

Bayer, Sonja; Breuer, Johannes; Lösch, Thomas; Goebel, Jürgen W.
Nutzung von Social-Media-Daten in der Bildungsforschung. Version 1.2

Frankfurt am Main : DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation 2021, 28 S. -
(forschungsdaten bildung informiert; 9)



Quellenangabe/ Reference:

Bayer, Sonja; Breuer, Johannes; Lösch, Thomas; Goebel, Jürgen W.: Nutzung von Social-Media-Daten in der Bildungsforschung. Version 1.2. Frankfurt am Main : DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation 2021, 28 S. - (forschungsdaten bildung informiert; 9) - URN: urn:nbn:de:0111-pedocs-221219 - DOI: 10.25656/01:22121

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-221219>

<https://doi.org/10.25656/01:22121>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-sa/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Abwandlungen und Bearbeitungen des Werkes bzw. Inhaltes anfertigen, solange sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und die daraufhin neu entstandenen Werke bzw. Inhalte nur unter Verwendung von Lizenzbedingungen weitergeben, die mit denen dieses Lizenzvertrags identisch, vergleichbar oder kompatibel sind. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-sa/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work or its contents in public and alter, transform, or change this work as long as you attribute the work in the manner specified by the author or licensor. New resulting works or contents must be distributed pursuant to this license or an identical or comparable license.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

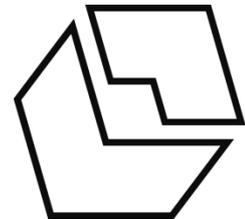
9

forschungsdaten
bildung **informiert**

Sonja Bayer, Johannes Breuer,
Thomas Lösch und Jürgen W. Goebel

Nutzung von Social-Media-Daten in der Bildungsforschung

Version 1.2 // April 2021



Impressum

forschungsdaten bildung informiert // Nr. 9

Herausgeber

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Verbund Forschungsdaten Bildung

Rostocker Straße 6

60323 Frankfurt am Main

verbund@forschungsdaten-bildung.de

Redaktion und Layout

Ramona Gietzen

Die Handreichung wurde in Zusammenarbeit mit der Rechtsanwaltskanzlei Goebel & Scheller erstellt. Der vorliegende Text beinhaltet Hinweise, die auf Erfahrungswerten beruhen und daher rechtlich nicht verbindlich sein können. Er beinhaltet keine juristischen Einzelfallprüfungen.

Wir bedanken uns bei Dr. Sebastian Karcher, Dr. Elisabeth Huber und Dr. Jan-Ocko Heuer für die hilfreichen und konstruktiven Rückmeldungen zu früheren Versionen der Handreichung.

Zitationsvorschlag: Bayer, S., Breuer, J., Lösch, T. und J. W. Goebel, 2021. Nutzung von Social-Media-Daten in der Bildungsforschung. *forschungsdaten bildung informiert* 9, Version 1.

www.forschungsdaten-bildung.de

1	EINFÜHRUNG	4
2	BEGRIFFSBESTIMMUNGEN.....	4
3	ZUGANG ZU SOCIAL-MEDIA-DATEN UND VERKNÜPFUNG MIT ANDEREN DATEN.....	6
3.1	ZUGANG	6
3.2	VERKNÜPFUNG.....	9
4	RECHTLICHE ASPEKTE DER VERARBEITUNG VON SOCIAL-MEDIA-DATEN.....	9
4.1	DATENSCHUTZRECHT	11
4.2	URHEBERRECHT.....	13
4.3	RECHT AM EIGENEN BILD.....	15
4.4	VERTRAGSRECHT	16
4.5	SONSTIGE RECHTSBEREICHE.....	17
5	INFORMIERTE EINWILLIGUNG.....	18
5.1	EINFÜHRUNG.....	18
5.2	INFORMIERTE EINWILLIGUNG FÜR SOCIAL-MEDIA-DATEN	19
6	ARCHIVIEREN UND PUBLIZIEREN VON SOCIAL-MEDIA-DATEN	20
6.1	ARCHIVE, REPOSITORIEN UND FORSCHUNGSDATENZENTREN	21
6.2	RECHTLICHE UND ETHISCHE ASPEKTE	22
6.3	MAßNAHMEN ZUM SCHUTZ PERSONENBEZOGENER UND SENSIBLER DATEN	23
6.4	DOKUMENTATION VON SOCIAL-MEDIA-DATEN	24
7	LITERATURVERZEICHNIS.....	25

1 Einführung

Social Media gewinnen im privaten, schulischen und beruflichen Leben zunehmend an Bedeutung. Insbesondere im Alter zwischen 12 und 18 Jahren liegt in Deutschland die Zahl jener, die täglich Social Media nutzen – je nach Dienst – zwischen 15 % (Facebook) und 64 % (Instagram) (Medienpädagogischer Forschungsverbund Südwest, 2020). Die besonderen Formen sozialer Interaktion und speziell auch die Daten, die diese generieren, eröffnen neue Einblicke und innovative Forschungsfragen für die Wissenschaft (Golla et al., 2018). Dies bietet auch für die Bildungs- und Sozialforschung neue Möglichkeiten für die Analyse sozialen Handelns (bspw. Dunn, 2019; Karakoyun, 2019; Maresova et al., 2020; McManus, 2019). Im Speziellen lassen sich mit Social-Media-Daten etwa individuelle Informations- und Bildungsprozesse oder auch Aspekte des kollektiven Wissensaufbaus erforschen (bspw. Chaka & Govender, 2020; Farr & Riordan, 2015). Social-Media-Daten bieten zudem potenzielle Vorteile für die Forschung im Vergleich zu anderen Erhebungsverfahren. Zunächst einmal sind sie im Vergleich zu anderen Datentypen (etwa aus Befragungs-, Experimental- oder Interviewstudien) schnell zugänglich. So können häufig mit wenig Aufwand große Fallzahlen erreicht werden, die Kosten für die Datenerhebung sind meist gering und Interaktionen (wie auch andere Prozesse) können über längere Zeiträume hinweg untersucht werden (Hagen et al., 2019). Insbesondere im Vergleich zu Selbstauskünften haben Social-Media-Daten spezifische Vorteile: Sie sind unter anderem nicht reaktiv und weniger anfällig für soziale Erwünschtheit. Zudem sind in kurzer Zeit viele Daten verfügbar (siehe z. B. Stier et al., 2020).

Trotz der diversen Vorteile ergeben sich erhebliche Anforderungen bei der Forschung mit Social-Media-Daten, die bspw. bei Befragungsdaten oder administrativen Daten in dieser Form nicht auftreten. Diese Publikation möchte Forschende dabei unterstützen, Anforderungen und Prozesse bei der Erhebung und Verwendung von Social-Media-Daten zu erkennen und zu reflektieren und somit einen produktiven und rechtssicheren Umgang mit diesen Daten zu ermöglichen bzw. zu erleichtern.

Die vorliegende Handreichung präsentiert zuerst einen allgemeinen Einstieg in die Definition von Social Media (Kapitel 2) und thematisiert verschiedene Zugangswege zu Social-Media-Daten sowie die Verknüpfung dieser Daten mit Daten aus anderen Erhebungsformaten (Kapitel 3). Anschließend wird vertiefend auf zentrale Rechtsbereiche eingegangen, die bei der Forschung mit Social-Media-Daten tangiert sein können (Kapitel 4). Die Beschreibung bezieht sich auf entsprechende Gesetze und Rechtsvorschriften, hat jedoch den Anspruch, anhand praktischer Bezüge die rechtlichen Grundlagen und ihre Bedeutung für die Forschungspraxis zu reflektieren. In Kapitel 5 wird der Aspekt des Datenschutzes konkret anhand des Themas der informierten Einwilligung diskutiert, wobei speziell auch auf Besonderheiten bei der Erhebung von Daten in und mit Social Media eingegangen wird. Die Handreichung gibt abschließend Hinweise und Empfehlungen für das Archivieren und Publizieren von Social-Media-Daten mit Hinweisen auf vertiefende Informationsquellen zu diesem Thema (Kapitel 6).

2 Begriffsbestimmungen

Im Forschungskontext sind Daten allgemein eine Grundlage, ein Werkzeug und ein Produkt des Forschungsprozesses, die zu Informationen werden, wenn ihnen eine Bedeutung zugeordnet wird (Bodendorf, 2006; Hagen et al., 2019). Sie sind damit eine Ressource, die erhoben, erfasst, bearbeitet,

gespeichert, archiviert und geteilt werden kann. Im Folgenden werden wir diesbezüglich allgemein von der „Verarbeitung von Daten“ sprechen.

Der Begriff Social Media ist in der Literatur unterschiedlich besetzt (Carr & Hayes, 2015) und reicht von weit gefassten Merkmalszuschreibungen als „bessere Möglichkeiten, Inhalte online zu veröffentlichen und zu bearbeiten sowie zum Austausch mit anderen“ (Schmidt, 2017, S. 11) bis hin zu sehr engen Definitionen, bspw. zugespißt auf die technische Ebene (vgl. Röhle, 2015). Eine übergreifende Definition, die auch den praktischen Forschungsbezug dieser Handreichung adressiert, leiten Carr und Heyes (2015) aus einer umfassenden Sichtung vielfältiger Definitionen von Social Media ab. Sie definieren Social Media als „Internet-based, disentrained, and persistent channels of masspersonal communication facilitating perceptions of interactions among users, deriving value primarily from user-generated content“ (Carr & Hayes, 2015, S. 8). Social Media sind demnach Internet-basierte Kanäle, die eine dauerhafte und an das Tool gebundene soziale Interaktion zwischen einem großen Kreis von Nutzenden ermöglichen. Die Inhalte des Kanals werden überwiegend von den Nutzenden generiert.

Die Ausgestaltung der Social-Media-Tools bzw. -Plattformen ist vielfältig und im steten Wandel. Als Orientierungshilfe können die Tools und Plattformen anhand ihrer Funktionalität kategorisiert werden. Eine gute Heuristik liefern Kaplan und Haenlein (2010) mit ihrer Unterscheidung in sechs Typen:

- Blogs und Mikroblogs (z. B. Twitter, Tumblr)
- Soziale Netzwerke (z. B. Facebook, MySpace)
- Inhaltsgemeinschaften (z. B. YouTube, Reddit, Twitch, TikTok, Instagram, Flickr, Vine, Spotify¹)
- Kollaborative Projekte (z. B. Wikipedia),
- Virtuelle Spiele (z. B. World of Warcraft)
- Virtuelle soziale Welten (z. B. Second Life)

Die Zuordnungen sind nicht immer eindeutig und können sich über die Zeit wandeln. Twitter wird in dieser Klassifikation bspw. als Mikroblog kategorisiert, bietet jedoch zugleich die Möglichkeit der sozialen Vernetzung (Social Media Research Group, 2016). Zudem können Funktionen von Social Media in anderen Diensten integriert sein. Im Kontext der Bildungsforschung sind bspw. digitale Plattformen relevante hybride Dienste, die zu Lehr- und Lernzwecken eingesetzt werden können, wie Lernapps, Massive Open Online Courses (MOOCs) oder Lernmanagement-Systeme (z. B. Moodle oder Ilias). Diese Lernplattformen bieten auch weitere Analysemöglichkeiten von Lern- und Leistungsprozessen über Log-Daten (siehe Ifenthaler, 2020). Im Zentrum dieser Publikation stehen jedoch der soziale Austausch und der gezielt bereitgestellte sowie geteilte Inhalt.

