

Digitalisierungsforschung: Ein Einblick in die Bandbreite der Forschung zu und mit digitalen Daten

Kohne, Julian; Schmitt, Josephine B.; Breuer, Johannes

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Kohne, J., Schmitt, J. B., & Breuer, J. (2023). Digitalisierungsforschung: Ein Einblick in die Bandbreite der Forschung zu und mit digitalen Daten. *easy_social_sciences*, 69, 3-10. <https://doi.org/10.15464/easy.2023.07>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>



Digitalisierungsforschung

Ein Einblick in die Bandbreite der Forschung zu und mit digitalen Daten

Julian Kohne, Josephine B. Schmitt & Johannes Breuer

„Daten sind das neue Öl“ – dieser plakative Slogan der Werbeindustrie verdeutlicht den Boom in der Auswertung und Nutzung digitaler Daten in den letzten beiden Jahrzehnten. Kommerzielle Anbieter nutzen diese Daten, um immer mehr und immer detailliertere Informationen über ihre Nutzenden und Kund*innen zu erheben und ihre Produkte so personalisieren und weiterentwickeln zu können. Aber digitale Daten können nicht nur zur Gewinnmaximierung genutzt werden. Auch die (sozial-)wissenschaftliche Forschung nutzt immer häufiger solche Daten und erforscht Verhalten im digitalen Kontext. Digitale Forschungsdaten bringen dabei, genau wie das „alte Öl“, enorme Potenziale, aber auch große Herausforderungen mit sich. In dieser Ausgabe von *easy_social_sciences* möchten wir daher einen Einblick in die vielfältigen disziplinären, methodischen und inhaltlichen Ansätze zur Forschung mit und über digitale Daten geben, um einen Eindruck zu vermitteln, wie sie genutzt werden können, um neue Forschungspotenziale zu erschließen.

“Data is the new oil” – this illustrative quote from the advertising industry underlines the boom in the analysis and use of digital data in recent decades. Commercial providers use this data to collect increasingly detailed information about their users and customers in order to develop and personalize their products and services. But digital data cannot only be used to maximize profits. (Social) scientific research is also making increasing use of such data and investigating behavior in a digital context. However, digital research data, just like the proverbial “oil”, comes not only with enormous potentials but also great challenges. In this issue of *easy_social_sciences*, we thus provide an overview of the diverse disciplinary, methodological, and substantive approaches to research with and about digital data. In doing so, we aim to underline how these kinds of data can be used to exploit new research opportunities.

Keywords: Digitale Forschungsdaten, Digitalisierung, digitale Transformation, Interdisziplinarität

„Daten sind das neue Öl“. Mit diesem Satz beschrieb der Data Scientist Clive Humby schon 2006 in einem Vortrag auf der Association of National Advertisers Conference die Potenziale von digitalen Daten für die Werbeindustrie (Humby & Palmer, 2006). Auch wenn die Metapher oftmals kritisiert wurde (vgl. Sonderegger, 2021), hat sich die darin enthaltene Prognose im Wesentlichen bewahrheitet.

Die Entwicklungen der letzten beiden Jahrzehnte können gewissermaßen in Anlehnung an den „Oil Boom“ des 20. Jahrhunderts auch als „Data Boom“ bezeichnet werden. Social-Media-Plattformen wie Facebook, reddit oder YouTube entwickelten Geschäftsmodelle, bei denen die Nutzung der Plattformen sich ausschließlich über zielgerichtete Werbung finanziert, während für Endnutzende keine

