstellungen) zu hosten, zu verwenden, zu verbreiten, zu modifizieren, auszuführen, zu kopieren, öffentlich vorzuführen oder anzuzeigen, zu übersetzen und abgeleitete Werke davon zu erstellen“³ (vgl. auch die gleichlautenden AGBs von Instagram⁴).

Bei Twitter heißt es ebenfalls: „Durch Übermittlung, Veröffentlichung oder Anzeigen von Inhalten auf oder über die Dienste gewähren Sie uns eine weltweite, nicht ausschließliche, unentgeltliche Lizenz (mit dem Recht zur Unterlizenzierung), diese Inhalte in sämtlichen Medien und über sämtliche Verbreitungswege, die gegenwärtig bekannt sind oder in Zukunft bekannt sein werden (zur Klarheit, diese Rechte umfassen z. B. das Kuratieren, Transformieren und Übersetzen) zu verwenden, zu vervielfältigen, zu reproduzieren, zu verarbeiten, anzupassen, abzuändern, zu veröffentlichen, zu übertragen, anzuzeigen und zu verbreiten. [...] Mit dieser Lizenz erteilen Sie uns die Erlaubnis, Ihre Inhalte weltweit verfügbar zu machen und dies auch Dritten zu ermöglichen.“ Darüber hinaus wird klar formuliert: „Sie haben verstanden, dass Sie durch die Nutzung der Dienste Ihre Einwilligung zur Erhebung und Nutzung (wie in der Datenschutzrichtlinie dargelegt) von diesen personenbezogenen Daten, einschließlich der Übertragung dieser Daten in die Vereinigten Staaten von Amerika, nach Irland und/oder in andere Länder für die Speicherung, Bearbeitung und Nutzung durch Twitter und seine verbundenen Unternehmen erteilt haben.“⁵

Deutlich wird, dass sich Nutzer*innen im Wesentlichen mit der Weiterverwendung ihrer Inhalte durch die Plattformen einverstanden erklären. Mit Blick auf Blogs könnte die Freischaltung der Kommentarfunktion ein Indiz dafür sein, dass mit dem Schreiben eine gewisse Öffentlichkeit angestrebt werde (Meier-Vieracker 2021). Konkret ist nun zu fragen, ob die Aktivität auf einer der genannten Social-Media-Plattformen als Zustimmung zur Weiterverwendung der Daten für wissenschaftliche Zwecke gelesen werden darf. Die Antwort ist so schlicht wie ernüchternd: Der Nutzung für wissenschaftliche Studien wird durch die Akzeptanz der Nutzungsbedingungen nicht zugestimmt, wenngleich eine wissenschaft-

³ <https://de-de.facebook.com/terms>, aktualisiert am 20.12.2020.

⁴ https://help.instagram.com/1215086795543252?helpref=page_content, aktualisiert am 20.12.2020.

⁵ <https://twitter.com/de/tos>, aktualisiert am 18.06.2020.

liche Nutzung deutlich sensibler mit den erhobenen Daten verfahren würde, als dies in den oben zitierten Geschäftsbedingungen skizziert ist. Zudem muss immer auch berücksichtigt werden, dass die Zustimmung zu diesen Geschäftsbedingungen dann notwendig ist, wenn ein Interesse an der Nutzung der Plattform besteht, so dass eine Entscheidung dagegen immer auch mit einem Ausschluss aus dieser kommunikativen Arena einherginge. Der Entscheidungsspielraum ist also deutlich eingeschränkt, eine Voraussetzung, die so für Entscheidungsspielräume im Vorfeld von wissenschaftlichen Studien nicht gegeben ist.

Der benannte eingeschränkte Entscheidungsspielraum führt häufig dazu, dass die Geschäftsbedingungen nicht eingehend gelesen werden, was in unterschiedlichen Zusammenhängen so auch klar thematisiert wird, z. B. in der Zeit online mit Bezug auf AGB generell: „Kaum einer liest sie, noch weniger verstehen sie. Und wer die AGB beim Online-Einkauf nicht akzeptiert, wird vom Konsum ausgeschlossen“ (Steuernagel, Frey & Friedrich 2017) oder im Teaser zu einem Blogbeitrag über die TikTok-AGB: „Wir alle akzeptieren sie, ohne sie vorher zu lesen. Doch was steht eigentlich in den Tik-Tok-AGB?“ (Lexa 2020). Dass aus Bildern und audiovisuellen Daten weitere persönlich relevante Informationen ausgelesen werden können, ist Allgemeinen Geschäftsbedingungen nicht zu entnehmen. Diesen Kenntnisstand müssen sich Nutzer*innen selbst erarbeiten. Es gibt also durchaus Anlass zur Annahme, dass die Nutzer*innen von Sozialen Medien nicht einfach als informiert eingestuft werden können. Darüber hinaus verbietet die gute wissenschaftliche Praxis eine Haltung, in der die Bedingungen von Medienunternehmen schlicht übernommen werden.

Auch die Sensibilisierung für potenziell problematische, transportierte Inhalte unabhängig von der jeweiligen Zeichenmodalität ist ein längerer und von vielen Faktoren beeinflusster Prozess. Zu diesen Faktoren können z. B. auch technische Uninformiertheit oder Gutgläubigkeit zählen. Es ist also grundsätzlich auch angesichts der verbreiteten Inhalte nicht davon auszugehen, dass sie bewusst weder für eine Öffentlichkeit noch für die Wissenschaft produziert werden (vgl. auch Eynon, Fry & Schroeder 2008 und Giaxoglou 2017).

Als relevantes Axiom im Bereich Social Media ist also die informierte Zustimmung besonders hervorzuheben. Ein zweites wichtiges Axiom ist die Frage der Anonymität respektive ggf. notwendigen Anonymisierung. Gerade bei öffentlichen Kommunikaten stellt sich die Frage, inwieweit eine Anonymisierung im Sinne der Urheber*innen ist, die einerseits als Autor*innen Rechte an Kommunikaten (Texten, Bildern, Text-Bild-Kombinationen) geltend machen können, denen andererseits aber auch aus ethischen Gründen die Schöpfungshöhe am eigenen Produkt zugestanden sein soll. Ein drittes wichtiges Axiom betrifft den Umgang mit unterschiedlichen semiotischen Ressourcen, etwa Bildern oder audiovisuellen Kommunikaten, die ebenso wie Textelemente konstitutiv für Social Media sind. Eine im Zuge einer

Anonymisierung vorgenommene Schwärzung beispielsweise ist eine erhebliche Veränderung am Kommunikat, die zu begründen wäre.