Uneinigkeit herrscht zudem bei der Zuordnung von Instant Messengern (z. B. WhatsApp, Telegram, Snapchat) zu Social-Media-Diensten. Instant Messengers ermöglichen es Nutzenden, sich in Echtzeit über einen oder auch mehrere Kommunikationskanäle miteinander zu verbinden. Ähnlich wie in

¹ Einige der hier genannten Dienste werden häufig auch unter den Begriff der Streaming-Dienste gefasst, bspw. Spotify, Twitch oder YouTube. Diese Dienste bieten zwar auch Möglichkeiten der Interaktion und Vernetzung, jedoch liegt der Fokus der Nutzung hier meist stärker auf der passiven Rezeption.

sozialen Netzwerken können die Nutzenden Kontaktlisten pflegen und andere Personen dort wahlweise hinzufügen oder ihnen die Autorisierung verweigern (Hornung & Müller-Terpitz, 2015). Im Sinne der obigen Definition können Instant Messenger demnach den Social-Media-Diensten zugeordnet werden.

In dieser Handreichung werden wir jedoch nicht näher auf Instant Messenger eingehen; vielmehr beziehen wir uns auf Social-Media-Dienste, die überwiegend den ersten drei Kategorien obiger Heuristik zuordenbar sind: Blogs und Mikroblogs, soziale Netzwerke und Inhaltsgemeinschaften, die auch als hybride Formen in anderen Diensten enthalten sein können. Während viele der in dieser Handreichung genannten Punkte auch auf andere Formen von Social Media (wie z. B. Instant Messenger) übertragbar sind, liegt der Fokus auf Plattformen aus den o. g. Kategorien, da diese bislang in der Forschung am häufigsten untersucht bzw. genutzt wurden und zudem speziell auch für die Bildungs- und Sozialforschung vielfältige relevante Daten liefern können.

3 Zugang zu Social-Media-Daten und Verknüpfung mit anderen Daten

3.1 Zugang

Ähnlich vielfältig wie Social-Media-Daten sind deren Zugangswege. Grundlegend gibt es folgende Zugangsmöglichkeiten (vgl. Breuer et al., 2020):

1. Daten selbst erheben
 - a. manuell (z. B. über Screenshots oder manuelle Kodierung)
 - b. über Application Programming Interfaces (APIs)
 - c. mittels Web Scraping
2. Direkte Kooperationen mit Plattformen,
3. Erwerb von Drittanbietenden (Data Reseller oder Marktforschungsfirmen)
4. Nutzung bereits erhobener und veröffentlichter Daten
5. Kooperation mit Nutzenden

Jeder dieser Zugangswege hat seine eigenen Vor- und Nachteile. Wichtige Dimensionen, auf denen sich die Optionen für den Zugang zu Social-Media-Daten unterscheiden, sind die benötigten Ressourcen (Kosten, Zeit, Fähigkeiten), Art und Umfang der Daten, Kontrolle über den Erhebungsprozess und Unabhängigkeit der Forschenden (Breuer et al., 2020).

Bei der **manuellen Erhebung** sind Forschende unabhängig und haben Kontrolle darüber, welche Daten sie erfassen. Allerdings eignet sich dieser Ansatz aufgrund des mit ihm verbundenen Aufwands nicht für

die Erhebung großer Datenmengen und ist daher eher für qualitative Forschungsansätze (v. a. small-n studies) relevant.²

Insbesondere in der quantitativen Forschung ist der Zugriff über **APIs** eine weit verbreitete Methode. APIs erlauben die Sammlung großer und zugleich strukturierter Datensätze. Zu bedenken ist bei dieser Methode jedoch, dass APIs in aller Regel Beschränkungen im Hinblick auf die Art, den Umfang und die Häufigkeit von Datenabfragen haben. In der Regel haben APIs, ähnlich wie die jeweiligen Plattformen selbst, Nutzungsbedingungen, die regulieren, wie sie genutzt werden können. Diese sogenannten Terms of Service (ToS) sind für Forschende nicht immer klar zu verstehen und können sich überdies ändern. Inwieweit bzw. wie (sehr) die akademische Forschung an die ToS von Plattformen und deren APIs gebunden ist und welche Konsequenzen es haben kann, wenn ToS von Forschenden (teilweise) ignoriert werden, ist eine anhaltende forschungsethische sowie rechtliche Diskussion. Auch wenn dies nicht auf den deutschen Kontext übertragbar ist, ist nach einem Gerichtsurteil aus den USA ein Verstoß gegen die ToS durch Forschende dort zumindest kein strafrechtlich relevantes Vergehen (nach dem Urteil fällt dies nicht unter den sogenannten „Computer Fraud and Abuse Act“; siehe American Civil Liberties Union, 2020). Nach deutschem Recht kann die Nichtbeachtung der ToS (Allgemeine Nutzungsbedingungen = AGB) allerdings zivilrechtliche Konsequenzen haben (Schadensersatz, Unterlassung; eventuell auch weitergehende Rechtsfolgen nach dem Urheberrechtsgesetz (UrhG)).

Unabhängig von den ToS können die Anbietenden allerdings auch die Funktionalitäten der APIs jederzeit ändern oder diese sogar ganz abschalten. Ein prominentes Beispiel ist die Reaktion von Facebook auf den Cambridge-Analytica-Skandal im Jahr 2018, im Zuge dessen die Datenzugangsmöglichkeiten über die APIs von Facebook drastisch reduziert wurden, so dass sie nicht mehr für die wissenschaftliche Forschung genutzt werden können. Einige Forschende sprechen in diesem Kontext von einer „APIcalypse“ (Bruns, 2019) bzw. dem Beginn eines „post-API age“ (Freelon, 2018).

Vor diesem Hintergrund argumentieren einige, dass **Web Scraping** (zur Definition der Methode siehe z. B. Rat für Sozial- und Wirtschaftsdaten [RatSWD], 2019 oder Mancosu & Vegetti, 2020) als Methode zur Sammlung von Social-Media-Daten (wieder) wichtiger wird (Freelon, 2018; Mancosu & Vegetti, 2020). Web Scraping ist sehr flexibel einsetzbar und nicht von den Vorgaben von APIs abhängig. Allerdings erfordert es in aller Regel Programmierkenntnisse und es gibt eine anhaltende Debatte um die rechtlichen Aspekte des Web Scrapings in der wissenschaftlichen Forschung, da dieses gemäß der ToS von Plattformen oftmals nicht erlaubt ist (Golla & von Schönfeld, 2019; RatSWD, 2019).

Für Forschende, die Social-Media-Daten selbst erheben wollen, gibt es eine Vielzahl an Tools, von denen viele frei verfügbar sind (viele davon ermöglichen oder erleichtern den Datenzugriff über APIs). Übersichten über solche Tools bieten bspw. das [Social Media Research Toolkit](#) vom Social Media Lab der

² Bei der manuellen Erhebung können sich zudem, je nach Art der Erfassung, Besonderheiten für die Weiterverarbeitung (und damit auch die Archivierung) der Daten ergeben, wenn bspw. mit Screenshots oder Videoaufzeichnungen gearbeitet wird – z. B. im Hinblick auf die Anonymisierung, Copyright-Fragen oder auch hinsichtlich praktischer Aspekte wie der Dateigröße.

Ryerson University³ oder die [Liste der Data Collection Tools](#) vom Social Media Observatory des Leibniz-Instituts für Medienforschung | Hans-Bredow-Institut (HBI)⁴.

Eine der möglichen Alternativen dazu, Daten selbst zu erheben, ist die **direkte Kooperation mit Plattformen** (etwa im Modell des sogenannten „embedded researchers“; siehe Breuer et al., 2020). Dies ermöglicht Forschenden i. d. R. umfangreicheren Zugriff auf Daten, jedoch können vertragliche Vereinbarungen mit den Plattformbetreibern die Nutzung der Daten und damit die Unabhängigkeit der Forschenden deutlich einschränken.⁵ Ähnliches gilt häufig auch für den **Erwerb von Daten von Drittanbietern**. Zudem wird hier der Datenerhebungsprozess nicht von den Forschenden kontrolliert und ist nicht selten in Teilen intransparent. Auch bei der **Nutzung bereits erhobener und veröffentlichter Social-Media-Daten** haben die Forschenden, die diese (wieder-)verwenden möchten, keinen Einfluss darauf, wie diese erhoben werden bzw. wurden. Es gibt sowohl fortlaufende Erhebungen (z. B. [GESIS Social Media Monitoring](#)⁶ oder [TweetsKB](#)⁷), Sammlungen i. d. R. thematischer Datensätze (z. B. [The DocNow Catalog](#),⁸ oder [TweetSets](#)⁹) sowie publizierte Datensätze zu spezifischen Studien bzw. Projekten (siehe z. B. Kinder-Kurlanda et al., 2017), die (wieder)verwendet werden können.

Eine weitere Option für den Zugang zu Social-Media-Daten ist die **Kooperation mit den Nutzenden** von Social-Media-Diensten. Halavais (2019) beschreibt dies als „ethical distributed research“ (S. 1567), mit dem es möglich ist, die Probleme zu vermeiden, die sich etwa durch die ToS von APIs für Forschende ergeben. Viele Plattformen (z. B. Facebook und Twitter) bieten ihren Nutzenden die Möglichkeit, ihre persönlichen Daten zu exportieren. Diese können sie dann mit Forschenden teilen. Eine Alternative zu einer solchen Datenspende kann z. B. auch die Verwendung von Browser-Plugins sein, die Studienteilnehmende verwenden können, um einen Teil ihrer Social-Media-Daten zu erfassen und den Forschenden zugänglich zu machen (siehe z. B. Haim & Nienierza, 2019). Der große Vorteil solcher Ansätze ist, dass sie das Einholen einer informierten Einwilligung erleichtern (siehe hierzu ausführlicher Abschnitt 5) und zudem transparenter für die Personen sind, deren Daten erhoben werden. Ein wesentlicher Nachteil ist der für solche Lösungen notwendige Aufwand im Hinblick auf die Umsetzung technischer Lösungen für das Teilen bzw. die Erfassung der Daten von Nutzenden.

Eine wesentliche Besonderheit in der Arbeit mit Social-Media-Daten ist, dass neben den Interessen und Rechten der Beforschten sowie Anforderungen und Prinzipien der Scientific Community auch Interessen und Vorgaben kommerzieller Anbieter berücksichtigt werden müssen. Dies gilt grundlegend für alle Zugangsoptionen, wenn auch in geringerem Maße für die Kooperation mit Nutzenden. Welchen Zugangsweg Forschende wählen, sollte im Wesentlichen von ihrer jeweiligen Forschungsfrage

³ Social Media Research Toolkit, <https://socialmediadata.org/social-media-research-toolkit/> (letzter Zugriff: 27.11.2020).

⁴ Social Media Observatory Data Collection Tools, <https://github.com/Leibniz-HBI/Social-Media-Observatory/wiki/Data-Collection-Tools> (letzter Zugriff: 27.11.2020).

⁵ Für die Zusammenarbeit zwischen Forschenden und Plattformen werden aktuell alternative Modelle diskutiert (King & Persily, 2019; Puschmann, 2019). Das prominenteste umgesetzte Modell ist Social Science One (<https://socialscience.one/> (letzter Zugriff: 18.01.2021)), bei dem eine unabhängige Organisation als Mediator zwischen Forschenden aus der Wissenschaft und dem Unternehmen, das die Daten hält (in diesem konkreten Fall Facebook), agiert.

⁶ GESIS Social Media Monitor, <http://mediamonitoring.gesis.org/> (letzter Zugriff: 27.11.2020).

⁷ TweetsKB, <https://data.gesis.org/tweetskb/> (letzter Zugriff: 27.11.2020).

⁸ The DocNow Catalog, <https://catalog.docnow.io/> (letzter Zugriff: 27.11.2020).

⁹ TweetSets, <https://tweetsets.library.gwu.edu/> (letzter Zugriff: 27.11.2020).

abhängen bzw. davon, welche (Art von) Social-Media-Daten sie zu deren Beantwortung benötigen. Zugleich müssen natürlich aber auch die verfügbaren Ressourcen (finanzielle Mittel, Zeit, Personal und Expertise) bei der Entscheidung berücksichtigt werden. Ein weiterer Aspekt, der bei der Entscheidung bedacht werden sollte, ist, dass jeder Zugangsweg zu spezifischen Verzerrungen (Biases) in den Daten führen kann (siehe Sen et al., 2019). Ein Beispiel ist die Selbstselektion bei Studien, in denen direkt mit Nutzenden kooperiert wird, um an Social-Media-Daten zu gelangen.