Kosten anfallen. Dieses Prinzip funktioniert, da die Plattformen Informationen über ihre Nutzenden sammeln können. Dies umfasst neben soziodemographischen Daten wie Alter, Geschlecht, Familienstand, Wohnort oder das Bildungsniveau auch Eigenschaften wie Hobbies, Interessen, politische Einstellungen oder Werte (z.B. Meta, 2022). Diese Informationen können bei der Nutzung der Plattformen entweder von Nutzenden selbst angegeben werden oder aber mithilfe von mathematischen Modellen aus dem Nutzungsverhalten abgeleitet werden. Laut Googles Chefökonom Hal Varian (2014) sind dabei vier Punkte entscheidend: Datenextraktion und Analyse, Personalisierung und Anpassbarkeit, stetiges Experimentieren und neue Vertragsformen durch bessere Überprüfbarkeit. Damit kann Werbung zielgerichteter und individueller personalisiert werden als zuvor. Dieses Geschäftsmodell erweist sich als äußerst profitabel, wenn es auf Millionen oder gar Milliarden Nutzende angewendet wird. Aus diesem Grund wurde hierfür u.a. die Bezeichnung „Überwachungskapitalismus“ (Englisch: surveillance capitalism) geprägt (Zuboff, 2015).

Aber nicht nur Social-Media-Plattformen haben das enorme Potenzial erkannt. Mittlerweile nutzen fast alle Anbieter von digitalen Services – von Betriebssystemen auf Computern und Smartphones (z.B. Microsoft 2023; Samsung, 2023), über Shopping-Plattformen wie Amazon oder eBay (Amazon, 2023; eBay, 2023), Entertainment-Anbietern wie Spotify (Spotify, 2023) oder Netflix (Netflix, 2022), bis zu speziellen Geräten wie Fitnesstrackern (z.B. Fitbit, 2023) – digitale Daten, um ihre Nutzenden zu quantifizieren und ihre Angebote damit zu optimieren und zu personalisieren. Dies muss nicht unbedingt immer im Interesse der Nutzenden selbst liegen: Das bekannteste Beispiel ist hier wohl das der personalisierten Werbung.

Die digitalen Daten, die Nutzende generieren, sind allerdings nicht nur interessant für die Unternehmen, die diese zu kommerziellen Zwecken erfassen. Auch die sozial-

wissenschaftliche Forschung ist inzwischen „auf Öl gestoßen“. Sie hat die Potenziale für die Forschung zu und mit der Vielzahl an digitalen Daten erkannt und hat begonnen diese zu nutzen (Keuschnigg et al., 2018; Wagner et al., 2021).

Digitale Forschungsdaten: Mehr als nur Big Data

Wenn es um neue Datentypen und die Idee von Daten als dem „neuen Öl“ geht, ist oft von Big Data die Rede. Obwohl der Begriff oft nicht genau definiert ist, können viele der neuen digitalen Forschungsdatentypen zu dieser Kategorie gezählt werden. Der Begriff Big Data umfasst aber nicht notwendigerweise *alle* digitalen Daten, die für die sozialwissenschaftliche Forschung von Interesse sind.

Während Big Data in der Regel durch fünf Merkmale definiert sind, die sich auf das Volumen, die Vielfalt, die Geschwindigkeit, die Genauigkeit und den Nutzen der Daten (im Englischen: volume, variety, velocity, veracity, and value) beziehen, gibt es auch digitale Daten, die für die Forschung wertvoll sind, aber nicht alle diese Merkmale aufweisen (s.u.). Zudem gibt es alternative Begriffe, die v.a. aus Sicht der Sozialwissenschaften, präziser sind als Big Data. So wird am GESIS – Leibniz-Institut für Sozialwissenschaften beispielsweise der Begriff „Digitale Verhaltensdaten“ genutzt, der für alle Spuren von menschlichem Verhalten steht, die durch die Nutzung digitaler Technologien entstehen oder nutzbar gemacht werden können (Wilke et al., 2021). Ein weiterer gängiger Begriff ist „digitale Spurdaten“, welche als „records of activity [...] undertaken through an online information system [...] that can be collected from a multitude of technical systems, such as websites, social media platforms, smartphone apps, or sensors“ definiert werden können (Stier et al., 2020a; Howison et al., 2011).