Best Practice

Welche Möglichkeiten gibt es nun, eine informierte Zustimmung der Nutzer*innen von Social Media einzuholen? Es bietet sich hier eine direkte Kontaktaufnahme an. Je nach Anlage der Studie (qualitativ/quantitativ) ist damit ein erheblicher Aufwand verbunden, etwa mit Blick auf die Kontaktadressen-Recherche (im besten Falle E-Mail-Adressen). Abhängig von der Plattform ist auch eine Adressierung über plattformeninhärente direkte Nachrichtenkanäle denkbar. Die Rücklaufquote ist jedoch schwer im Vorhinein zu kalkulieren. Unsere Annahme ist: Je besser der*die Forschende bereits auf der Plattform vernetzt ist, desto höher ist die Chance, auch über die direkte Adressierung mit Nutzer*innen in Kontakt treten zu können. Es gibt aber weitere Faktoren, die zu berücksichtigen sind. So beschränken manche Nutzer*innen die Möglichkeiten der direkten Kontaktaufnahme. Auf Twitter gibt es die Funktion, selbst die Antwortmöglichkeiten auf Tweets einzuschränken, wenngleich eine Kontaktaufnahme sichtbar auf der Timeline ohnehin nicht ratsam ist. Zu berücksichtigen ist bei allen Versuchen der Kontaktaufnahme, dass für ein- und dieselbe Studie sehr unterschiedliche Strategien zum Einsatz kommen und dokumentiert werden müssen. Es muss sichergestellt werden, dass trotz dieser Heterogenität allen „Studienteilnehmer*innen“ die gleichen Informationen zur Verfügung gestellt werden.

Beispiel 1: Verzicht

Je nach Forschungsinteresse kann die oben beschriebene Vorgehensweise durchaus auch als unverhältnismäßig eingestuft werden. Siever (2015: 24), die sich in ihrer Dissertation zu multimodaler Kommunikation im Social Web am Beispiel der Foto-Community Flickr widmet, verzichtet in ihrer produktorientierten Analyse gänzlich auf Bilder, auf denen identifizierbare Personen abgebildet sind. Im Mittelpunkt ihres Interesses steht die Praktik des Social Tagging, also Fragen zu kommunikativen Funktionen der Notizenkommunikation, zu pragmatischen Notiz-Bild-Relationen und zu Notiz-Dialogen. Es geht ihr darum, die laienhafte Dokumentation von Bildern zu untersuchen und die Einbettung von Sprache in einem multimodalen Kontext zu beschreiben. Vor diesem Hintergrund ist die Entscheidung, Personenabbildungen aus dem Untersuchungsmaterial zu exkludieren, ethisch praktikabel und wissenschaftlich nachvollziehbar.

Beispiel 2: Kontaktaufnahme

Gerade bei höchst persönlichen Texten kann der oben problematisierte Verweis auf eine prinzipiell öffentliche Zugänglichkeit keine Option sein, forschungsethische Implikationen auszublenden. In seinen Arbeiten zu Blogs von Glioblastom-Patient*innen und/oder deren Angehörigen verweist Meier-Vieracker auf die unterschiedlichen Motive für die krankheitsbegleitenden Erzählungen, die in den Texten der inzwischen verstorbenen Schreiber*innen teilweise explizit formuliert wurden: „Information von Freunden und Kollegen, [...] Unterstützung anderer Betroffener, [...] individuelle [...] Bewältigung ihrer je eigenen Krankheits- und Trauererfahrung“ (Meier-Vieracker 2021: 172). Meier-Vieracker arbeitet mit einem Korpus aus 11 Blogs, für zwei Blogs ist es ihm gelungen, Kontakt mit den Autor*innen aufzunehmen und die Erlaubnis für die wissenschaftliche Analyse einzuholen. Bei den anderen neun Blogs ließen sich keine Kontaktdaten finden oder Anfragen wurden nicht beantwortet, also auch nicht ausdrücklich abgelehnt, so dass diese Daten mit Blick auf die öffentliche Verfügbarkeit in das Korpus integriert wurden. Dass ein Blog nach der Kontaktaufnahme offline ging, wurde von Meier-Vieracker als „verweigerte Zustimmung“ gewertet. Entsprechend sind diese Daten aus dem Korpus extrahiert worden. In wissenschaftlichen Publikationen zu diesen Korpus-Daten wird mit Referenz auf die forschungsethisch hohen Anforderungen an den Umgang mit höchstpersönlichen Daten konsequent anonymisiert, URLs werden nicht angegeben. Die Kontaktaufnahme mit Blog-Autor*innen liegt auch den Arbeiten von Marx (2019) sowie Marx und Tienken (2021) über Pflege und Trauer zugrunde. Die Kontaktaufnahme führte nicht nur dazu, dass die Daten umfangreich linguistisch analysiert werden konnten und können, sondern auch zu Kooperationen im Bereich Lehre und Third Mission.⁶

Mit Blick auf die Frage der ggf. notwendigen Anonymisierung kann diese Frage direkt in diesem Rahmen mit den Urheber*innen thematisiert werden. Diese Vorgehensweise sollte in Publikationen transparent geschildert werden.

Beispiel 3: Spenden-Korpora

Ein eleganter Weg zur Akquise von Social-Media Daten sind sogenannte Datenspenden. Die informierte Zustimmung kann dabei geradezu klassisch sichergestellt werden, weil Proband*innen vorher über die Studienzwecke informiert werden und ihre Einwilligung aufbauend auf diesen Informationen geben können. Auf diese Weise ist z. B. das DiDi-Korpus entstanden, das aus ca. 40.000 Facebook-Texten besteht (Glaznieks & Frey 2020). Dass diese Vorgehensweise

⁶ So waren die Autor*innen des Blogs bereits Gäste in unterschiedlichen akademischen Lehrveranstaltungen. Eine wissenschaftlich gerahmte Vernissage ist für das Jahr 2022 geplant.

auch die Erhebung von verlässlichen Metadaten wie Alter, Geschlecht, regionale Herkunft u. a. ermöglicht, ist ein willkommener Nebeneffekt. Auch für die Erhebung von nicht zugänglichen Messengerdienst-Daten ist diese Methode ein ethisch gut vertretbarer Weg. So wird die Mobile Communication Database (MoCoDa und MoCoDa2) sukzessive aus Datenspenden generiert (Beißwenger et al. 2020). Bevor Spender*innen Daten in das System einspeisen können, stimmen sie der weiteren Nutzung für die Wissenschaft aktiv zu.

Durch die Datenspenden wird die Identität der Spender*innen für Forscher*innen zu keinem Zeitpunkt zugänglich, Datenspende*rinnen haben selbst die Möglichkeit, die für die Analyse wichtigen Metadaten in eine Eingabemaske zu integrieren. Die Frage der Anonymisierung wird somit in diesem Verfahren umgangen.