3.2 Verknüpfung

Social-Media-Daten enthalten häufig nur begrenzte Informationen über wesentliche Merkmale der Personen, die diese generiert haben. Dies betrifft etwa soziodemographische Merkmale aber auch Einstellungen. Zudem erfassen sie in aller Regel nur sehr eingeschränkt Dimensionen des Offline-Verhaltens. Um diesen Limitationen zu begegnen, kann es sinnvoll sein, Social-Media-Daten mit anderen Daten zu verknüpfen. Dies können etwa Daten aus Interviews oder Umfragen sein (siehe Stier et al., 2020).

Menchen-Trevino (2013) trifft für sogenannte digitale Spurdaten, zu denen Social-Media-Daten gehören, die Unterscheidung zwischen sogenannten „horizontal“ und „vertical“ (S.330) digital traces. Bei „horizontal digital traces“ geht es um die Erfassung in der Breite. Ein Beispiel wäre die Sammlung aller Tweets zu einem bestimmten Hashtag über die Twitter API. Demgegenüber gehen „vertical digital traces“ in die Tiefe und erfassen umfangreiche Daten für eine geringere Anzahl an Beobachtungseinheiten (i. d. R. Individuen). Beispiele hierfür wären die Verknüpfung von Befragungs- oder Interviewdaten mit (verschiedenen) Social-Media-Daten für dieselben Befragten oder teilnehmende Beobachtung am Social-Media-Verhalten der Nutzenden.

Neben dieser Verknüpfung auf Individualebene gibt es allerdings auch die Möglichkeit der Verknüpfung von aggregierten Social-Media-Daten mit Befragungs- oder Interviewdaten. Beispielsweise könnten aggregierte Social-Media-Daten für bestimmte Zeiträume oder Regionen (z. B. das Sentiment von Tweets zu einem bestimmten Thema) mit Befragungsdaten kombiniert werden (Stier et al., 2020). Social-Media-Daten können sowohl parallel mit Interview- oder Befragungsdaten erfasst werden (ex-ante Verknüpfung) oder es können Daten aus zuvor unabhängig voneinander erhobenen Datensätzen nachträglich verknüpft werden (ex-post Verknüpfung). Welche Form der Verknüpfung möglich und sinnvoll ist, hängt wiederum von der Art der verfügbaren Daten sowie der jeweiligen Forschungsfrage ab.

4 Rechtliche Aspekte der Verarbeitung von Social-Media-Daten

Leitlinien guter wissenschaftlicher Praxis und Stellungnahmen zum Umgang mit Forschungsdaten geben Forschenden allgemeine Orientierungshilfen für den Forschungsprozess und das Datenmanagement (Deutsche Forschungsgemeinschaft [DFG], 2015; DFG, 2019; Deutsche Gesellschaft für Erziehungswissenschaft [DGfE] et al., 2020; Deutsche Gesellschaft für Soziologie [DGS], 2019). Diese haben auch für die Forschung mit Social-Media-Daten ihre Gültigkeit. Darüber hinaus birgt die Forschung mit Social-Media-Daten jedoch neue Herausforderungen. Dies betrifft insbesondere den Bereich rechtlicher Fragen.

Es ist technisch möglich, dass Forschende verbale Äußerungen, Bilder oder Informationen über Dritte aus Social Media für Forschungszwecke nutzen, auch ohne Wissen und Erlaubnis derjenigen, welche die teils hochsensiblen Informationen teilen, jener, die mit den Daten in Bezug gebracht werden können oder jener, die Rechte an bestimmten Inhalten haben (Williams et al., 2017). Allerdings kann durch die Verarbeitung und Publikation solcher Daten durch Forschende den betroffenen Personen u. U. erheblicher Schaden zugefügt werden. Darüber hinaus kann eine unrechtmäßige Verarbeitung von Daten negative Konsequenzen für die Forschenden haben. In diesem Abschnitt gehen wir daher auf zentrale Rechtsbereiche ein, die bei der Verarbeitung von Social-Media-Daten tangiert sein können. Auch ethische Aspekte gilt es zu berücksichtigen (Williams et al., 2017). Diese werden in der Handreichung ebenfalls an einigen Stellen explizit diskutiert (etwa in den Abschnitten 5 und 6).

Bislang existiert kein einheitlicher Rechtsrahmen für Social Media. Dennoch ist dieser Bereich stark reguliert durch eine Fülle von unterschiedlichen Rechtsvorschriften, die es zu beachten gibt. Je nach Schwerpunkt des Sachverhalts und insbesondere des in Rede stehenden Rechts sind jeweils ein anderes Gesetz bzw. meist mehrere Gesetze parallel heranzuziehen (Roggenkamp & Krieg, 2015). Zentral, jedoch nicht erschöpfend, für die Forschung mit Social-Media-Daten sind folgende Rechtsbereiche:

- Datenschutzrecht (bei der Verwendung personenbezogener Daten)
- Urheberrecht (darunter können Texte, Bilder oder auch Skizzen, Tafelbilder und ganze Profile fallen)
- Recht am eigenen Bild (wenn Abbildungen von Personen Teil der erhobenen Daten sind)
- Vertragsrecht (bspw. Nutzungsbedingungen und AGB von Plattformbetreibenden)

Aus Sicht von Datennutzenden beschränken diese Vorschriften die Möglichkeiten der Verarbeitung von Daten, schützen jedoch zugleich die Betroffenen bzw. Rechteinhabenden. Demgegenüber stehen Forschungsprivilegien (GG Artikel 5 Abs. 3), welche (wissenschaftlich) Forschenden wiederum größere Freiheiten einräumen, insbesondere auch im Umgang mit personenbezogenen Daten und dem Urheberrecht. In diesem Spannungsfeld sind die Möglichkeiten und Grenzen der Verarbeitung von Daten als Abwägungselemente zwischen den verschiedenen Rechtsgrundlagen und den beteiligten Rechtspositionen zu sehen.

Die folgenden Ausführungen beschreiben dieses rechtliche Spannungsfeld mit einem Fokus auf Social-Media-Daten. Zudem wird auf Besonderheiten der Bildungsforschung eingegangen. Bspw. sind häufig Kinder und Jugendliche Gegenstand der Bildungsforschung, deren Daten nach der Datenschutz-Grundverordnung (DSGVO) einen besonderen Schutz genießen.

Es sei jedoch betont, dass nicht immer gesicherte Antworten aus den Gesetzen abgeleitet werden können. Vielmehr ist für die Abwägungen der jeweils spezifische Kontext entscheidend. Bei konkreten Fragen können Datenschutz- und Datenmanagementbeauftragte unterstützen. Diese Positionen sind an

Forschungseinrichtungen sowie zunehmend auch bei Förderlinien und Forschungsdateneinrichtungen angesiedelt¹⁰. Im Zweifelsfall sollte juristischer Rat hinzugezogen werden.

4.1 Datenschutzrecht

Social-Media-Daten weisen generell eine große Dichte personenbezogener Daten auf. Neben direkten Identifikatoren können indirekte Identifikatoren einen Rückschluss auf die betroffene Person erlauben. Mit direkten Identifikatoren werden Informationen oder Merkmale bezeichnet, durch die unmittelbar auf eine bestimmte Person geschlossen werden kann. Typische Beispiele sind der Name, das Bild der betroffenen Person oder die Adresse. Aber auch E-Mail-Adressen, IP-Adressen oder Telefonnummern können einen Rückschluss auf eine einzelne Person ermöglichen. Indirekte Identifikatoren sind Informationen oder Merkmale, mit denen man in Kombination oder durch Anspielung externer Informationen eine Person indirekt identifizieren kann (siehe hierzu auch Rocher et al., 2019). Insbesondere mit zunehmenden Möglichkeiten der Datenverarbeitung und Verfeinerung von Suchalgorithmen sind Daten immer einfacher re-personalisierbar. Bei der Verarbeitung von Social-Media-Daten ist somit der rechtssichere Umgang mit personenbezogenen Daten zentral (Meyermann & Porzelt, 2019; Watteler & Ebel, 2019).

Als Element des allgemeinen Persönlichkeitsrechts unterliegen personenbezogene und personenbeziehbare Daten dem Recht auf informationelle Selbstbestimmung (GG Art. 1 Abs. 1 und 2 Abs. 1; BVerfGE 65, 1). Diese Daten unterliegen damit einem besonderen Schutz und dürfen nur unter bestimmten Voraussetzungen verarbeitet werden. Grundlage für jedwede Verarbeitung personenbezogener Daten ist dabei seit Mai 2018 die DSGVO. Erweiterte Regelungen gelten dabei für besondere Kategorien personenbezogener Daten, die besonders sensibel sind und eines besonderen Schutzes bedürfen, weil ihre Verarbeitung erhebliche Risiken für die betroffenen Personen mit sich bringen kann (bspw. politische Meinungen, religiöse und weltanschauliche Überzeugungen, ethnische Herkunft, sexuelle Überzeugung; Art 9, DSGVO). Daneben gibt es das Bundesdatenschutzgesetz (BDSG) sowie länderspezifische Gesetze (z. B. HDSIG, BayDSG), welche die datenschutzrechtlichen Regelungen der DSGVO konkretisieren und teilweise erweitern. Zentral für Erhebungen an Schulen sind darüber hinaus Vorschriften etwa in Spezialgesetzen wie den Schulgesetzen der jeweiligen Länder (z. B. HSchG, BayEUG).

Bei der Anwendung der Datenschutzgesetze ist stets zunächst die DSGVO maßgeblich. Spezialgesetze wie die Schulgesetze können die DSGVO ergänzen. Die allgemeinen Datenschutzgesetze von Bund und Ländern kommen erst dann zur Anwendung, wenn die Regelungen der DSGVO Spielräume oder sogenannte Öffnungsklauseln aufweisen (Meyermann & Porzelt, 2019).

4.1.1 Grundlagen für die Verwendung personenbezogener Daten

Datenschutzgesetze sind einschlägig, sobald personenbezogene Daten lebender, „natürlicher Personen“ verarbeitet werden. Daten von Institutionen fallen bspw. nicht unter diese Rechtsgrundlage (Art. 4 Nr. 1 DSGVO). Die Verarbeitung personenbezogener Daten ist nur unter bestimmten Bedingungen

¹⁰ Bspw. das Beratungsangebot des VerbundFDB, <https://www.forschungsdaten-bildung.de/beratung?la=de> (letzter Zugriff: 19.03.2021).

rechtmäßig. Für die wissenschaftliche Forschung primäre Bedingung ist die Einwilligung der betroffenen Person in Form einer informierten Einwilligung (Art. 4 Nr. 2 DSGVO; zur Gestaltung der informierten Einwilligung siehe Verbund Forschungsdaten Bildung, 2019). Auch wenn Personen unter Pseudonymen in Social Media präsent sind, ist es insbesondere bei sozialen Netzwerken möglich, aus Daten wie Bildern, Nachrichten, Vorlieben und Kontakten auf die konkrete Person zu schließen (Hornung, 2015). Auch für die Verarbeitung von Daten, unter denen Personen nicht mit ihrem Namen direkt identifizierbar sind, bedarf es daher einer informierten Einwilligung (siehe auch Abschnitt 5 zur informierten Einwilligung in diesem Dokument). Soweit ein Personenbezug aus den erhobenen Daten ausgeschlossen werden kann, ist für die Verarbeitung der Daten keine informierte Einwilligung erforderlich. Wenn ein solcher Personenbezug nur mit großem Aufwand hergestellt werden kann, den kein Mensch vernünftigerweise betreiben würde, gelten die Daten als anonym (siehe dazu die Legaldefinition etwa in § 2 Abs. 4 HDSIG).