Damit sind einerseits auch kleinere, homogenere Datensätze gemeint, die für die

Sozialwissenschaften interessant sind, andererseits werden Big Data ausgeschlossen, aus denen wenige oder keine Informationen über Menschen gewonnen werden können. So kann beispielsweise ein einzelner Twitter-Verlauf (jetzt X) einer politischen Persönlichkeit oder eines Parteiaccounts als digitaler Forschungsgegenstand relevant sein – selbst, wenn nicht alle Merkmale von Big Data auf ihn zutreffen. Andererseits gibt es viele Datensätze, auf welche die Definition von Big Data zutrifft, die aber für sozialwissenschaftliche Fragestellungen typischerweise nicht relevant sind. Dazu gehören beispielsweise astronomische Beobachtungen, physikalische Messungen oder Sensordaten, die in Produktionsmaschinen zur Prozessoptimierung gesammelt werden.

Neue Daten und Methoden als Treibstoff für die Forschung

Clive Humbys Metapher ist im Kontext der Forschung in mehreren Hinsichten überraschend treffend. Denn digitale Daten bringen nicht nur neue Möglichkeiten mit sich, sondern auch neue Herausforderungen. Genau wie für das sprichwörtliche Öl braucht es nämlich für digitale Forschungsdaten neue Werkzeuge und Wissen, um diese Daten „zu fördern“. Zusätzlich sind „rohe“ digitale Forschungsdaten, ähnlich wie Rohöl, meist nicht direkt nutzbar, sondern müssen aufwändig aufbereitet und weiterverarbeitet werden, bevor sie überhaupt ausgewertet – und damit genutzt – werden können. Dazu gehört beispielsweise die Anreicherung mit anderen Datensätzen. So wie Motorenbenzin viele Zusätze enthält, um optimal auf den jeweiligen Motor abgestimmt zu sein, ist es für digitale Forschungsdaten in vielen Fällen sinnvoll oder sogar notwendig, diese um Umfragedaten oder qualitative Daten zu erweitern, um ihr volles Potenzial zu entfalten (Stier et al., 2020a).

Nicht zuletzt kann man mit Kerosin keinen Holzofen betreiben. Das bedeutet, dass für potenteren Treibstoff (in diesem Fall mehr

und detailliertere Informationen) geeignete Methoden entwickelt werden müssen, um diesen effizient nutzen zu können. Für digitale Forschungsdaten sind dies oft Methoden aus den Bereichen Data Science und Machine Learning, die es ermöglichen, größere und unstrukturiertere Daten zu verarbeiten und komplexere Zusammenhänge abzubilden als dies die „klassischerweise“ in den Sozialwissenschaften genutzten Methoden der deskriptiven und Inferenzstatistik können. Machine Learning ist dabei oft viel stärker datengetrieben als die sozialwissenschaftliche Inferenzstatistik. Statt ein Modell zu entwickeln und dann zu testen, ob dieses Modell zu den gesammelten Daten passt, geht Machine Learning typischerweise von den gesammelten Daten aus, und *lernt* das beste Modell, dass die Zusammenhänge in den Daten abbilden kann. Statt die linearen Zusammenhänge einiger weniger Faktoren auf kleinen Datensätzen zu modellieren, können diese komplexeren Modelle das Zusammenspiel von tausenden oder Millionen Einflussfaktoren in hochkomplexen Interaktionen darstellen.

Mit dieser Steigerung der Möglichkeiten im Umgang mit komplexen Datenstrukturen geht oft allerdings auch ein Verlust an Transparenz und Interpretierbarkeit einher. Neuronale Netzwerke sind beispielsweise als Machine-Learning-Ansatz gut geeignet, um komplexe Zusammenhänge zur Kategorisierung von Merkmalen zu erlernen, da sie anhand von beobachteten Daten „selbst lernen“, welche Einflussfaktoren in welcher Kombination die größte Rolle bei der Vorhersage spielen. Die erlernten Muster sind allerdings häufig mathematisch so komplex, dass sie für Menschen nicht sinnvoll interpretierbar sind (z.B. Rudin, 2019). Konkret bedeutet dies, dass es mit solchen Modellen möglich ist, mathematisch gute Kategorisierungen zu entwickeln, ohne im Detail zu verstehen, *warum* eine bestimmte Kombination aus Merkmalen zu einem bestimmten Ergebnis führt (vgl. Abbildung 1). Diese und ähnliche Ansätze werden daher auch als „black box“ bezeichnet (Rudin, 2019).