Im Hinblick auf quantitative Zugänge zu Diskursen und natürlich auch die Bereitstellung ganzer Korpora ist Anonymisierung keine Frage der Abwägung. Ein Rechtsgutachten, das vom DFG-Netzwerk „Empirikom-Empirische Erforschung computervermittelter Kommunikation“ im Zusammenhang mit der Veröffentlichung des Dortmunder Chatkorpus in der Clarin-D-Infrastruktur eingeholt wurde, ist in dieser Hinsicht sehr eindeutig. So ist nach „Ansicht der Gutachter [...] eine Personenbeziehbarkeit nach dem strengen Datenschutzverständnis erst dann nicht mehr gegeben, wenn die Herstellung des Personenbezugs einen für die jeweilige Stelle klar unverhältnismäßigen Aufwand erfordern würde, von dem nicht zu erwarten ist, dass er getrieben würde“ (Beißwenger et al. 2020: 168–169). Weiterhin wird mit Bezug auf das Chatkorpus angemerkt, dass Standardmaßnahmen, wie die Randomisierung von Hostnamen; Ersetzung der Nicknames, Ortsnamen, Expertennamen; Anonymisierung, Entfernung oder Permutierung der Zeitstempel, zur Minimierung rechtlicher und ethischer Risiken durchgeführt werden sollten (Beißwenger et al. 2020: 169). Den Hinweis, dass Personen mit Zugriff auf die Originaldaten nach wie vor auch den Bezug zu Personen herstellen könnten, nehmen wir zum Anlass, auf eine „Best Practice“ von Hartmann (2020) zu verweisen, der Rohdaten zwar veröffentlicht hat, diese aber aus rechtlichen und ethischen Gründen nur als bag-of-word-Listen und n-grams zugänglich macht, was eine interessante Art der Anonymisierung ist, die sich als direkte Verlängerung aus den Methoden der Analyse ergibt.

4.2 Fallbeispiel 2: Wikipedia

Fallweise und mediale Konstellation

Die Online-Enzyklopädie Wikipedia ist eine lexikonartige Sammlung von Informationen und zugleich ein laufendes soziales Projekt, um die Erstellung der Einträge

zu koordinieren. Wikipedia ist also zugleich frei verfügbares Nachschlagewerk und offenes, auf Freiwilligenengagement basierendes Editier- und Schreibvorhaben. Entsprechend mehrschichtig gestaltet sich die fallweise und mediale Konstellation. Im Folgenden wird ein Ansatz vorgeschlagen, um Wikipedia zu erschließen und in ihren rechtlichen bzw. ethischen Spezifika zu reflektieren.⁷

Und zwar wird vorgeschlagen, Wikipedia als Ansammlung miteinander verbundener Stationen zu betrachten, an denen die Autor*innen in einzelnen Handlungs- bzw. Interaktionsepisoden tätig werden. Stationen sind, um mit Giddens (1997: 171) zu sprechen, Orte, „an denen sich die Routineaktivitäten verschiedener Individuen überschneiden“. Stationen unterscheiden sich in ihrer Öffentlichkeit bzw. Privatheit und insbesondere die längerfristig am Projekt beteiligten User*innen teilen einen stationenweise differenzierten ‚sense of audience‘. Solche Stationen sind zum Beispiel ein einzelner Artikel oder die Diskussionsseite eines Artikels, die Seite zur Diskussion von Löschkandidaten oder die Seite, auf der über Autor*innen diskutiert wird, die für eines der Ämter im Projekt kandidieren. Auch offline stattfindende Stammtische oder Hackathons können Stationen darstellen. Im Sinn von Nissenbaums (2011: 2) Konzept von Privatheit als „contextual integrity“ unterschieden sich die Stationen gemäß der hier erwartbaren Akteure, der Arten von Informationen und Kommunikationsnormen. Das heißt, obwohl in Wikipedia alle Seiten ohne Zugangsschutz gleich öffentlich und auffindbar waren, hat sich im Projekt ein Verständnis von verschiedenen öffentlichen Bereichen mit differenzierten Aufmerksamkeitshorizonten und Leser- bzw. Teilnehmerkreisen konventionalisiert. Neben einem potentiell sehr großen Publikum, für das die Stationen also prinzipiell erreichbar sind, ist das empirische Publikum von Wiki-Seiten variabel. Von diesen potentiellen und empirischen Publika sind wiederum das intendierte Publikum, also die vorgestellte Leser- und Autorenschaft einer Seite bzw. eines dort gemachten Edits sowie das mit bestimmten Botschaften adressierte Publikum zu unterscheiden. Öffentlichkeit in Wikipedia kann dann als abgestufte *publicness* von Stationen verstanden werden, deren potentielle Wahrnehmungschancen im Prinzip zwar gleich hoch sind, ihre empirische Rezeptionswahrscheinlichkeit sowie ihre intendierten und adressierten Publika indessen variieren.⁸

⁷ Der folgende Teil ist übertragen bzw. übernommen aus Pentzold (2015 und 2017).

⁸ Die empirische Öffentlichkeit einzelner Wiki-Seiten und Botschaften ist zudem potentiell skalierbar und kann relativ zu verschiedenen Faktoren auch in kurzer Zeit stark zunehmen, etwa wenn sie in (massen-)medialen Publikationen auftauchen. Spoerri (2007) Untersuchung der Top 100 an aufgerufenen Seiten zeigte etwa, dass zwischen September 2006 und Januar 2007 25% der monatlich meist aufgerufenen Seiten auch nur in diesem Monat darunter zählten und dass nur 39 in allen Monaten prominent waren.

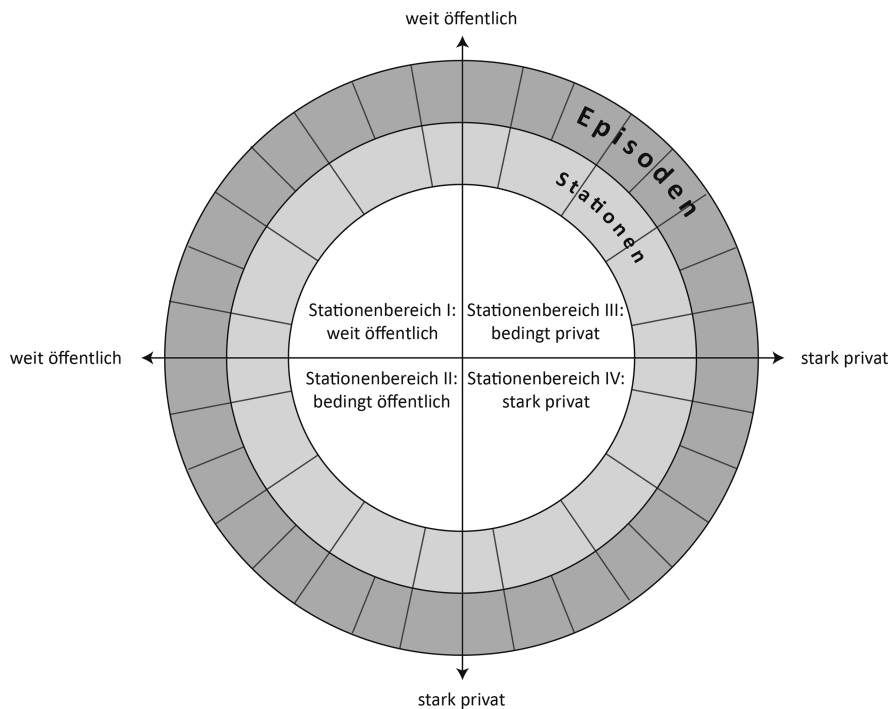


Abb. 2: Stationen in Wikipedia. Quelle: Pentzold (2015: 71).

Im Prinzip lassen sich vier Stationenbereiche entsprechend ihrer vergleichsweise großen bzw. kleinen empirischen sowie breiten oder engen intendierten und adressierten Publika trennen (Abb. 2).