Personenbezogene Daten Dritter sind ebenfalls bei Forschungsvorhaben in und mit Social Media zu beachten. Wenn Personen Daten Dritter in ihren Profilen bereitstellen, so handelt es sich zum einen um personenbezogene Daten des Profilinhabenden, etwa die Information über eine persönliche Beziehung zu dem Dritten. Zudem können die Daten Dritter ebenfalls einen Personenbezug aufweisen, insbesondere wenn die dritte Person ebenfalls Nutzende der entsprechenden Plattform ist (Doppelbezug; Hornung, 2015). Konkrete Beispiele hierfür sind Retweets bei Twitter oder personenbeziehbare Informationen über Dritte in schriftlichen Beiträgen. Für die Verarbeitung personenbezogener Daten Dritter muss ebenfalls eine informierte Einwilligung der betroffenen Person eingeholt werden.

4.1.2 Verwendung personenbezogener Daten ohne Einwilligung

In bestimmten Fällen ist jedoch auch eine Datenverarbeitung ohne informierte Einwilligung rechtmäßig. Drei Bedingungen werden im Folgenden näher beschrieben, wobei für die Anwendung in der Praxis die jeweiligen Spezifika des konkreten Verwendungsfalls ausschlaggebend sind. Zudem sollte im konkreten Fall, auch durch die zuständige Ethikkommission, geprüft werden, ob die Erhebung ohne informierte Einwilligung ethisch vertretbar ist. Fachgesellschaften fordern in jeweiligen Ethik-Richtlinien primär dazu auf, die Personen, deren Daten verarbeitet werden, um deren Einwilligung zu bitten (siehe bspw. [Ethikkodex der DGfE¹¹](#)).

1. **Frei zugängliche Profildaten**, d.h. Externe können ohne eigenen Account auf der entsprechenden Plattform die personenbezogenen oder personenbeziehbaren Daten einsehen (öffentliches Profil): In diesen Fällen kann eine stillschweigende Erlaubnis bezüglich der personenbezogenen Daten angenommen werden, deren Verarbeitung zu wissenschaftlichen Zwecken ohne informierte Einwilligung möglich ist. Ausgenommen sind (geschützte) Daten Dritter. Für die Verarbeitung der personenbezogenen Social-Media-Daten muss also maßgeblich differenziert werden, welche Privatsphäreinstellungen Nutzende vorgenommen haben. Sofern die Daten nur für einen begrenzten Kreis von ausgewählten Personen verfügbar sind, ist eine ausschließlich persönliche Zwecksetzung anzunehmen und die Erhebung für Forschungszwecke nur mit zusätzlichen rechtlichen Mitteln wie einer informierten Einwilligung möglich (Hornung, 2015; siehe auch Abschnitt 6.2).

¹¹ Ethik-Kodex der Deutschen Gesellschaft für Erziehungswissenschaft, https://www.dgfe.de/fileadmin/OrdnerRedakteure/Satzung_etc/Ethikkodex_2016.pdf (letzter Zugriff: 10.12.2020)

2. **Forschungsprivileg** (Art. 89 Abs. 1 DSGVO): In bestimmten Fällen kann das Einholen einer informierten Einwilligung dem Forschungsvorhaben entgegenstehen. Für solche Fälle sieht die DSGVO Ausnahmen für wissenschaftlich Forschende vor, die eine Verarbeitung personenbezogener Daten ohne informierte Einwilligung der betroffenen Person ermöglicht oder es erlaubt, die informierte Einwilligung nachträglich einzuholen. Im Einzelfall ist eine Abwägung zwischen den Interessen des oder der Betroffenen und denen des oder der Forschenden vorzunehmen (siehe Meyermann & Porzelt 2019, S. 23).

4.1.3 Verwendung personenbezogener Daten ohne Einwilligung

Bei der Verarbeitung personenbezogener Social-Media-Daten gibt es je nach Datenquelle und Forschungsdesign weitere rechtliche Aspekte, die berücksichtigt werden müssen. Dies betrifft nicht nur die Abwägung zwischen dem Schutz personenbezogener Daten und dem Forschungsprivileg, sondern auch die Wahl der Plattform sowie der Tools für die Erhebung von Daten (siehe hierzu Abschnitt 3.1). Manche Anbietende verarbeiten personenbezogene Daten ohne Einwilligung der Kommunikationspartner*innen. Ein Beispiel ist das Auslesen des Telefonverzeichnisses bei WhatsApp. Telefonnummern werden ohne Einwilligung der jeweiligen Inhabenden an WhatsApp übermittelt. Daten können zudem außerhalb der EU verarbeitet werden und unterliegen somit nicht dem Schutz der DSGVO. Darüber hinaus können Daten an Dritte weitergegeben werden. Insbesondere für die Generierung neuer Daten (bspw. Befragungen) sind die Bestimmungen der verwendeten Plattformen und Tools einzubeziehen.

Das „Recht auf Vergessenwerden“ (Art. 17 DSGVO) ist insbesondere mitzudenken, wenn für den Forschungsprozess oder zum Zwecke der Weiterverarbeitung Profile gespeichert werden. Wird ein Profil mit Kenntnis des Forschenden gelöscht, kann es nötig werden, auch die gesicherten Daten beim Forschenden oder der archivierenden Einrichtung zu löschen. Dies ist eine Abwägungssache zwischen dem Recht des oder der Betroffenen auf Vergessenwerden, was eine Löschung der Daten bedeutet, und dem Forschungsprivileg. Überwiegen die Interessen der Forschenden in Abwägung zu den Interessen des Betroffenen, kann eine Löschung nicht erforderlich sein. Ein Beispiel hierfür könnte die Nutzung und Archivierung gelöschter Tweets von Politiker*innen¹² oder anderen Personen des öffentlichen Interesses sein. Speziell etwa bei kontroversen Äußerungen von Politiker*innen auf Social Media kann die Dokumentationspflicht bzw. das öffentliche Interesse an solchen Daten das Recht auf Vergessenwerden überwiegen (Personen der Zeitgeschichte).

4.2 Urheberrecht

Die Urheber*innen von Werken der Literatur, Wissenschaft und Kunst genießen für ihre Werke Schutz (§ 1 UrhG). Welche Werke geschützt sind, ist in § 2 UrhG festgelegt. Übertragen auf die Forschung mit Social-Media-Daten können bspw. Videos, Fotos, Texte, Sprachnachrichten, Grafiken oder auch ganze Profile oder Chatverläufe urheberrechtlich geschützt sein. Ausschlaggebend für den urheberrechtlichen Schutz ist, dass das Werk eine persönliche geistige Schöpfung darstellt. Dies bedeutet:

¹² Die Plattform Politwoops (<https://www.politwoops.co.uk/>) (letzter Zugriff: 19.03.2021) sammelt systematisch gelöschte Tweets von Politiker*innen sowie Daten zu diesen.

- 1) Es liegt ein Schöpfungsakt einer handelnden Person vor (im Gegensatz dazu ist ein Stein in der Natur keine Schöpfung).
- 2) Das Werk weist einen geistigen Gehalt auf (im Gegensatz zu Routineerzeugnissen).
- 3) Die Individualität des*der Urhebers*in kommt zum Ausdruck.
- 4) Eine bestimmte Gestaltungshöhe wird erreicht.
- 5) Es hat eine Formgebung stattgefunden.

Auch die (Weiter-)Bearbeitung und Umgestaltung von vorhandenen Werken ist urheberrechtlich geschützt (Esser, 2015).

Speziell für geschriebene Texte ist auf ein Urteil des Europäischen Gerichtshofs zu verweisen, der in einem urheberrechtlichen Streitfall entschieden hat, dass selbst Auszüge im Umfang von elf Wörtern urheberrechtlich geschützt sein können. Dabei hat es sich bei dem Streitfall um die kommerzielle Vervielfältigung von Zeitungsartikeln mittels eines automatisierten Verfahrens und deren anschließenden elektronischen Verarbeitung gehandelt (Urheberrechte – Informationsgesellschaft – Richtlinie 2001/29/EG – Art. 2 und 5 – Werke der Literatur und Kunst – Begriff der Vervielfältigung – ‚Teilweise‘ Vervielfältigung – Vervielfältigung kurzer Auszüge aus Werken der Literatur – Zeitungsartikel – Vorübergehende und flüchtige Vervielfältigungen – Technisches Verfahren, das im Einscannen von Artikeln und deren anschließender Umwandlung in eine Textdatei, einer elektronischen Verarbeitung der Vervielfältigung, der Speicherung eines Teils der Vervielfältigung und deren Ausdruck besteht, C-5/08, Europäischer Gerichtshof 16. Juli 2009). Jedoch fällt auch die Verarbeitung einzelner Tweets und ähnlichem unter die Verwendung von urheberrechtlich geschützten Werken (Hagen et. al, 2019). Ausschlaggebend ist jedoch nicht nur die Länge, sondern auch die Originalität der Texte. Für Beiträge in dem sozialen Netzwerk Twitter entschied das Landesgericht Bielefeld bspw., dass ein bloßer Sprachwitz nicht für den urheberrechtlichen Schutz ausreicht (LG Bielefeld, 3. 1. 2017 – 4 O 144/16, MMR 2017, 641 ff).

4.2.1 Verwendung von urheberrechtlich geschützten Werken

Werden Werke, die dem Urheberrecht unterliegen, zu Forschungszwecken verarbeitet, ist dies meistens als Vervielfältigung, Verbreitung und/oder öffentliche Zugänglichmachung anzusehen, für die es entsprechende Verwertungsrechte des Rechteinhabers bedarf. Der Rechteinhaber ist im Idealfall die Person, welche das entsprechende Werk in Social Media verfügbar macht. Es können aber auch Dritte Rechteinhabende sein, wenn die betroffene Person Werke von Dritten übernommen hat. Verwertungsrechte können bspw. in Form von Lizenzen allgemein vergeben oder es können individuelle vertragliche Vereinbarungen getroffen werden. Bspw. können Forschende in der informierten Einwilligung von den Personen, deren Daten gesammelt werden, auch das Recht erbitten, urheberrechtlich geschützte Daten verarbeiten zu dürfen (siehe auch Abschnitt 5).

4.2.2 Verwendung urheberrechtlich geschützter Daten ohne Verwertungsrechte

In wenigen Fällen ist eine Verarbeitung von urheberrechtlich geschützten Daten möglich. Drei Ausnahmen werden im Folgenden näher beschrieben.

1. Forschungsprivileg: Das Urheberrecht räumt der nicht-kommerziellen Forschung bestimmte Privilegien ein, indem zu diesem Zwecke Werke bis zu 15 % vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden dürfen. Voraussetzung ist, dass diese für einen bestimmten,

abgegrenzten Kreis von Personen für deren eigene wissenschaftliche Forschung genutzt werden sollen (§ 60 Abs. 1 UrhG). Für die eigene wissenschaftliche Forschung dürfen darüber hinaus bis zu 75 % eines Werkes vervielfältigt werden. Zu beachten ist, dass eine Verarbeitung, Archivierung oder Weitergabe über die individuelle Nutzung hinaus nicht oder nur in geringerem Umfang (bis 15 %) zulässig ist. Die Spezifikation von bestimmten Prozentanteilen eines Werkes ist für abgeschlossene Texte und einzelne Bilder relativ eindeutig festzulegen. Bei fortgeschriebenen Werken, wie etwa Posts oder gesamte Profile, sind die Kriterien für die Festlegung bestimmter Prozente hingegen bislang ungeklärt.

2. Web Scraping für die wissenschaftliche Forschung: Laut einem Gutachten des RatSWD (2019) ist unter bestimmten Voraussetzungen Data Mining, wie es beim Web Scraping zur Anwendung kommt, auch für potenziell urheberrechtlich geschützte Daten für die wissenschaftliche Forschung zustimmungsfrei.