Digitale Forschungsdaten sind allerdings

nicht nur „das neue Öl“, wenn es um die Herausforderungen geht, die sich aus ihnen ergeben, sondern auch was ihre Möglichkeiten betrifft. Sie stellen für die Forschung einerseits eine Innovation in einzelnen, isolierten Disziplinen oder für ganz bestimmte Forschungsfragen dar. Andererseits fungieren sie als Innovationsmotor über Themenbereiche und disziplinäre Grenzen hinweg. So gibt es in fast allen (sozial-)wissenschaftlichen Disziplinen, von der Soziologie, über die Politikwissenschaft, die Kommunikations- und Medienwissenschaft oder die Psychologie immer mehr Forschungsarbeiten zu und mit digitalen Daten (Edelmann et al., 2020). Dabei ermöglichen diese Daten durch die Abbildung tatsächlichen Verhaltens in hoher Granularität einen neuen Blick auf etablierte Forschungsthemen wie z.B. Rassismus (z.B. Chaudhry, 2015), Sexismus (z.B. Samory et al., 2021), Extremismus (z.B. Aldera et al., 2021), Gruppendynamiken (z.B. Vega et al., 2021), Informationssuche und Nachrichtenkonsum (z.B. Stier et al., 2020b) oder Meinungsbildung (z.B. Kozitsin, 2022). Weiterhin ergeben sich durch digitale Daten neue Forschungsfragen, beispielsweise zu Phänomenen wie Filterblasen (z.B. Ross Arguedas, 2022), Interaktionen mit digitalen Technologien im Alltag (z.B. Faelens et al., 2021) oder dem Einfluss von Suchvorschlägen und Rankingalgorithmen auf menschliches Verhalten (z.B. Santos et al., 2021).

Wichtig ist dabei, dass digitale Forschungsdaten und Analysemethoden keinesfalls etablierte Datentypen und Methoden vollständig ersetzen können (Wilke et al., 2021). Diese Ansätze können sich im Gegenteil sehr gut gegenseitig ergänzen (Stier et al., 2020a). Umfragedaten erfassen oft subjektive Meinungen, Einstellung und Werte, Experimente können kausale Zusammenhänge offenlegen und qualitative Ansätze bieten mehr Möglichkeiten für tiefere Einblick in Prozesse (etwa des Verstehens) und Theorieentwicklung als die quantitative Arbeit mit digitalen Forschungsdaten. Im Gegenzug erfassen digitale Daten tatsächliches Verhalten, haben eine höhere



Abbildung 1

Cartoon zur problematischen Nutzung von Machine-Learning-Algorithmen. Statt interpretierbarer, theoriegeleiteter Hypothesen arbeiten diese Modelle oft mit rein datengetriebenen Vorhersagen. Sie können dafür allerdings komplexere Muster in großen Datensätzen besser abbilden als „klassische“ statistische Ansätze.

Quelle: <https://m.xkcd.com/1838/>

zeitliche und räumliche Auflösung, können oft in großem Umfang automatisiert erhoben werden und sind dabei zum Teil weniger umständlich in der Erhebung.

Eine letzte Gemeinsamkeit zwischen dem „Oil Boom“ und der Nutzung digitaler Forschungsdaten stellt das Potenzial für Schaden dar, der aus unsachgemäßem Gebrauch resultieren kann und die Notwendigkeit von klaren Regeln, die dies verhindern sollen. So wie in fossilen Brennstoffen eine enorme Menge Energie gespeichert ist, die nur zur richtigen Zeit, am richtigen Ort und zum richtigen Zweck freigesetzt werden sollte, enthalten digitale Daten oft persönliche und möglicherweise sensible Informationen über uns und unser soziales Umfeld, deren Nutzung u.U. „explosiv“ sein kann (um im Bild zu bleiben). So wie es für den Umgang mit fossilen Brennstoffen industrielle Sicherheitsstandards gibt, gibt es daher auch für digitale Forschungsdaten Leitlinien für einen ordnungsgemäßen Umgang. Dazu

gehören neben einem gesetzlichen Rahmen, wie ihn zum Beispiel die europäische Datenschutzgrundverordnung (DSGVO) vorgibt, auch ethische Richtlinien und Empfehlungen innerhalb der wissenschaftlichen Forschung zum Umgang mit solchen Daten (z.B. franzke et al., 2020).