- (1) weit öffentliche Bereiche der Tagungen, der Wikipedia-Artikel und der Wiki-Seiten, über die das Organisieren des Projekts abläuft,
- (2) öffentliche Bereiche der Workshops und Diskussionen zu Wikipedia-Artikeln,
- (3) der private Bereich der Stammtische, Exkursionen, Nutzerseiten, Nutzerdiskussionsseiten und anderweitig betriebenen Webpräsenzen (z. B. Blogs, Twitter-Profile, Webseiten, andere Wiki-Projekte) und
- (4) der sehr private Bereich von persönlichen Gesprächen und Mitteilungen zwischen User*innen.

Identifikation relevanter Axiome und Regularien

Das erste forschungsethisch zu reflektierende Spezifikum von Wikipedia-Stationen stellt sich bezüglich der Öffentlichkeit des Beteiligens einerseits und der Anonymität der Beteiligten andererseits (Sveningsson-Elm 2009).⁹ Forschungspraktisch ist die in Wikipedia gewährleistete freie Zugänglichkeit der Daten, also das Fehlen technologisch implementierter Zugriffsbarrieren und Vertraulichkeitsreglements, von Vorteil. Wikipedia ist zum einen ein öffentliches und selbstdokumentierendes Projekt und daher der überwiegende Teil der Wiki-Seiten mit enzyklopädischen Inhalten, Aktivitätsaufzeichnungen, Diskussionen und Selbstauskünften ohne Hindernisse auch unangemeldeten Besucher*innen zugänglich. Zum anderen geht ein Teil der hochaktiven Nutzenden offen mit ihrer Teilnahme und Autorenschaft um. Manche Autor*innen geben auf ihren Benutzerseiten an, bei welchen Artikeln sie mitgewirkt hatten, Nutzernamen werden aus dem Projekt herausgetragen und tauchen beispielsweise in der journalistischen Berichterstattung und in online-medialen Publikationen auf.

Die meisten Beiträge in Wikipedia liegen archiviert vor und sind in der internen Wiki-Suche als auch in externen Suchmaschinen referenziert. Selbst wenn daher Pseudonyme oder IP-Adressen nochmals maskiert werden würden, können wörtliche Zitate über Suchfunktionen ohne Weiteres auf die jeweiligen angemeldeten User*innen bzw. die in der Datenbank abgespeicherte IP-Adresse rückgeführt, mit anderen Daten kombiniert und eventuell zu anderen als den mit der Studie verfolgten Zwecken ausgewertet werden (Tilley & Woodthorpe 2011). Darüber hinaus ist die öffentliche Dokumentation, welcher Account welchen Edit getätigt hat, gemäß der auf der Plattform gültigen Auslegung der Creative-Commons-Attribution-ShareAlike 3.0 Unported (CC-BY-SA) eine lizenzrechtliche Verpflichtung. Entsprechend wurden die Editor*innen in den Nutzungsbedingungen der Wikimedia Foundation (WMF) als US-amerikanischer Trägerorganisation der Plattform darauf hingewiesen, dass lizenzkonform Nennung ihres Nutzernamens bzw. ihrer IP-Adresse

- i „Durch einen Hyperlink auf die Seite(n) (sofern möglich) oder die Angabe der URL der Seite(n), die Sie weiternutzen (da jede Seite über eine Versionsgeschichte verfügt, in der sämtliche Autoren und Bearbeiter aufgeführt sind);
- ii Durch Hyperlink (sofern möglich) oder URL auf eine andere, stabile Online-Kopie, die frei zugänglich ist, die Lizenz erfüllt und die die Namensnennung der Autoren auf eine Weise gewährleistet, die gleichwertig zu der auf den Projekt-Webseiten ist; oder

⁹ Dieser Abschnitt ist übernommen und z. T. übertragen aus Pentzold (2015: 73–79).

- iii Durch eine Liste aller Autoren (bitte beachten Sie allerdings, dass jede Liste von Autoren gefiltert werden kann, um sie um sehr kleine oder irrelevante Beiträge zu bereinigen).“
erfolgen müsste (WMF:Terms of Use 2021).¹⁰

Forschungsethisch ist indessen zu reflektieren, dass den Beitragenden versichert wird, sie haben ein Recht auf Anonymität. Ungeachtet der codegemäß nivellierten Zugänglichkeit und der an die Editierenden herangetragenen Hinweise über die Öffentlichkeit und Zurechenbarkeit ihrer Tätigkeiten kann forschungsethisch nicht übergangen werden, dass die User*innen ihre Beiträge pro Station an verschiedenen große und verschieden zusammengesetzte Publika adressieren und nicht ohne Weiteres davon auszugehen ist, dass sie hierbei zum Beispiel auch eine (fachwissenschaftliche) Öffentlichkeit im Blick haben (Hudson & Bruckman 2004).

Hinzu kommt, dass Wikipedia neben der Option, per IP-Adresse oder pseudonym tätig zu werden, Verfahren bereitstellt, um Klarnamen zu löschen und Pseudonyme von bürgerlichen Namen zu trennen. So können Administrator*innen einzelne kritische Versionen entfernen lassen und Oversighters Edits, also einzelne gespeicherte Eingriffe in das Wiki verbergen; außerdem können mit einer IP registrierte Eingriffe nachträglich Benutzernamen zugeordnet werden und für das unerlaubte ‚Outing‘ verantwortlich gemachte Nutzer*innen können gesperrt werden; Unterlassungs- und Schadenersatzklagen wurden bereits angedroht. Mit ihrer 2008 vom *Board of Trustees* erlassenen, inzwischen aber in dieser Form widerrufenen, *Data Retention Policy* erklärte die WMF zudem, nur „the least amount of personally identifiable information consistent with maintenance of its services, with its privacy policy, or as required by state or federal legal provisions under United States of America law“ (WMF:Data Retention Policy 2008) von den Nutzer*innen erheben zu wollen.

Das Gewährleisten informationeller Selbstbestimmung der Wikipedia-Autor*innen umfasst somit auf der Plattform das Aufklären der Nutzenden über ihre datenschutzrelevanten öffentlichen Aktionen und das Zusichern von Anonymität unter der

10 In der Privacy Policy der WMF war dahingehend zu lesen: „All Projects of the Wikimedia Foundation are collaboratively developed by its users using the MediaWiki software. Anyone with Internet access (and not otherwise restricted from doing so) may edit the publicly editable pages of these sites with or without logging in as a registered user. By doing this, editors create a published document, and a public record of every word added, subtracted, or changed. This is a public act, and editors are identified publicly as the author of such changes. All contributions made to a Project, and all publicly available information about those contributions, are irrevocably licensed and may be freely copied, quoted, reused and adapted by third parties with few restrictions.“ WMF:Privacy Policy (2011).