Schutzdauer: Das Urheberrecht erlischt nach einer bestimmten Zeitdauer. In verschiedenen Staaten gelten hierbei unterschiedliche Regelungen. Ausschlaggebend ist die Regelung des Landes, in dem das Werk genutzt wird, nicht die Regelung des Landes, in dem das Werk geschaffen wurde. Für in Deutschland genutzte Werke endet der urheberrechtliche Schutz nach 70 Jahren, für Lichtbilder nach 50 Jahren und für Datenbanken nach 15 Jahren (§ 64–95 UrhG). Ähnliche Schutzfristen gelten auch innerhalb der Europäischen Union.

4.3 Recht am eigenen Bild

In Social Media werden häufig Bilder und Videos geteilt. Für die Forschung mit diesen Daten ergeben sich Einschränkungen über das Urheberrecht hinaus, sobald Personen abgebildet sind. Das Kunsturhebergesetz (KUG) räumt der Person Einfluss- und Entscheidungsmöglichkeiten über das Anfertigen und die Verwendung von Bildern und Videos ein, auf denen sie abgebildet ist, das sogenannte Recht am eigenen Bild (§§ 22 ff. KUG; Müller-Terpitz, 2015). Bilder dürfen somit nur mit Einwilligung des Abgebildeten verbreitet oder öffentlich zur Schau gestellt werden. Die Einwilligung gilt im Zweifel als erteilt, wenn die*der Abgebildete dafür, dass sie*er sich abbilden ließ, eine Entlohnung erhalten hat (§ 22 KUG).

Für die Forschung bedeutet dies, dass Fotos, Videos o. ä. (im Folgenden zusammenfassend Bilder genannt), die dem KUG unterliegen, für interne Forschungszwecke gesammelt und verarbeitet werden dürfen. Die Daten dürfen jedoch nicht weitergegeben oder gezeigt werden. Dies bedeutet auch, dass Bilder, die in Social Media frei verfügbar sind, nicht im Rahmen von Erhebungen gezeigt werden dürfen. Es dürften bspw. keine Lehr-Lernvideos aus YouTube im Rahmen einer Interventionsstudie als Lehrmaterial eingesetzt werden.

Für das KUG gelten wiederum Ausnahmeregelungen, unter denen Bilder von Personen ohne Einwilligung verbreitet und zur Schau gestellt werden dürfen. Hierbei gelten keine gesonderten Ausnahmeregelungen für die Forschung, stattdessen sind allgemeingültige Ausnahmeregelungen in § 23 des KUG festgelegt:

1. Bilder aus dem Bereiche der Zeitgeschichte
2. Bilder, auf denen die Personen nur als Beiwerk¹³ neben einer Landschaft oder sonstigen Örtlichkeiten erscheinen
3. Bilder von Versammlungen, Aufzügen und ähnlichen Vorgängen, an denen die dargestellten Personen teilgenommen haben
4. Bilder, die nicht auf Bestellung angefertigt sind, sofern die Verbreitung oder Zurschaustellung einem höheren Interesse der Kunst dient

Die Ausnahmen finden nur Anwendung, wenn durch eine Verbreitung und Zurschaustellung des Bildes kein berechtigtes Interesse der*des Abgebildeten oder, falls diese*r verstorben ist, ihrer*seiner Angehörigen verletzt wird. Zudem bieten die Ausnahmeregelungen wiederum Interpretationsspielraum und müssen im Einzelfall geprüft werden. Rechtshandbücher geben dazu unterstützende Hinweise (z. B. Götting & Becker, 2008 oder Ricker et al., 2012). In jedem Fall kommt es aber auf die besonderen Umstände des Einzelfalls an.

4.4 Vertragsrecht

Die meisten Social-Media-Dienste knüpfen den Zugang zu ihren Angeboten an AGBs, Nutzungs- und Teilnahmebedingungen (Bräutigam & von Sonnleithner, 2015). Die Nutzung von Social-Media-Diensten für private, kommerzielle oder Forschungszwecke ist somit bestimmten vertragsrechtlichen Regelungen unterworfen (§§ 145ff BGB). Diese Terms of Service (ToS) können für die gesamte Plattform gelten und um weitere ToS für einzelne (Unter-)Dienste ergänzt sein (bspw. Application Programming Interfaces (APIs), siehe dazu auch Abschnitt 3.1).

Die ToS setzen einen starken Rahmen, welche Forschungsaktivitäten möglich und erlaubt sind. Sie können bspw. regulieren, welche Daten in welchem Umfang von Externen genutzt werden können. Dabei sind die ToS nicht immer klar verständlich und können sich über die Zeit ändern, u. a. da der stete technische und wirtschaftliche Wandel zu fortlaufenden Modifikationen des Leistungsumfangs von Plattformen (sowie auch ihrer APIs) führen kann (Bräutigam & von Sonnleithner, 2015). Für Forschungsvorhaben empfiehlt sich daher, die bei Datenerhebung geltenden ToS mit Datumsvermerk zu archivieren, als Nachweis für die zur Erhebung geltenden Bedingungen (zum Thema der Dokumentation siehe auch Abschnitt 6.4). Forschende sind wie alle anderen Nutzenden des Social-Media-Dienstes an das Vertragsrecht und somit die ToS gebunden. Nicht nur deshalb empfiehlt es sich, die Inhalte der ToS gründlich nachzuvollziehen, eventuell auch mithilfe externer juristischer Beratung.

In bestimmten Fällen können einzelne Regelungen der ToS von Anbietenden außerhalb der EU nicht mit den in Deutschland und der EU geltenden verbraucher*innenschützenden Gesetzen in Einklang stehen. Bspw. musste Facebook seine Nutzungsbedingungen anpassen, da einzelne Regelungen und

¹³ Als Beiwerk kann eine Person angesehen werden, wenn die Landschaft oder sonstige Örtlichkeit den Gehalt des Bildes prägt. Die Personenabbildung muss derart untergeordnet sein, dass sie auch entfallen könnte, ohne dass Gegenstand und Charakter des Bildes sich dadurch verändern würde; sie darf insoweit nicht selbst Thema des Bildes sein (Ernst, 2010, S. 180; Ricker et al., 2012, Kap. 43 Rn. 28; Schertz, 2008, Rn. 63).

Voreinstellungen vom Landesgericht Berlin als gesetzeswidrig eingestuft wurden (Az 160 341/[9] 15). Grundsätzlich gilt, dass Angebote in Deutschland von Dienstleistenden aus anderen Nationen den in Deutschland geltenden Rechten unterworfen sind. Aus Sicht von Nutzenden und Forschenden gilt daher, die geltenden gesetzlichen Regelungen einzuhalten. Dies betrifft auch das Mindestalter der Nutzenden. Social-Media-Anbietende wie Snapchat, Twitter oder Facebook geben (mit Stand der Publikation) ein Mindestalter von 13 Jahren für ihre Nutzung an. Jüngere Personen dürfen weder einen Account eröffnen noch den Service nutzen. Nach geltendem Recht liegt die Altersgrenze zur Eröffnung eines Accounts bei Social-Media-Diensten in Deutschland und der EU hingegen bei 16 Jahren. Personen unter 16 Jahren ist es nur mit Einwilligung der Eltern (der gesetzlichen Vertretenden) erlaubt, diese Dienste zu nutzen, da personenbezogene Daten verarbeitet werden, die u. a. bei der Eröffnung eines Kontos erforderlich sind, sowie die Daten durch die Anbietenden häufig für Werbezwecke und zur Profilbildung genutzt werden (siehe Erwägungsgrund 38 und Art. 8 DSGVO). Häufig findet dies in den ToS der Plattformen keine Erwähnung und wird selten von den Betreibern geprüft. Als Besonderheit der Genehmigung kann im Falle von Social Media auch die (ausdrückliche) Duldung des Verhaltens, also der Nutzung eines Social-Media-Dienstes, gesehen werden. Für die Forschung ergeben sich daraus rechtliche Unsicherheiten, insbesondere wenn Personen unter 16 Jahren zur Zielgruppe gehören.

4.5 Sonstige Rechtsbereiche

Im Einzelfall können bei der Übernahme von Social-Media-Daten für Forschungszwecke auch andere Rechtsbereiche tangiert sein. Beispielhaft werden untenstehend weitere Rechtsbereiche aufgezählt, die im Bedarfsfall zu berücksichtigen sind. Die Aufzählung beansprucht keine Vollständigkeit, sondern ist als allgemeiner Hinweis zu verstehen.

- Markenrecht: Bspw., wenn auf einem Bild oder in einem Video eine gesetzlich geschützte Marke (etwa das Logo eines Software-Anbietenden) erkennbar ist. Allerdings wird insoweit i. d. R. kein markenmäßiger Gebrauch der Marke vorliegen, so dass insofern das Markenrecht dann doch nicht eingreift.¹⁴
- Strafrecht: Bspw., wenn übernommene Inhalte strafbar sind, wie z. B. Beleidigungen, Verleumdungen oder Pornographie.
- Firmenrecht: Bspw., wenn der Name einer Firma übernommen wird.
- Allgemeines Persönlichkeitsrecht einer (natürlichen) Person.

Grundrechte, bspw. das Persönlichkeitsrecht und das Recht auf freie Meinungsäußerung aus Art. 2 Abs. 1 GG.

¹⁴ Zum „markenmäßigen Gebrauch“ einer Marke siehe bspw. <https://www.ra-plutte.de/markenmaessige-benutzung-alles-wichtige-im-ueberblick/> (letzter Zugriff: 06.04.2021).

5 Informierte Einwilligung

5.1 Einführung

Wie bereits zuvor erläutert, ist die Verarbeitung personenbezogener Daten nur unter bestimmten Voraussetzungen rechtmäßig (siehe Kapitel 4.1). In der Forschung von zentraler Bedeutung ist die explizite und informierte Zustimmung der Personen, deren Daten verarbeitet werden, dazu, dass die eigenen personenbezogenen Daten verarbeitet werden: die sogenannte informierte Einwilligung nach Art. 6 Abs. 1 lit. a) DSGVO.

Die informierte Einwilligung ist eine „freiwillig für den bestimmten Fall, in informierter Weise und unmissverständlich abgegebene Willensbekundung“ (Art. 4 Abs. 11 DSGVO). Die Einwilligung ist „informiert“, da sie über alle Modalitäten der Datenerhebung informiert (Art. 12 DSGVO). Die Inhalte der informierten Einwilligung sind in Art. 13 der DSGVO geregelt. Diese umfasst im Kontext der Bildungsforschung bspw. die beteiligten Forschenden und deren Affiliation, den Titel des Forschungsprojekts, Zeitpunkt und Ablauf der Datenerhebung oder die Prozesse der Datenverarbeitung. Unverzichtbar ist überdies die Berücksichtigung der Rechte der Betroffenen wie die Freiwilligkeit der Datenverarbeitung, das Recht auf Widerruf und die Folgenlosigkeit bei Verweigerung oder Widerruf. Zudem sind die Zwecke, für die die personenbezogenen Daten verarbeitet werden sollen, anzugeben.¹⁵

Sollen die erhobenen Daten im Sinne der guten wissenschaftlichen Praxis für mindestens zehn Jahre archiviert und ggf. veröffentlicht werden, muss dies bereits in der informierten Einwilligung angegeben werden. Eine informierte Einwilligung kann zudem als eine Art Vereinbarung verstanden werden, durch die Rechte an urheberrechtlich geschütztem Material geregelt werden können. In einer informierten Einwilligung kann bspw. erklärt werden, dass Fotos, die Betroffene auf ihren Profilen veröffentlicht haben, für die definierten Forschungszwecke verarbeitet und gemeinsam mit allen weiteren Daten in einem Forschungsdatenzentrum veröffentlicht werden dürfen (siehe auch Kapitel 4.2.1).

Eine [Checkliste zur Erstellung rechtskonformer informierter Einwilligungen](#)¹⁶ sowie [allgemeine Formulierungsbeispiele](#)¹⁷ und [Formulierungsbeispiele in leichter Sprache](#)¹⁸ gibt es vom Verbund

¹⁵ Rechtlich noch nicht gesichert ist ein sogenannter Broad Consent, welcher eine möglichst breite Verarbeitung der personenbezogenen Daten für über das konkrete Projekt hinausgehende Zwecke ermöglicht, bspw. auch ohne Angabe von Datenempfangenden und Verwendungsarten.