Digitale Forschungsdaten als „das neue Öl“ bringen also sowohl zahlreiche Chancen und Potenziale als auch spezifische Herausforderungen und Risiken für die sozialwissenschaftliche Forschung mit sich. Aber was bedeutet das konkret für einzelne Forschungsprojekte? Das lässt sich für die Digitalisierungsforschung als methodisch, theoretisch und disziplinär enorm heterogenes Feld nicht pauschal beantworten (zur allgemeinen Entwicklung des Feldes siehe auch Schmitt et al., 2023). Mit der vorliegenden Ausgabe von *easy_social_sciences* wollen wir trotzdem den Versuch unternehmen, den Lesenden anhand von ausgewählten Forschungsarbeiten die Vielfalt an digitalen Forschungsdaten, -fragen und -methoden sowie die damit verbundenen Potenziale und Herausforderungen näher zu bringen.

Ein Einblick in die Digitalisierungsforschung aus fünf Perspektiven

In der letzten Ausgabe *easy* 68 „Digitale Gesellschaft(en)“ haben wir uns bereits der Digitalisierungsforschung gewidmet und eine Auswahl an Querschnittsthemen aus der Vogelperspektive betrachtet. Unser Fokus spannte sich von einer initialen Begriffserklärung über die Identifizierung relevanter Forschungsfragen, Erläuterungen zur Qualität neuer Datentypen sowie der Potenziale und Herausforderungen bei der Verknüpfung verschiedener Datentypen bis zur Kommunikation von Ergebnissen der Digitalisierungsforschung in die Gesellschaft. Die einzelnen Beiträge haben dabei bewusst Themen aufgegriffen, die die Forschung eher von einer

Metaperspektive betrachten und sind weniger auf einzelne, inhaltliche Forschungsarbeiten eingegangen.

Im Gegensatz dazu möchten wir in dieser Ausgabe inhaltliche Arbeiten in den Mittelpunkt stellen, die sich aus verschiedenen disziplinären, methodischen und inhaltlichen Perspektiven mit Fragen der Digitalisierungsforschung sowie der Nutzung digitaler Forschungsdaten auseinandersetzen. Dazu haben wir Forschende aus verschiedenen sozialwissenschaftlichen Bereichen eingeladen, uns einen Einblick in ihre Forschungsprojekte zu geben.

Im ersten Beitrag, „Digitale Methoden in der Politikwissenschaft – Auf dem Weg in virtuelle Welten?“, beschreibt *Isabelle Borucki* aus politikwissenschaftlicher Perspektive die Kategorisierung von analogen, digitalen und nativ-digitalen Daten und Methoden und erläutert, wie eine Integration von neuen Daten und Methoden in etablierte Forschungsprozesse erfolgreich sein kann. Der Beitrag unterstreicht vor allem, dass digitale Forschungsdaten und -methoden nicht in Konkurrenz zu etablierten Ansätzen stehen, sondern im Gegenteil große Kombinationspotenziale aufweisen.

Im zweiten Beitrag, „The Dark Web. A Brief Introduction“, nimmt *Felix Soldner* die Lesenden mit in das sogenannte Dark Web – also den Teil des Internets, der nur über Verschlüsselungs- und Anonymisierungsdienste wie z.B. das TOR-Netzwerk zugänglich ist. Wenngleich die Zugänglichkeit zu diesen Daten für Forschende erschwert ist, ermöglicht der Blick ins Dark Web Einsichten in Interaktionen und Prozesse, die sich Forschenden oftmals entziehen. Dazu gehören vor allem kriminelle Handlungen wie der Handel mit illegalen Waren und Dienstleistungen oder Fälschungen, aber auch Interaktionen von politischen Aktivist*innen in Ländern, in denen Zensur herrscht.