Erwartung, dass die Nutzer*innen kompetent von ihren Persönlichkeitsrechten Gebrauch machen (Schmidt 2012). Der erlangte bzw. beibehaltene Grad an Anonymität steht so gesehen in der Verfügung der einzelnen Beitragenden.¹¹

Aufbauend auf den forschungspraktischen Möglichkeiten der freien Zugänglichkeit der Inhalte und der forschungsethischen Forderung nach Anonymität der Autor*innen betrifft die zweite forschungsethische Schwierigkeit das Einholen informierter Zustimmung von anonym bzw. pseudonym agierenden Nutzer*innen, die Beiträge auf der Plattform leisten und die dabei unzuverlässig über ihre individuelle Nutzerseite erreichbar sind.

Grundsätzlich problematisch ist dabei, dass informierte Zustimmung forschungsethisch plausibel als Aushandlung stattfinden soll, in der das Vorgehen und die Erwartungen der Forschenden mit den Ansprüchen der in die Untersuchung einbezogenen Personen im Fortgang der Studie abzugleichen sind. Auf diese Weise ist der Prozesshaftigkeit des methodischen Erkenntnisvorgangs Rechnung getragen und zugleich die Freiwilligkeit der Teilnahme und die angemessene Fundierung der Entscheidung der studierten Akteure, fortwährend teilzunehmen, berücksichtigt (Hopf 2005: 592). In der Praxis ist dies vor allem im Sammeln von Dokumenten und dem retrospektiven Beobachten dokumentierter Aktionen nur bedingt möglich. In beiden Fällen können die betroffenen anonymen oder pseudonymen Editor*innen nur nachträglich vermittels ihrer möglicherweise vorhandenen Nutzerseite angefragt und um Zustimmung gebeten werden, wobei diese Kontaktaufnahme dann wiederum nicht privat bleibt, sondern in ihrer Form öffentlich ist, Öffentlichkeit für die erfolgte bzw. eventuell zu erfolgende Beobachtung und Analyse herstellt und so bereits eine Verbindung zwischen der Studie und den Nutzer*innen schafft. Erschwert wird der Vorgang noch dadurch, dass die Nutzenden dynamischer IP-Adressen keinen dauerhaften Zugriff auf ihre Benutzer- und ihre Benutzerdiskussionsseite haben, dass Autor*innen unter statischen IP-Adressen ihre Diskussionsseite häufig nicht verfolgen und viele der einmal aktiven Accounts verwaist sind. In allen drei Fällen bleiben die Anfragen erwartungsgemäß unbeantwortet.

¹¹ Die WMF Privacy Policy machte Einschränkungen dieser Zusicherung bzgl. der Nutzung von E-Mails, Mailinglisten und dem Open Ticket Request System (OTRS), mit dem Anfragen an das Projekt organisiert wurden.

Best Practice

In der Mitte der Studie, deren forschungsethische Kasuistik hier skizziert wird, stand die Frage nach den Praktiken und den institutionellen Bedingungen online-medialen, gelingenden und produktiven, also das Generieren informationeller Güter bezweckenden, Zusammenarbeitens.¹² Diese wurde exemplarisch und exponiert am Beispiel der freien Gemeingüterproduktion studiert, genauer gesagt am Fall der Online-Enzyklopädie Wikipedia (Pentzold 2016). Die kontrastive Erhebung kombinierte mehr oder minder öffentliche Stationen gemäß den verfügbaren Zahlenwerten zu Besucher*innen, Editierenden und Edits mit im Projekt ausgebildeten und an neue Nutzer*innen weitergegebenen Überzeugungen, welche Stationen stark oder schwach in der Aufmerksamkeit der Wikipedia-internen und der Wikipedia-externen Öffentlichkeit ständen. Heuristisch konnte über das quantitative Bestimmen unterschiedlich großer empirischer Wikipedia-Öffentlichkeiten hinaus das Gruppieren der Stationen in softwareseitig und funktional getrennte Namensräume und der damit zusammenhängenden vagen Einteilung verschieden öffentlicher, weil unterschiedlich intensiv editierter und von User*innen aufgerufener, Stationen eingesetzt werden. Die fallweisen Entscheidungsraster orientieren sich an McKee und Porter (2008).

Im Blick auf die forschungsethischen Entscheidungen, um Anonymität wenn möglich zu wahren – Darf eine Situation beobachtet werden bzw. darf ein Dokument gesampelt werden? Darf das Material zitiert werden? –, waren hier pro Stationenbereich die Öffentlichkeitsdimensionen zwischen weit öffentlich, bedingt öffentlich, bedingt privat und stark privat sowie die namentliche bzw. anonyme Kennung der Akteure zwischen namentlich-persönlich, pseudonym-identifizierbar, pseudonym-nicht identifizierbar und anonym heuristisch zu bestimmen. Darüber hinaus war als zusätzlicher Faktor der Grad an Interaktion von Feld und Beobachter*in zwischen synchron wechselseitig, asynchron wechselseitig, einseitig und nicht vorhanden einzustufen.

Unberücksichtigt blieben in diesem Entscheidungsraster und im folgenden Entscheidungsraster somit andere mögliche Faktoren wie die potenzielle Vulnerabilität der einbezogenen Subjekte oder die Sensibilität der jeweiligen Episode, da diese nicht für Stationenbereiche generell bestimmt werden konnten. Das Entscheidungsraster und die Entscheidungsheuristik, exemplarisch für den weit öffentlichen Stationenbereich, gestaltete sich folglich wie in Abb. 3 dargestellt. Ein Beispiel für diesen Stationenbereich wäre die Episode einer Wikipedia-internen Reviewdiskussion, bei der ein*e Autor*in einen für gewöhnlich hauptsächlich