¹⁶ Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen des VerbundFDB, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_1.pdf (letzter Zugriff: 10.12.2020).

¹⁷ Formulierungsbeispiele für informierte Einwilligungen des VerbundFDB, https://www.forschungsdaten-bildung.de/files/fdbinfo_4.pdf (letzter Zugriff: 10.12.2020).

¹⁸ Formulierungsbeispiele für informierte Einwilligungen in leichter Sprache des VerbundFDB, https://www.forschungsdaten-bildung.de/files/fdbinfo_5.pdf (letzter Zugriff: 10.12.2020).

Forschungsdaten Bildung (VerbundFDB), vom [RatSWD](#)¹⁹ sowie bei anderen FDZen²⁰ oder lokalen Beauftragten für [Forschungsdatenmanagement](#)^{21, 22}.

Neben den rechtlichen Rahmenbedingungen für die informierte Einwilligung (siehe vorheriger Abschnitt) sollte diese auch forschungsethischen Standards entsprechen. Diese können durch eine Ethikkommission oder auch Richtlinien wissenschaftlicher Fachgesellschaften definiert sein. Für die Forschung mit Social-Media-Daten allgemein bieten in dieser Hinsicht etwa die Ethical Guidelines der Association of Internet Researchers (AOIR; Franzke et al., 2020) hilfreiche Orientierungspunkte.

5.2 Informierte Einwilligung für Social-Media-Daten

Für die Arbeit mit Social-Media-Daten gibt es einige spezielle Aspekte, die für die informierte Einwilligung relevant sind. Ganz wesentlich ist zunächst die Herausforderung, informierte Einwilligungen für Social-Media-Daten einzuholen. Eine Möglichkeit besteht darin, die informierte Einwilligung für die Erhebung von Social-Media-Daten in anderen Befragungen oder Interviews einzuholen und dort die Teilnehmenden zu fragen, ob bestimmte Teile ihrer Social-Media-Daten zu Forschungszwecken gesammelt und genutzt werden dürfen (siehe hierzu auch Abschnitt 3.2). Ein Beispiel für ein Studiendesign, in dem Befragungs- und Social-Media-Daten auf diese Weise verknüpft wurden (inklusive einer Vorlage für die Formulierung der informierten Einwilligung), liefern Sloan et al. (2020). In dieser Studie wurden Befragungs- und Twitter-Daten verknüpft, aber die Informationen aus der informierten Einwilligung können auch für andere Datentypen und Studiendesigns angepasst werden. Allgemein gilt: Werden Daten verknüpft, sollte explizit angegeben werden, welche Daten dies betrifft und zu welchem Zweck dies geschieht. Eine praktische Herausforderung hierbei ist, den Teilnehmenden die relevanten Informationen zu geben, ohne sie mit technischen Details zu überfordern oder zu verwirren. Die Lösung, die Sloan et al. (2020) hierfür gefunden haben, ist die Bereitstellung zusätzlicher Informationen über Help Screens (zusätzliche, optional einblendbare Erklärungstexte), welche bei Onlineumfragen technisch relativ einfach implementierbar sind. Praktische Hinweise zur Gestaltung und Einholung einer informierten Einwilligung für Social-Media-Daten bietet neben dem Artikel von Sloan et al. (2020) und den Ethical Guidelines der AOIR (Franzke et al., 2020) auch der „Report on legal and ethical framework and strategies related to access, use, re-use, dissemination and preservation of social media data“ (Hagen et al., 2019) aus dem Projekt SERISS (Synergies for Europe’s Research Infrastructures in the Social Sciences).

Wenn die Forschenden nicht in direktem Kontakt mit den Social-Media-Nutzenden stehen, deren Daten sie verarbeiten (wollen), ist das Einholen einer informierten Einwilligung deutlich schwieriger. In Studien, in denen Social-Media-Daten für eine kleine Anzahl von Personen gesammelt werden (z. B. manuell, über Application Programming Interfaces (APIs) oder Web Scraping), sollte es prinzipiell

¹⁹ Die informierte Einwilligung als Voraussetzung für die (Nach-)Nutzung von Forschungsdaten des RatSWD, https://www.forschungsdaten-bildung.de/files/RatSWD_WP_264.pdf (letzter Zugriff: 10.12.2020).

²⁰ Erläuterungen zur Verwendung der von Qualiservice bereitgestellten Vorlagen für die informierte Einwilligung, <https://doi.org/10.26092/elib/192> (letzter Zugriff: 21.01.2021).

²¹ FDM-Kontakte, <https://www.forschungsdaten.org/index.php/FDM-Kontakte> (letzter Zugriff: 09.12.2020).

²² Spezielle Formulierungsbeispiele und Guidelines für die Formulierung der informierten Einwilligung für Social-Media-Studien werden im nachfolgenden Abschnitt diskutiert.

möglich sein, diese zu kontaktieren, um eine informierte Einwilligung einzuholen, bevor die Daten ausgewertet oder veröffentlicht werden (siehe hierzu auch Kapitel 4.1.2). Gegebenenfalls kann die Kontaktaufnahme auch erst nach der Datenerhebung erfolgen, sofern davor nicht klar ist, welche Personen in den gesammelten Daten vertreten sein werden.

Bei sehr umfangreichen Datensammlungen (im Sinne von „horizontal digital traces“ nach Menchen-Trevino, 2013) ist es in vielen Fällen praktisch nicht möglich, eine informierte Einwilligung einzuholen (zu den rechtlichen Möglichkeiten siehe Kapitel 4.1). In solchen Fällen ist es dann umso wichtiger, geeignete Maßnahmen zu ergreifen, welche die personenbezogenen Daten sowie die betroffenen Personen angemessen schützen. Dies betrifft insbesondere die Speicherung, Archivierung und Publikation der Daten.

6 Archivieren und Publizieren von Social-Media-Daten

Forschungsdaten im Sinne der guten wissenschaftlichen Praxis für mindestens zehn Jahre zu archivieren und zu publizieren, wird von der wissenschaftlichen Community und von Forschungsfördernden zunehmend gefordert. So wird bspw. im Rahmenprogramm empirische Bildungsforschung vom BMBF standardmäßig gefordert, Forschungsdaten nach Projektende zu publizieren. Die [Leitlinien zu guter wissenschaftlicher Praxis der DFG](#)²³ sowie die Richtlinien vieler wissenschaftlicher Fachgesellschaften (bspw. der [Deutschen Gesellschaft für Erziehungswissenschaften \(DGfE\)](#), der [Gesellschaft für Empirische Bildungsforschung \(GEBF\)](#) und der [Gesellschaft für Fachdidaktik \(GfD\)](#)²⁴, des [Fachkollegiums 111 „Sozialwissenschaften“](#)²⁵ und der [Deutsche Gesellschaft für Psychologie \(DGPs\)](#)²⁶, siehe auch die regelmäßig aktualisierte Übersicht der [DFG](#)²⁷) empfehlen ebenso die Veröffentlichung von Forschungsdaten.

Bei der Veröffentlichung ist es wichtig, dass Daten nachvollziehbar dokumentiert werden, so dass sie den sogenannten [FAIR-Kriterien](#)²⁸ entsprechen, gemäß derer Forschungsdaten auffindbar (findable), zugänglich (accessible), interoperabel (interoperable) und wiederverwendbar (reusable) sein sollen.

Während es für andere Forschungsdaten wie quantitative Surveys, Leistungsmessungen, Interviewtranskriptionen oder audiovisuelle Aufzeichnungen Verfahren zur Archivierung und

²³ Leitlinien zur Sicherung guter wissenschaftlicher Praxis der DFG, https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf (letzter Zugriff: 27.11.2020).

²⁴ Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung, https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/stellungnahme_forschungsdatenmanagement.pdf (letzter Zugriff: 06.12.2020).

²⁵ Stellungnahme des Fachkollegiums 111 „Sozialwissenschaften“ zum Forschungsdatenmanagement in der Soziologie, der Politikwissenschaft und der Kommunikationswissenschaft, https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/fachkollegium111_forschungsdatenmanagement.pdf (letzter Zugriff: 06.12.2020).

²⁶ Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen, <https://psyarxiv.com/hcxtm/> (letzter Zugriff: 27.11.2020).

²⁷ Umgang mit Forschungsdaten, https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/ (letzter Zugriff: 06.12.2020).

²⁸ FAIR Principles, <https://www.go-fair.org/fair-principles/> (letzter Zugriff: 09.12.2020).

Veröffentlichung gibt (bspw. beim [VerbundFDB](#)²⁹), ist dies für Social-Media-Daten nicht der Fall. Für diese Medien ergeben sich zudem hinsichtlich der Publikation und Archivierung einige spezifische An- und Herausforderungen. Diese betreffen praktische, rechtliche und ethische Aspekte sowie die Dokumentation der Daten.

Im Vergleich zu Befragungs- oder Interviewdaten sind Social-Media-Daten i. d. R. wesentlich umfangreicher und Datensätze dementsprechend häufig um ein Vielfaches größer (sowohl was die Fallzahlen als auch die Anzahl der Variablen anbelangt). Auch die Geschwindigkeit, mit der die Daten produziert werden, ist im Vergleich zu anderen Datentypen, die in der Bildungsforschung gemeinhin genutzt werden, deutlich höher. Social-Media-Daten werden kontinuierlich generiert und können zudem von Nutzenden verändert oder gelöscht werden. Diese drei Eigenschaften sind wesentliche Gründe dafür, weshalb die Archivierung von Social-Media-Daten für Datenrepositorien eine besondere Herausforderung darstellt. Gerade für komplexe, sehr umfangreiche oder langfristige bzw. kontinuierliche Sammlungen von Social-Media-Daten müssen daher u. U. neue Lösungen für die Archivierung sowie den Datenzugang bei den Repositorien entwickelt werden (siehe hierzu z. B. Hemphill, 2019; Van Atteveldt et al., 2020).

6.1 Archive, Repositorien und Forschungsdatenzentren

Zum aktuellen Zeitpunkt gibt es noch keine Infrastruktur, die ein auf Daten aus Social Media spezialisiertes Angebot zur Veröffentlichung von Daten anbietet. Es befinden sich jedoch Angebote im Aufbau. Beim Inter-university Consortium for Political and Social Research (ICPSR) in den USA wird aktuell ein dezidiertes Social Media Archive (SOMAR) entwickelt (siehe Hemphill, 2019; Hemphill et al., 2018) und andere Archive arbeiten ebenfalls an der Entwicklung entsprechender Lösungen und Angebote. Zudem gibt es einige Archive und Repositorien, bei denen bereits ausgewählte Social-Media-Daten aufgenommen und veröffentlicht wurden.

Als allgemeine Angebote können bspw. Zenodo und das Open Science Framework (OSF) genannt werden. Bei diesen liegt der Fokus auf der Selbstarchivierung. Das OSF ist zudem völlig offen, was Daten- und Dateiformate anbelangt, erfordert keine systematische Dokumentation und nutzt keine strukturierten Metadaten, was die Auffindbarkeit (Findability) und potenziell auch die Nachnutzbarkeit (Reusability) reduziert. Die Nutzung solcher Plattformen für die eigenständige Archivierung ist niedrigschwellig, jedoch liegt dieser Prozess (inkl. der Dokumentation der Daten) dort gänzlich in den Händen der Forschenden.