Im dritten Beitrag, „Jung, weiblich und extrem rechts. Die narrative Kommunikation weiblicher Akteurinnen auf der Plattform Instagram“, zeigt *Sandra Kero* auf, wie Daten der Plattform Instagram – insbesondere Bil-

der aus Instagram-Stories – genutzt werden können, um detaillierte Einblicke in die Kommunikationsstrategien rechter Accounts zu erhalten. Dabei betrachtet sie eine Gruppe, die bisher im öffentlichen Diskurs und der wissenschaftlichen Forschung kaum beachtet wird: Frauen, die rechte Inhalte erstellen und teilen. Methodisch verbindet sie digitale Daten der Plattform mit qualitativer Inhaltsanalyse und quantitativer Auswertung.

Im vierten Beitrag, „Periods in the Public Eye – Investigating Risk Perceptions of Data Sharing in Reproductive Health Applications“ von *Annika Deubel* und *Pauline Heger*, geht es im weitesten Sinne um „Sensoren“ – allerdings nicht für die Erfassung von öffentlich verfügbaren Daten, sondern von persönlichen Gesundheitsinformationen. Konkret betrachten die Autor*innen basierend auf Daten einer Online-Befragung den Zusammenhang zwischen der Nutzung von Apps zur Erfassung des eigenen Menstruationszyklus und den Einstellungen der Nutzenden zum Datenschutzrisiko, welches diese Apps bei der Verarbeitung von sensiblen, gesundheitsbezogenen Daten mit sich bringen.

Im letzten Beitrag der Ausgabe, „Making Sense of the Big Data Mess. Why Interdisciplinarity Matters in Smart Cities“, verlassen wir mit den Autor*innen *Niklas Frechen*, *Pauline Heger*, *Christoph Bieber* und *Mennatullah Heddawy* das Internet und erhalten einen Überblick über digitale Daten, die von Sensoren wie Stromzählern, Wetterstationen oder Ampelschaltungen im Alltag von Stadtbewohnenden erhoben werden. Die Autor*innen beschreiben mit kritischem Blick die Potenziale dieser vielfältigen Datentypen in sogenannten „Smart Cities“ und wie diese durch interdisziplinäre Zusammenarbeit und öffentliche Zugänglichkeit zur politischen Entscheidungsfindung genutzt werden können.

Wir hoffen, mit diesen vielfältigen Beiträgen einen Einblick in die Themen und Methoden der Digitalisierungsforschung geben zu können, und ein Bewusstsein dafür zu schaffen, wie digitale Forschungsdaten über Fachdisziplinen hinweg die Erforschung

sowohl neuer als auch etablierter Fragen ermöglichen, welche Herausforderungen sich dabei für die Forschenden ergeben, und wie mit diesen produktiv umgegangen werden kann. Wir können dabei natürlich nur einen kleinen Einblick in ein dynamisches, thematisch sowie methodisch breites und sehr heterogenes Forschungsfeld liefern, hoffen aber trotzdem, den Leser*innen dieser Ausgabe die Forschung zu und mit digitalen Daten ein Stück näher bringen zu können.