12 Dieser Teil ist übernommen und z. T. übertragen aus Pentzold (2015: 65–66, 75–76 u. 77–79).

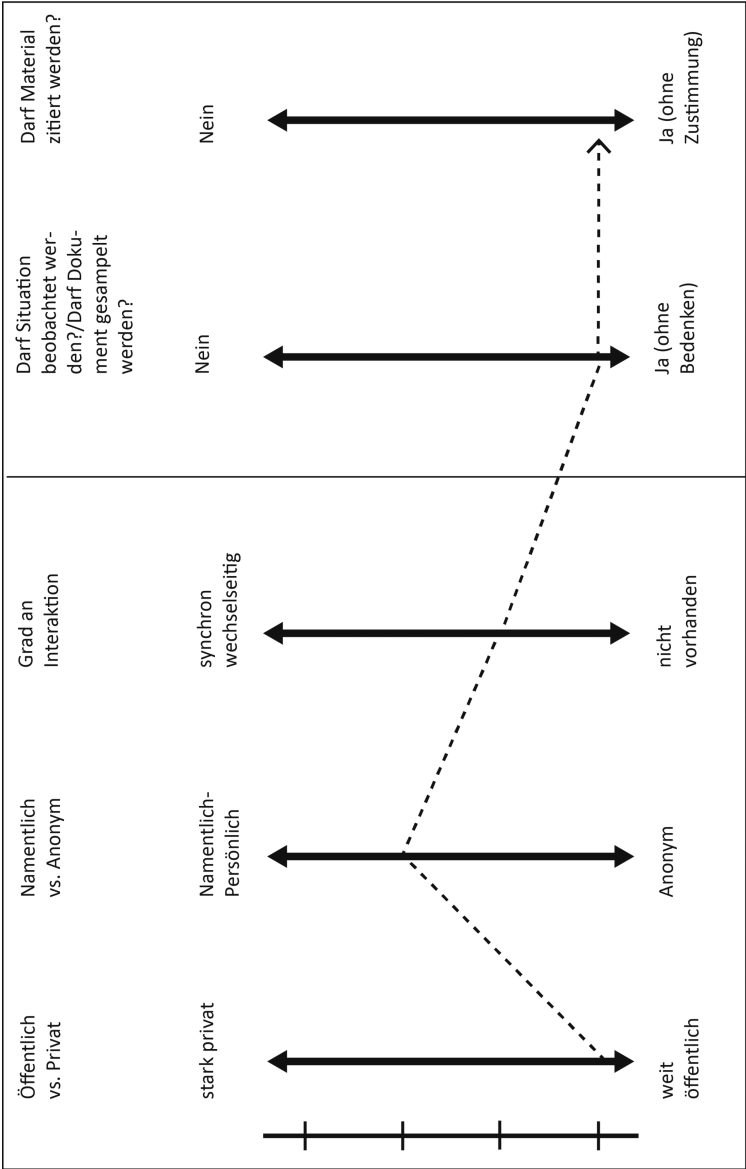


Abb. 3: Entscheidungsraster Öffentlichkeit vs. Anonymität. Quelle: Pentzold (2015: 76).

von ihm editierten Artikel auf der Station einer eigens dafür eingerichteten Wiki-Seite zur Diskussion stellt, um so Feedback von anderen Autor*innen einzuholen und den Artikel zu verbessern.

Zusammenfassend wurden folgende forschungsethische Entscheidungen getroffen: Es wurden wörtliche Belegstellen in der Ergebnisdarstellung ohne Zustimmung nur aus den beiden Stationenbereichen zitiert, die von den Nutzer*innen als stark bis bedingt öffentlich eingestuft wurden. Konkret wurde also darauf verzichtet, ohne Zustimmung Zitate aus den Benutzer- und Benutzerdiskussion-Namensräumen, den Stammtischen und anderen plattformenunabhängigen Webpräsenzen anzuführen, wenn diese erkennbar von den jeweiligen Benutzer*innen nicht für größere (Wikipedia-)Publika längerfristig öffentlich gehalten wurden. Das von der Lizenz geforderte Nennen der jeweiligen Autor*innen wurde dadurch abgegolten, dass im Forschungsbericht ein Link auf die jeweilige zitierte Wiki-Seite angegeben wurde.

Das Verfahren zur Sicherung von informierter Zustimmung erfolgte analog: Hinsichtlich der forschungsethischen Entscheidungen, um informierte Zustimmung wenn nötig zu gewähren – Kann im teilnehmenden Beobachten informierte Zustimmung eingeholt werden? Muss sie eingeholt werden? –, waren hier ebenso wie im obigen Entscheidungsraster pro Stationenbereich die Öffentlichkeitsdimensionen zwischen weit öffentlich, bedingt öffentlich, bedingt privat und stark privat sowie die namentliche bzw. anonyme Kennung der Akteure zwischen namentlich-persönlich, pseudonym-identifizierbar, pseudonym-nicht identifizierbar und anonym heuristisch zu bestimmen.

Darüber hinaus war als zusätzlicher Faktor der Grad an Interaktion von Feld und Beobachter*in zwischen synchron wechselseitig, asynchron wechselseitig, einseitig und nicht vorhanden einzustufen. Das Entscheidungsraster und die Entscheidungsheuristik, exemplarisch für den bedingt privaten Stationenbereich, gestaltete sich entsprechend wie in Abb. 4 dargestellt. Ein Beispiel für diesen Stationenbereich wäre die Episode einer Wikipedia-internen Diskussion, bei der angemeldete und unangemeldete Autor*innen die Station der Diskussionsseite eines*r Autors*in gebrauchten, um die Berechtigung einer Benutzersperre wegen des Vorwurfs, ein*e Editor*in habe Einträge vandalisiert, zu debattieren.

Zusammenfassend wurden folgende forschungsethische Entscheidungen getroffen: Es wurde eine offene Beobachtung durchgeführt, insoweit die Plattform ein Offenbaren des beobachtenden Teilnehmers und seiner Ziele zuließ. Das hieß, als Pseudonym wurde der Klarname gewählt und auf der entsprechend verlinkten Nutzerseite fanden sich Informationen zu dem Forschungsvorhaben und es wurde die Möglichkeit gegeben, per E-Mail auch außerhalb von Wikipedia in Kontakt zu treten. Jeder Beitrag in Wikipedia wurde angemeldet getätigt und jeder Beitrag zu Diskussionsseiten signiert, sodass die Aktivitäten des Accounts nachverfolgt werden konnten. Wenn Interaktionen mit angemel-

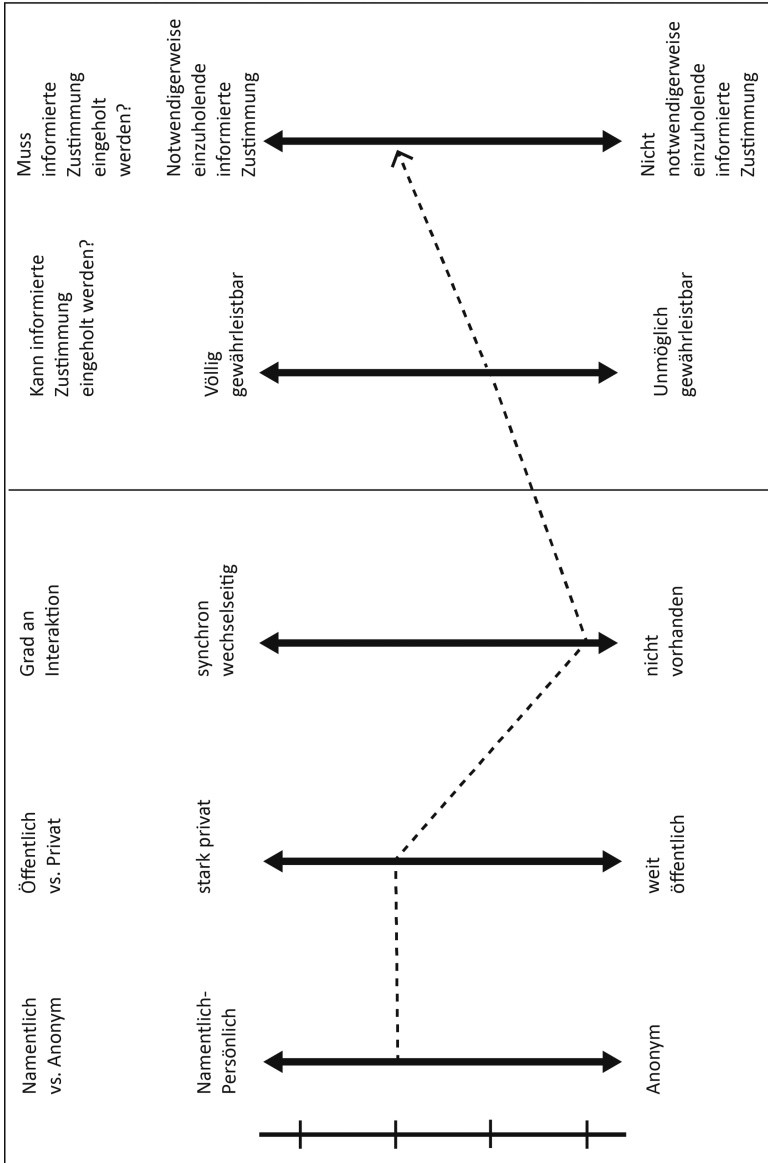


Abb. 4: Entscheidungsraiser Anonymität vs. Informierte Zustimmung. Quelle: Pentzold (2015: 78).

deten Nutzer*innen stattfanden, wurden diese auf das Forschungsvorhaben aufmerksam gemacht und den angefragten Autor*innen wurde das Angebot gemacht, den Kontakt per Mail oder Telefon fortzusetzen.