Es gibt jedoch auch Repositorien, die eine systematische Dokumentation erfordern, strukturierte Metadaten nutzen und den Forschenden Unterstützung bei der Archivierung ihrer Daten anbieten. Einige davon sind auch für die Archivierung von Social-Media-Daten geeignet. Für Daten mit psychologischem Schwerpunkt kommt das Angebot PsychArchives des Leibniz-Instituts für Psychologie (ZPID) in Frage. Auch hier können Daten verschiedener Struktur und somit auch Daten aus Social Media veröffentlicht werden. Differenzierte Möglichkeiten der Zugriffsbeschränkung befinden sich derzeit jedoch noch im Aufbau. Für Daten aus Social Media, die im Kontext qualitativer Ansätze gewonnen wurden, kommt eine

²⁹ VerbundFDB – Daten teilen, <https://www.forschungsdaten-bildung.de/daten-teilen> (letzter Zugriff: 09.12.2020).

Veröffentlichung über das FDZ Qualiservice in Frage, das themenungebunden alle Arten von qualitativen sozialwissenschaftlichen Forschungsmaterialien archiviert und unter von den Datengebenden festzulegenden Bedingungen für wissenschaftliche Nachnutzungen bereitstellt.

Bei GESIS – Leibniz-Institut für Sozialwissenschaften besteht bereits Expertise im Umgang mit Social-Media-Daten. Die Struktur der Archivierungsangebote bei GESIS ist ausreichend flexibel, um auch Daten aus Social Media aufnehmen zu können. Auch Möglichkeiten zur Zugriffsbeschränkung sind gegeben. Dementsprechend sind bei GESIS bereits einige Social-Media-Datensätze archiviert. So gibt es dort bspw. Twitter-Daten zu den deutschen Bundestagswahlen 2013 (Kaczmarek & Mayr, 2015) und 2017 (Stier et al., 2018), Facebook-Daten zur österreichischen Nationalratswahl 2013 (Müller et al., 2017) sowie eine umfangreiche Sammlung georeferenzierter Twitter-Daten aus den USA (siehe Kinder-Kurlanda et al., 2017). Trotz dieser Erfahrungen und der diversen archivierten Social-Media-Datensätze sind sowohl bei GESIS als auch bei Qualiservice zum aktuellen Zeitpunkt noch keine spezialisierten Angebote etabliert, so dass praktische, rechtliche und ethische Fragen im Hinblick auf die Veröffentlichung der Daten sowie Fragen bezüglich der Dokumentation noch auf Einzelfallbasis geklärt werden müssen.

6.2 Rechtliche und ethische Aspekte

Wie in Kapitel 4 beschrieben, muss insbesondere bezogen auf den Datenschutz und das Urheberrecht geklärt sein, dass Forschende die Befugnis haben, Daten zu veröffentlichen. Eine Besonderheit bei Social-Media-Daten ist, dass die Veröffentlichung der Daten durch die Nutzungsbedingungen (Terms of Service (ToS)) der Plattformen oder ihrer Application Programming Interfaces (APIs), sofern diese zur Datengewinnung genutzt wurden, eingeschränkt werden kann. So dürfen bspw. bei Twitter die Tweets in aller Regel nicht als Texte, sondern nur die sogenannten Tweet-IDs archiviert werden. Diese Tweet-IDs können dann genutzt werden, um die Tweets über die Twitter API erneut zu sammeln; ein Prozess, der Rehydration (Rehydrierung) genannt wird. Da gelöschte Tweets oder Tweets von gelöschten Accounts so nicht wieder erfasst werden können, entspricht diese Regelung zwar dem Recht auf Vergessenwerden, reduziert aber die Reproduzierbarkeit von Forschung sowie die Nachnutzungsmöglichkeiten der Daten.

Unabhängig von Vorgaben der Plattformanbietenden – hinsichtlich der Nutzung der Daten – ist ein zentrales Thema im Hinblick auf die Veröffentlichung von Social-Media-Daten der Schutz personenbezogener Daten. Dieser wirft sowohl rechtliche (siehe Abschnitt 4) als auch ethische Fragen auf. Ein Beispiel für eine forschungsethische Frage in diesem Kontext ist bspw. diejenige nach der Öffentlichkeit von Informationen bzw. Daten. Selbst wenn z. B. ein Profil oder bestimmte Inhalte technisch bzw. rechtlich gesehen öffentlich sind, muss sich dies nicht mit den Sichtweisen und Erwartungen der Nutzenden decken. Diese haben i. d. R. eine „imagined audience“ (Marwick & Boyd, 2011), zu der Forschende, die ihre Daten nutzen, für die meisten eher nicht gehören. Zwar geben die Nutzungsbedingungen vieler Plattformen an, dass die Daten durch Dritte genutzt werden können, jedoch lesen die meisten Nutzenden diese nicht oder vergessen deren Inhalte bei der konkreten Nutzung. Dies bedeutet, dass davon ausgegangen werden kann, dass den meisten Nutzenden nicht bewusst ist, dass ihre Daten (auch) für die Forschung genutzt werden können. Daher ist es umso wichtiger, Maßnahmen zu ergreifen, um die Daten so gut wie möglich zu schützen.

6.3 Maßnahmen zum Schutz personenbezogener und sensibler Daten

Grundsätzlich gibt es verschiedene Möglichkeiten, personenbezogene und (potenziell) sensible Social-Media-Daten zu schützen. Wichtig dabei ist, zu bedenken, dass viele der Methoden zur Anonymisierung, die für andere Datenarten entwickelt wurden, bei Social-Media-Daten aufgrund ihrer besonderen Eigenschaften nur begrenzt funktionieren. Generell können die Daten so reduziert oder bearbeitet werden, dass eine Identifikation von Personen nicht direkt bzw. nur unter großem Mehraufwand (z. B. durch die Verknüpfung mit weiteren Daten) möglich ist. Generell bestehen hier Abwägungen zwischen dem Erhalt von Informationen und dem Schutz personenbezogener Daten. Die Palette der Optionen zur Datenreduktion reicht von offensichtlichen Maßnahmen wie dem Ersetzen von Benutzernamen durch generisch IDs³⁰ oder dem Löschen bestimmter Variablen aus den zu veröffentlichenden Datensätzen über die Ersetzung von Rohdaten (wie z. B. Texte) durch Annotationen (z. B. Anzahl an Wörtern, identifizierte Themen etc.) bis hin zur Aggregation der Daten, so dass diese nicht mehr auf Individualniveau vorliegen. Diese Maßnahmen führen i. d. R. zu einer Reduktion der Informationen und können dadurch die Nutzbarkeit der Daten beeinträchtigen.

Ein weiterer Mechanismus, der eingesetzt werden kann, um die Daten zu schützen, ist die Beschränkung des Zugriffs. Dabei wird der Zugriff auf die Daten nur unter bestimmten Restriktionen gewährt. Bspw. können Personen nur dann auf die Daten zugreifen, wenn ihre Identität geklärt ist und sie sich vertraglich zur Einhaltung des Datenschutzes erklären. Viele Datenarchive (wie z. B. das von GESIS) bieten die Möglichkeit, verschiedene Zugangsklassen zu vergeben. Zudem können besonders sensible Daten über speziell gesicherte Angebote zugänglich gemacht werden. Dies können bspw. sogenannte „Safe Rooms“ vor Ort in Datenarchiven (wie z. B. das Secure Data Center bei GESIS oder das Safe Center beim Qualiservice) sein. Viele Archive bieten zudem auch Lösungen für den gesicherten Datenzugriff aus der Ferne an (einen sogenannten „secure remote access“) bzw. entwickeln diese – speziell auch für neue Datentypen wie Social-Media-Daten.

Diese beiden Ansätze (Veränderung/Reduktion der Daten und Zugangsbeschränkung) können selbstverständlich auch kombiniert werden. So kann es bspw. eine mögliche Variante sein, einen reduzierten oder aggregierten Datensatz frei zur Verfügung zu stellen und den vollen Datensatz nur für bestimmte Nutzengruppen in speziell gesicherten (physischen oder virtuellen) Umgebungen bereitzustellen. Angesichts der praktischen, rechtlichen und ethischen Herausforderungen beim Zugänglichmachen von Social-Media-Daten werden zurzeit unterschiedliche neue Lösungsideen für den Datenzugang diskutiert und erprobt (siehe dazu z. B. Van Atteveldt et al., 2020). Speziell zu forschungsethischen Fragen im Hinblick auf die Veröffentlichung von Social-Media-Daten gibt es mittlerweile einige Publikationen, die Forschenden Hinweise und konkrete Empfehlungen bieten (z. B. Bishop & Gray, 2017; Mannheimer & Hull, 2018; Thomson, 2016; Williams et al., 2017).

³⁰ Diese Maßnahme allein genügt jedoch nur selten zum Schutz der Daten, speziell wenn es sich um Textdaten handelt, da viele Plattformen eine Volltextsuche ermöglichen, über die sich User Names identifizieren lassen.

6.4 Dokumentation von Social-Media-Daten

Damit die Daten auffindbar (findable) und nachnutzbar (resusable) sind, ist eine ausreichende Dokumentation nötig. Während für andere Datentypen bereits Metadatenstandards existieren (für Befragungsdaten werden bspw. die Metadaten schemata der DDI – Data Documentation Initiative genutzt), gibt es diese für Social-Media-Daten bislang nicht. Zwar lassen sich viele der Metadatenelemente aus etablierten Standards für andere Datentypen (z. B. DDI für Umfragedaten) auch auf Social-Media-Daten anwenden (generelle Hinweise zur Dokumentation von Forschungsdaten finden sich bspw. beim [VerbundFDB](#)³¹ oder bei [CESSDA](#)³²), jedoch gibt es einige Informationen, die für die Dokumentation von Social-Media-Daten relevant sind, welche in diesen Standards nicht auftauchen. Hierzu gehören bspw. detaillierte Angaben zu den spezifischen Methoden der Datensammlung (z. B. APIs, Web Scraping). Hier ist es im Sinne der Reproduzierbarkeit und Replizierbarkeit der Forschung auch relevant, Versionsnummern (von APIs oder Software, die zur Datenerhebung genutzt wurden) sowie den Code für die Datensammlung (z. B. über Web Scraping oder APIs) oder die exakten Suchanfragen (bspw. bei der Verwendung von Programmen mit grafischer Benutzeroberfläche) zu dokumentieren. Weitere Aspekte betreffen die genaue Beschreibung der konkreten Datentypen.

Die Vielfalt und rasante Entwicklung von Social-Media-Plattformen macht die Entwicklung von Standards, die für alle Arten von Social-Media-Daten funktionieren, besonders herausfordernd. Dennoch wird aktuell in verschiedenen Kontexten – wie z. B. dem Consortium of European Social Science Data Archives (CESSDA) – an der Entwicklung von Dokumentations- und Metadatenstandards für Social-Media-Daten gearbeitet. Noch offen ist dabei, inwiefern die Dokumentation der Daten einer Nachnutzbarkeit über forschungsmethodische Ausrichtungen hinweg genügen soll oder kann. So ist unklar, ob eine Dokumentation im Sinne einer quantitativen Erhebung eines sehr großen Social-Media-Datensatzes einer Nachnutzung mittels qualitativer Verfahren genügt (oder umgekehrt).

³¹ Metadaten set des VerbundFDB, https://www.forschungsdaten-bildung.de/files/fdbinfo_8_Metadaten set_v1.0.pdf (letzter Zugriff: 09.12.2020).

³² CESSDA Data Management Expert Guide – Documentation and metadata, <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata> (letzter Zugriff: 09.12.2020).