Referenzen

- Aldera, S., Emam, A., Al-Qurishi, M., Alrubaian, M. & Alothaim, A. (2021). Online extremism detection in textual content: A systematic literature review. *IEEE Access*, 9, 42384–42396. <https://doi.org/10.1109/ACCESS.2021.3064178>
- Amazon (2023, 30. Juni). Amazon.com Privacy Notice. Abgerufen am 10.07.2023, von <https://www.amazon.com/gp/help/customer/display.html?nodeId=GX7NJQ4ZB8MHFRNJ>
- Borucki, I. (2023). Digitale Methoden in der Politikwissenschaft. Auf dem Weg in virtuelle Welten. *easy_social_sciences*, 69, 11–17. <https://doi.org/10.15464/easy.2023.08>
- Chaudhry, I. (2015). #Hashtagging hate: Using Twitter to track racism online. *First Monday*, 20(2). <https://doi.org/10.5210/fm.v20i2.5450>
- Deubel, A & Heger, P. (2023). Periods in the Public Eye. Investigating Risk Perceptions of Data Sharing in Reproductive Health Applications. *easy_social_sciences*, 69, 37–44. <https://doi.org/10.15464/easy.2023.11>
- eBay (2023, 24. März). eBay User Privacy Notice. Abgerufen am 10.07.2023, von <https://www.ebayinc.com/company/privacy-center/privacy-notice/>
- Edelmann, A., Wolff, T., Montagne, D. & Bail, C. A. (2020). Computational social science and sociology. *Annual Review of Sociology*, 46, 61–81. <https://doi.org/10.1146/annurev-soc-121919-054621>
- Faelens, L., Hoorelbeke, K., Soenens, B., Van Gaevelen, K., De Marez, L., De Raedt, R. & Koster, E. H. (2021). Social media use and well-being: A prospective experience-sampling study. *Computers in Human Behavior*, 114, 106510. <https://doi.org/10.1016/j.chb.2020.106510>
- Fitbit (2023, 6. Juni). Fitbit Privacy Policy. Abgerufen am 10.07.2023, von <https://www.fitbit.com/global/us/legal/privacy-policy#how-we-use-info>

- Frechen, N., Heger, P., Bieber C., & Hendawy, M. (2023). Making Sense of the Big Data Mess. Why Interdisciplinarity Matters in Smart Cities. *easy_social_sciences*, 69, 45–52. <https://doi.org/10.15464/easy.2023.12>
- 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>
- Howison, J., Wiggins, A. & Crowston, K. (2011). Validity issues in the use of social network analysis with digital trace data. *Journal of the Association for Information Systems*, 12(12), 767–797. <https://doi.org/10.17705/1jais.00282>
- Humby, C. & Palmer, M. (2006, 3. November). Data is the new oil. Abgerufen am 10.07.2023, von https://ana.blogs.com/maestros/2006/11/data_is_the_new.html
- Keuschnigg, M., Lovsjö, N. & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, 1, 3–14. <https://doi.org/10.1007/s42001-017-0006-5>
- Kero, S. (2023). Jung, weiblich und extrem rechts. Die narrative Kommunikation weiblicher Akteurinnen auf Instagram. *easy_social_sciences*, 69, 28–36. <https://doi.org/10.15464/easy.2023.10>
- Kozitsin, I. V. (2022). Formal models of opinion formation and their application to real data: Evidence from online social networks. *The Journal of Mathematical Sociology*, 46(2), 120–147. <https://doi.org/10.1080/0022250X.2020.1835894>
- Meta (2022, 11. Januar). Facebook Datenrichtlinie. Abgerufen am 10.07.2023, von <https://www.facebook.com/about/privacy/update/printable>
- Microsoft (2023, April). Data collection summary for Windows. Abgerufen am 10.07.2023, von <https://privacy.microsoft.com/en-us/data-collection-windows>
- Netflix (2022, 1. November). Netflix Privacy Statement. Abgerufen am 10.07.2023, von <https://help.netflix.com/legal/privacy>
- Ross Arguedas, A., Robertson, C., Fletcher, R. & Nielsen, R. (2022). Echo chambers, filter bubbles, and polarisation: A literature review. Reuters Institute for the Study of Journalism.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Samory, M., Sen, I., Kohne, J., Flöck, F. & Wagner, C. (2021). “Call me sexist, but...”: Revisiting sexism detection using psychological scales and adversarial samples. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1), 573–584. <https://doi.org/10.1609/icwsm.v15i1.18085>
- Samsung (2023). Samsung Privacy Policy Overview. Abgerufen am 10.07.2023, von <https://privacy.samsung.com/policy/overview>
- Santos, F. P., Lelkes, Y. & Levin, S. A. (2021). Link recommendation algorithms and dynamics of polarization in online social networks. *Proceedings of the National Academy of Sciences*, 118(50), e2102141118. <https://doi.org/10.1073/pnas.2102141118>
- Schmitt, J. B., Kohne, J. & Breuer, J. (2023). Digitalisierungsforschung. Wie wir die digitale Transformation wissenschaftlich erfassen können. *easy_social_sciences*, 68, 4–11. <https://doi.org/10.15464/easy.2023.01>
- Soldner, F. (2023). The Dark Web. A Brief Introduction. *easy_social_sciences*, 69, 18–27. <https://doi.org/10.15464/easy.2023.09>
- Sonderegger, P. (2021, 4. März). Data hits peak metaphor. Abgerufen am 10.07.2023, von <https://paulsonderegger.com/2021/03/04/data-hits-peak-metaphor/>
- Spotify (2023, 22. Februar). Datenschutzrichtlinie von Spotify. Abgerufen am 10.07.2023, von <https://www.spotify.com/de/legal/privacy-policy/>
- Stier, S., Breuer, J., Siegers, P. & Thorson, K. (2020a). 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>
- Stier, S., Kirkizh, N., Froio, C. & Schroeder, R. (2020b). Populist attitudes and selective exposure to online news: A cross-country analysis combining web tracking and surveys. *The International Journal of Press/Politics*, 25(3), 426–446. <https://doi.org/10.1177/1940161220907018>
- Varian, H. R. (2014). Beyond big data. *Business Economics*, 49(1), 27–31. <https://doi.org/10.1057/be.2014.1>
- Vega, L., Mendez-Vazquez, A. & López-Cuevas, A. (2021). Probabilistic reasoning system for social influence analysis in online social networks. *Social Network Analysis and Mining*, 11(1). <https://doi.org/10.1007/s13278-020-00705-z>
- Wagner, C., Strohmaier, M., Olteanu, A., Kıcıman, E., Contractor, N. & Eliassi-Rad, T. (2021). Measuring algorithmically infused societies. *Nature*, 595, 197–204. <https://doi.org/10.1038/s41586-021-03666-1>
- Wilke, R., Knoblauch, H., Kohne, J., Miller, B., Strohmaier, M., Wagner, C., Wolf, C., Hanekop, H., Heuer, J-O, Hollstein, B. & Mozygamba, K. (2021). Symposium Forschungsdateninfrastruktur. *Soziologie*, 50(4), 430–472.
- Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of information technology*, 30(1), 75–89. <https://doi.org/10.1057/jit.2015.5>