5 Fazit

Unsere Ausführungen setzen an einem Dilemma an, dem die Analyse digitaler Diskurse gegenübersteht: Es gibt das berechtigte Interesse, bei der Datenerhebung und Datenanalyse juristisch konform und ethisch integer zu agieren. Gleichzeitig ist unser Forschungsgegenstand so dynamisch, dass es kaum möglich ist, längerfristige Handreichungen und Entscheidungsvorlagen zu geben. Vielmehr ist die Adressierung rechtlicher und ethischer Fragen an Ableitungen geknüpft, ggf. auch von gesetzlichen Grundlagen, die nicht dezidiert unseren Gegenstand regeln, aber Approximationen ermöglichen. Vergleichbar verhält es sich auch mit ethischen Fragen. Wir empfehlen, diese Aspekte transparent zu dokumentieren und potenziellen Forschungsteilnehmer*innen und Urheber*innen mit größtmöglicher Zugewandtheit zu begegnen, ggf. ethische Fragen mit ihnen gemeinsam abzuwägen und im Zweifelsfall selbstverständlich so zu handeln, dass ihre Persönlichkeit geschützt bleibt, selbst wenn rechtliche Vorgaben eine Veröffentlichung legitimieren. In diesem Vorgehen zeigt sich einmal mehr die Besonderheit bei der linguistischen Arbeit an digitalen Diskursen: Die Vorgehensweise von Nutzer*innen und die Vorgehensweise von Forschenden können in gewinnbringender Weise miteinander verschränkt und sogar verwoben werden.

Literatur

- AolR. 2019. *Association of Internet Researchers. Ethical Guidelines 3.0*. <https://aoir.org/ports/ethics3.pdf>. (18 June, 2021.)
- Apel, Jochen, Nils Föhles & Jonas Kratzke. 2017. *Forschungsdatenmanagement. Workshop*. <https://doi.org/10.11588/heidok.00023215>. (18 June, 2021.)
- Baym, Nancy. 2010. *Personal Connections in the Digital Age*. Cambridge: Polity.
- Beaulieu, Anne & Adolfo Estalella. 2012. Rethinking research ethics for mediated settings. *Information, Communication & Society* 15(1). 23–42.
- Beißwenger, Michael, Evelyn Ziegler, Marcel Fladrich & Wolfgang Imo. 2020. Mobile Communication Database 2 (MoCoDa2). In *Netzwerktag: Digitalisierung in den Geisteswissenschaften: Digital forschen | Digitales erforschen | Lehre digital gestalten*. https://duepublico2.uni-due.de/receive/duepublico_mods_00071353. (18 June, 2021.)

- Bennett, Lance & Barbara Pfetsch. 2018. Rethinking Political Communication in a Time of Disrupted Public Spheres. *Journal of Communication* 2(68). 243–253.
- Boellstorff, Tom, Bonnie Nardi, Cathy Pearce & T.L. Taylor (2012): *Ethnography and Virtual Worlds. A Handbook of Method*. Princeton: Princeton University Press.
- boyd, danah & Kate Crawford. 2012. Critical Questions for Big Data. *Information, Communication & Society* 15(5). 662–679.
- Bruckman, Amay. 2002. Studying the amateur artist. *Ethics and Information Technology* 4(3). 217–231.
- Bruns, Axel. 2019. *Are Filter Bubbles Real?* Cambridge: Polity.
- Buchanan, Elizabeth & Charles Ess. 2009. Internet Research Ethics and the Institutional Review Board: Current practices and issues. *Computers and Society* 39(3). 43–49.
- Castells, Manuel. 2009. *Communication Power*. Oxford: Oxford University Press.
- De la Durantaye, Katharina & Benjamin Raue. 2020. Urheberrecht und Zugang in einer digitalen Welt – Urheberrechtliche Fragestellungen des Zugangs für Gedächtnisinstitutionen und die Digital Humanities. *Recht und Zugang* 1(1). 83–107.
- Deppermann, Arnulf & Martin Hartung. 2012. Was gehört in ein nationales Gesprächskorpus? Kriterien, Probleme und Prioritäten der Stratifikation des „Forschungs- und Lehrkorpus Gesprochenes Deutsch“ (FOLK) am Institut für Deutsche Sprache (Mannheim). In Ekkehard Felder, Marcus Müller & Friedemann Vogel (eds.), *Korpuspragmatik*, 414–450. Berlin: De Gruyter.
- Eisenegger, Mark, Marlis Prinzing, Patrick Ettinger, & Roger Blum (eds.). 2021. *Digitaler Strukturwandel der Öffentlichkeit*. Wiesbaden: Springer VS.
- Ess, Charles. 2009. *Digital Media Ethics*. Cambridge: Polity.
- Eynon, Rebecca, Jenny Fry & Ralph Schroeder. 2008. The Ethics of Internet Research. In Nigel Fielding, Raymond M. Lee & Grant Blank (eds.), *The SAGE Handbook of Online Research Methods*, 23–41. Los Angeles et al.: Sage.
- Fraas, Claudia, Stefan Meier & Christian Pentzold. 2012. *Online-Kommunikation*. München: Oldenbourg.
- Gatto, Maristella. 2014. *Web as Corpus. Theory and Practice*. London: Bloomsbury.
- Glaznieks, Aivars & Jennifer-Carmen Frey. 2020. Das DiDi-Korpus: Internetbasierte Kommunikation aus Südtirol. In Konstanze Marx, Henning Lobin & Axel Schmidt (eds.), *Deutsch in Sozialen Medien. Interaktiv – multimodal – vielfältig*, 353–354. Berlin et al.: De Gruyter.
- GESIS. 2021. *Datenarchiv für Sozialwissenschaften*. <https://www.gesis.org/institut/abteilungen/datenarchiv-fuer-sozialwissenschaften>. (18 June, 2021.)
- Giaxoglou, Korina. 2017. Reflections on internet research ethics from language-focused research on web-based mourning: revisiting the private/public distinction as a language ideology of differentiation. *Applied Linguistics Review* 8 (2–3). 229–250.
- Giddens, Anthony. 1997. *Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung*. Frankfurt a. M. & New York: Campus.
- Güllich, Elisabeth & Lorenza Mondada. 2012. *Konversationsanalyse. Eine Einführung am Beispiel des Französischen*. Tübingen: Niemeyer.
- Hartmann, Stefan. 2020. Identifying discourse patterns in online comments on „politically correct“ language use. *Yearbook of the German Cognitive Linguistics Association* 8(1). 27–44.
- Heesen, Hendrik & Leonie Jüngels. 2021. Der Regierungsentwurf der Text und Data Mining-Schranken (§§ 44b, 60d UrhG-E). *Recht und Zugang* 1(2). 45–55.