7 Literaturverzeichnis

- American Civil Liberties Union. (2020, March 28). *Federal court rules 'big data' discrimination studies do not violate Federal anti-hacking law* [Press release]. <https://www.aclu.org/press-releases/federal-court-rules-big-data-discrimination-studies-do-not-violate-federal-anti>
- Bishop, L., & Gray, D. (2017). Ethical Challenges of Publishing and Sharing Social Media Research Data. In K. Woodfield (Eds.), *Advances in Research Ethics and Integrity: Vol. 2. The Ethics of Online Research* (pp. 159-187). Emerald Publishing Limited. <https://doi.org/10.1108/S2398-601820180000002007>
- Bodendorf, F. (2006). *Daten- und Wissensmanagement* (2. Aufl.). Springer VS. <https://doi.org/10.1007/3-540-28682-9>
- Bräutigam, P., & von Sonnleithner, B. (2015). Vertragliche Aspekte der Social Media. In G. Hornung & R. Müller-Terpitz (Hrsg.), *Rechtshandbuch Social Media* (S. 35-77). Springer VS. https://doi.org/10.1007/978-3-642-38192-8_3
- Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*, 22(11), 2058-2080. <https://doi.org/10.1177/1461444820924622>
- 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>
- Carr, C. T., & Hayes, R. A. (2015). Social Media: Defining, Developing, and Divining. *Atlantic Journal of Communication*, 23(1), 46-65. <https://doi.org/10.1080/15456870.2015.972282>
- Chaka, J. G., & Govender, I. (2020). Implementation of mobile learning using a social network platform: Facebook. *Problems of Education in the 21st Century*, 78(1), 24-47. <https://doi.org/10.33225/pec/20.78.24>
- Deutsche Forschungsgemeinschaft. (2015). *Leitlinien zum Umgang mit Forschungsdaten*. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf
- Deutsche Forschungsgemeinschaft. (2019). *Guidelines for Safeguarding Good Research Practice. Code of Conduct*. <https://doi.org/10.5281/ZENODO.3923602>
- Deutsche Gesellschaft für Erziehungswissenschaft, Gesellschaft für Empirische Bildungsforschung & Gesellschaft für Fachdidaktik. (2020). *Empfehlungen zur Archivierung, Bereitstellung und Nachnutzung von Forschungsdaten im Kontext erziehungs- und bildungswissenschaftlicher sowie fachdidaktischer Forschung*. https://www.dgfe.de/fileadmin/OrdnerRedakteure/Stellungnahmen/2020.03_Forschungsdatenmanagement.pdf
- Deutsche Gesellschaft für Soziologie. (2019). *Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie*. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/dgs_stellungnahme_forschungsdaten.pdf
- Dunn, M. B. (2019). Using Social Network Analysis in the Classroom: An Experiential Activity and Tool to Enhance a Sense of Community. *Management Teaching Review*, 4(3), 204-218. <https://doi.org/10.1177/2379298118796600>
- Ernst, S. (2010). Google StreetView: Urheber- und persönlichkeitsrechtliche Fragen zum Straßenpanorama. *Computer und Recht*, 3, 178-184.
- Esser, R. (2015). Strafrechtliche Aspekte der Social Media. In G. Hornung & R. Müller-Terpitz (Hrsg.), *Rechtshandbuch Social Media* (S. 203-321). Springer VS. https://doi.org/10.1007/978-3-642-38192-8_7
- Farr, F., & Riordan, E. (2015). Tracing the reflective practices of student teachers in online modes. *ReCALL*, 27(1), 104-123. <https://doi.org/10.1017/S0958344014000299>
- Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>

- Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>
- Golla, S. J., Hofmann, H., & Bäcker, M. (2018). Connecting the Dots: Sozialwissenschaftliche Forschung in Sozialen Online-Medien im Lichte von DS-GVO und BDSG-neu. *Datenschutz und Datensicherheit*, 2.
- Golla, S. J., & von Schönfeld, M. (2019). Kratzen und Schürfen im Datenmilieu – Web Scraping in sozialen Netzwerken zu wissenschaftlichen Forschungszwecken. *Kommunikation & Recht*, 22(1), 15–21.
- Hagen, S., Bishop, L., Koščík, M., Vavra, M., Štebe, J., Ryan, L., Payne, E., Løvlie, A., Rød, L.-M., Straume, Ø., & Høgteveit Myhren, M. (2019). *Report on legal and ethical framework and strategies related to access, use, re-use, dissemination and preservation of social media data* (Report No. 654221). Synergies for Europe’s Research Infrastructures in the Social Sciences. https://seriss.eu/wp-content/uploads/2019/11/D6.3-Report-on-legal-and-ethical-framework-and-strategies..._FINAL.pdf
- Haim, M., & Nienierza, A. (2019). Computational observation: Challenges and opportunities of automated observation within algorithmically curated media environments using a browser plug-in. *Computational Communication Research*, 1(1), 79–102. <https://doi.org/10.5117/CCR2019.1.004.HAIM>
- Halavais, A. (2019). Overcoming terms of service: A proposal for ethical distributed research. *Information, Communication & Society*, 22(11), 1567–1581. <https://doi.org/10.1080/1369118X.2019.1627386>
- Hemphill, L. (2019). *Updates on ICPSR’s Social Media Archive (SOMAR)*. Inter-university Consortium for Political and Social Research. <https://doi.org/10.5281/ZENODO.3612677>
- Hemphill, L., Leonard, S. H., & Hedstrom, M. (2018). *Developing a Social Media Archive at ICPSR. Proceedings of Web Archiving and Digital Libraries (WADL’18)*. <http://hdl.handle.net/2027.42/143185>
- Hornung, G. (2015). Datenschutzrechtliche Aspekte der Social Media. In G. Hornung & R. Müller-Terpitz (Hrsg.), *Rechtshandbuch Social Media* (S. 79–130). Springer VS. https://doi.org/10.1007/978-3-642-38192-8_4
- Hornung, G., & Müller-Terpitz, R. (Hrsg.). (2015). *Rechtshandbuch Social Media*. Springer VS. <https://doi.org/10.1007/978-3-642-38192-8>
- Ifenthaler, D. (2020). Learning Analytics im Hochschulkontext – Potenziale aus Sicht von Stakeholdern, Datenschutz und Handlungsempfehlungen. In: R. A. Fürst (Hrsg.), *Digitale Bildung und Künstliche Intelligenz in Deutschland* (S. 519–535). Springer VS. https://doi.org/10.1007/978-3-658-30525-3_22
- Kaczmirek, L., & Mayr, P. (2015). *Deutsche Bundestagswahl 2013: Nutzung von Twitter durch Kandidaten* (GESIS Datenarchiv ZA5973; Version 1.0.0) [Data set]. GESIS. <https://doi.org/10.4232/1.12319>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Karakoyun, F. (2019). Preservice Teachers’ Personality Traits and Social Network Use Purposes. *Contemporary Educational Technology*, 10(4), 399–415. <https://doi.org/10.30935/cet.634190>
- Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717736336>
- King, G., & Persily, N. (2019). A New Model for Industry–Academic Partnerships. *PS: Political Science & Politics*, 53(4), 703–709. <https://doi.org/10.1017/S1049096519001021>
- Mancosu, M., & Vegetti, F. (2020). What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data. *Social Media + Society*, 6(3). <https://doi.org/10.1177/2056305120940703>
- Mannheimer, S., & Hull, E. A. (2018). Sharing selves: Developing an ethical framework for curating social media data. *International Journal of Digital Curation*, 12(2), 196–209. <https://doi.org/10.2218/ijdc.v12i2.518>

- Maresova, P., Hruska, J., & Kuca, K. (2020). Social Media University Branding. *Education Sciences*, 10(3), 74. <https://doi.org/10.3390/educsci10030074>
- Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. <https://doi.org/10.1177/1461444810365313>
- McManus, K. (2019). Relationships between social networks and language development during study abroad. *Language, Culture and Curriculum*, 32(3), 270–284. <https://doi.org/10.1080/07908318.2019.1661683>
- Medienpädagogischer Forschungsverbund Südwest (Hrsg.). (2020). *JIM-Studie 2019: Jugend, Information, Medien*. https://www.mpfs.de/fileadmin/files/Studien/JIM/2019/JIM_2019.pdf
- Menchen-Trevino, E. (2013). Collecting vertical trace data: Big possibilities and big challenges for multi-method research. *Policy & Internet*, 5(3), 328–339. <https://doi.org/10.1002/1944-2866.poi336>
- Meyermann, A., & Porzelt, M. (2019). *Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung: eine Handreichung* (forschungsdaten bildung informiert, 6). https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-6_v2.pdf
- Müller, W. C., Bodlos, A., Dolezal, M., Eder, N., Ennser-Jedenastik, L., Kaltenegger, M., Meyer, T. M., Praprotnik, K., & Winkler, A. K. (2017). *AUTNES Content Analysis of Party Facebook Pages 2013* (GESIS Datenarchiv ZA6882; Version 1.0.0) [Data set]. GESIS. <https://doi.org/10.4232/1.12699>
- Müller-Terpitz, R. (2015). Persönlichkeitsrechtliche Aspekte der Social Media. In G. Hornung & R. Müller-Terpitz (Hrsg.), *Rechtshandbuch Social Media* (S. 163–201). Springer VS. https://doi.org/10.1007/978-3-642-38192-8_6
- 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>
- Rat für Sozial- Und Wirtschaftsdaten. (2019). *Big Data in den Sozial-, Verhaltens- und Wirtschaftswissenschaften: Datenzugang und Forschungsdatenmanagement. Mit Gutachten „Web Scraping in der unabhängigen wissenschaftlichen Forschung“*. <https://doi.org/10.17620/02671.39>
- Ricker, R., Weberling, J., Löffler, M. (2012). *Handbuch des Presserechts* (6. Aufl.). C. H. Beck.
- Rocher, L., Hendrickx, J. M., de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(3069). <https://doi.org/10.1038/s41467-019-10933-3>
- Roggenkamp, J. D., & Krieg, H. (2015). *Rechtliche Aspekte sozialer Medien*. Springer- VS. https://doi.org/10.1007/978-3-658-03895-3_19-1
- Röhle, T. (2015). Die Software sozialer Medien. In J.-H. Schmidt & M. Taddicken (Hrsg.), *Handbuch Soziale Medien* (S. 1–22). Springer VS. https://doi.org/10.1007/978-3-658-03895-3_18-1
- Schertz, C. (2008). § 12. Das Recht am eigenen Bild. In H.-P. Götting & B. v. Becker (Hrsg.), *Handbuch des Persönlichkeitsrechts* (S. 208–250). C. H. Beck.
- Schmidt, J.H. (2017). *Social Media* (2. Aufl.). Springer VS.
- Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2019). *A Total Error Framework for Digital Traces of Humans*. ArXiv. <http://arxiv.org/abs/1907.08228>
- Sloan, L., Jessop, C., Al Baghal, T., & Williams, M. (2020). Linking Survey and Twitter Data: Informed Consent, Disclosure, Security, and Archiving. *Journal of Empirical Research on Human Research Ethics*, 15(1–2), 63–76. <https://doi.org/10.1177/1556264619853447>
- Social Media Research Group. (2016). *Using social media for social research: An introduction*. https://dera.ioe.ac.uk/26600/1/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf

- Stier, S., Bleier, A., Bonart, M., Mörsheim, F., Bohlouli, M., Nizhegorodov, M., Posch, L., Maier, J., Rothmund, T., & Staab, S. (2018). *Monitoring sozialer Medien im Bundestagswahlkampf 2017* (GESIS Datenarchiv ZA6926; Version 1.0.0) [Data set]. GESIS. <https://doi.org/10.4232/1.12992>
- Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, 38(5), 503-516. <https://doi.org/10.1177/0894439319843669>
- Thomson, S. D. (2016). *Preserving social media*. Digital Preservation Coalition. <https://doi.org/10.7207/twr16-01>
- Van Atteveldt, W., Althaus, S., & Wessler, H. (2020). The Trouble with Sharing Your Privates: Pursuing Ethical Open Science and Collaborative Research across National Jurisdictions Using Sensitive Data. *Political Communication*. <https://doi.org/10.1080/10584609.2020.1744780>
- Verbund Forschungsdaten Bildung. (2019). *Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen* (fdbinfo, 1). https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_1.pdf
- Watteler, O., Ebel, T. (2019). Datenschutz im Forschungsdatenmanagement. In: Jensen, U., Netscher, S., & Weller, K. (Hrsg.): *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten* (S. 57-80). Barbara Budrich.
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. *Sociology*, 51(6), 1149-1168. <https://doi.org/10.1177/0038038517708140>