- Hopf, Christel. 2005. Forschungsethik und qualitative Forschung. In Uwe Flick, Ernst von Kardoff & Ines Steinke (eds.), *Qualitative Forschung. Ein Handbuch*, 589–600. Reinbek b. Hamburg: Rowohlt.
- Hudson, James & Amy Bruckman. 2004. „Go away“: Participant objections to being studied and the ethics of chatroom research. *The Information Society* 20. 127–139.
- Klaus, Elisabeth & Ricarda Drüeke (eds.). 2017. *Öffentlichkeiten und gesellschaftliche Aushandlungsprozesse*. Bielefeld: transcript.
- Knuchel, Daniel & Janine Luth. 2018. Zugängliche Daten im Netz. Herausforderungen beim Korpusaufbau mit besonderer Berücksichtigung rechtlicher und ethischer Aspekte. *Osnabrücker Beiträge zur Sprachtheorie* (OBST) 92. 31–44.
- Lexa, Carsten. 2020. Durchgelesen: Das steht in den AGB von Tik Tok. Blogbeitrag *Basic Thinking* vom 23. Juli 2020 <https://www.basicthinking.de/blog/2020/07/23/tik-tok-agb-analyse/>. (18 June, 2021.)
- Marx, Konstanze. 2019. Von #Gänsehaut bis #esreicht – Wie lässt sich ein Territorium neuer Sagbarkeit konturieren? Ein phänomenologischer Zugang. In Ludwig Eichinger & Albrecht Plewnia, A. (eds.), *Neues vom heutigen Deutsch. Empirisch-methodisch-theoretisch. Jahrbuch 2018*, 245–264. Berlin & New York: De Gruyter.
- Marx, Konstanze & Susanne Tienken. 2021. Trost und Trösten. In Christian Braun (ed.), *Sprache des Sterbens, Sprache des Todes*, 141–164. Berlin & Boston: De Gruyter.
- McKee, Heidi & James Porter. 2008. *The Ethics of Digital Writing Research: A Rhetorical Approach*. New York: Peter Lang.
- Meier-Vieracker, Simon. 2021. „Ach, heute wurde mir zum ersten Mal ein Sitzplatz angeboten im Tram“ – Musterhaftigkeiten in digitalen Sterbenarrationen. In Christian Braun (ed.), *Sprache des Sterbens, Sprache des Todes*, 165–184. Berlin & Boston: De Gruyter.
- Meier-Vieracker, Simon. I. Dr. immer noch chemo – Zeitlichkeit in digitalen Krankheitserzählungen. In Marina Iakushevich, Marina & Yvonne Ilg & Theresa Schnedermann (eds.), *Linguistik und Medizin*. Berlin & Boston: De Gruyter (= Linguistik – Impulse & Tendenzen).
- Münch, Vanessa. (Dissertation in Vorbereitung): Multimodale Analyse metapragmatischer Positionierung in sozialen Medien. Ein Vergleich des deutschen und französischen Körperdiskurses.
- Neidhard, Friedhelm. 1994. Öffentlichkeit, öffentliche Meinung und soziale Bewegungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Sonderheft 34. Opladen: Westdeutscher Verlag.
- Nissenbaum, Helen. 2011. *Privacy in Context. Technology, Policy, and the Integrity of Social Life*. Palo Alto: Stanford University Press.
- Papacharissi, Zizi. 2010. *A Private Sphere*. Cambridge: Polity.
- Pentzold, Christian. 2015. Forschungsethische Prämissen und Problemfelder teilnehmenden Beobachtens auf Online-Plattformen. In Axel Maireder, Julian Ausserhofer, Christina Schumann & Monika Taddicken (eds.), *Digitale Methoden in der Kommunikationswissenschaft*, 53–77. Digital Communication Research.
- Pentzold, Christian. 2016. *Zusammenarbeiten im Netz. Praktiken und Institutionen internetbasierter Kooperation*. Wiesbaden: Springer VS.
- Pentzold, Christian. 2017. „What are these researchers doing in my Wikipedia?\": Ethical Premises and Practical Judgment in Internet-Based Ethnography. *Ethics and Information Technology* 19(2). 143–155.

- QualiService. 2021. *Leitlinien*. <https://www.qualiservice.org/de/leitlinien.html>. (18 June, 2021.)
- Rolfes, Louis & Luisa Wendel. 2018. Die Veröffentlichung von Korpora amtlicher Werke zu Forschungszwecken aus urheber- und datenbankrechtlicher Sicht. *Zeitschrift für Urheber- und Medienrecht* 2018. 590–600.
- Schmidt, Jan. 2012. Persönliche Öffentlichkeiten und informationelle Selbstbestimmung im Social Web. In Jan Schmidt & Thilo Weichert (eds.), *Datenschutz. Grundlagen, Entwicklungen und Kontroversen*, 215–225. Bonn: Bundeszentrale für politische Bildung.
- Siever, Christina. 2015. *Multimodale Kommunikation im Social Web*. Frankfurt a.M.: Peter Lang.
- Steuernagel, Armin, Bruno Frey & Jonas Friedrich. 2017. *Zum Lügen gezwungen*. Zeit online vom 1. März 2017, 14:13 <https://www.zeit.de/wirtschaft/2017-02/allgemeine-geschaeftsbedingungen-verbraucherschutz-zertifizierung-alternativen>. (18 June, 2021.)
- Spoerri, Anselm. 2007. What is Popular on Wikipedia and Why? *First Monday* 12 (4). <http://firstmonday.org/ojs/index.php/fm/article/view/1765/1645>. (18 June, 2021.)
- Sveningsson-Elm, Malin. 2009. How Do Various Notions of Privacy Influence Decisions in Qualitative Internet Research? In Annette Markham & Nancy Baym (eds.), *Internet Inquiry*, 69–87. London: Sage.
- Tilley, Liz & Kate Woodthorpe. 2011. Is it the end for anonymity as we know it? A critical examination of the ethical principle of anonymity in the context of 21st century demands on the qualitative researcher. *Qualitative Research* 11(2). 197–212.
- Welcherer, Peter. 2020. *Journalistische Praxis: Digitale Recherche. Verifikation und Fact Checking*. Wiesbaden: Springer.
- WMF Data Retention Policy. 2008. http://wikimediafoundation.org/wiki/Data_retention_policy. (12 December, 2008.)
- WMF Privacy Policy. 2011. http://meta.wikimedia.org/wiki/Privacy_policy. (20 December, 2011.)
- WMF:Terms of Use. 2012. http://wikimediafoundation.org/wiki/Terms_of_Use. (18 June, 2021.)