Julian Kohne

GESIS – Leibniz-Institut für Sozialwissenschaften

E-Mail julian.kohne@gesis.org

Julian Kohne ist wissenschaftlicher Mitarbeiter bei GESIS, Köln und Doktorand der Psychologie an der Universität Ulm. Bei GESIS arbeitet er an der Entwicklung zugänglicher Smartphone-Datenerfassungsinfrastrukturen für Sozialwissenschaftler*innen. In seinem Promotionsprojekt entwickelt er transparente Methoden zur Datenspende für WhatsApp-Chatprotokolle und untersucht, wie unterschiedliche soziale Beziehungen in Online-Kommunikationsmustern ausgedrückt werden. Sein übergreifendes Forschungsinteresse gilt der Nutzung digitaler Verhaltensdaten, insbesondere der Text- und Netzwerkdaten, um sozialwissenschaftlicher Forschungsfragen zu bearbeiten. <https://www.juliankohne.com>

Josephine B. Schmitt

Center for Advanced Internet Studies (CAIS)

E-Mail Josephine.Schmitt@cais-research.de

Josephine B. Schmitt ist wissenschaftliche Koordinatorin am Center for Advanced Internet Studies (CAIS). Dort befasst sie sich u.a. mit der Erforschung und Entwicklung von innovativen Konzepten für die interdisziplinäre Zusammenarbeit in der Digitalisierungsforschung. Sie forscht weiterhin zu Inhalt, Verbreitung und Wirkung von Hate Speech, extremistischer Propaganda und (politischen) Informations- und Bildungsangeboten im Internet. <https://orcid.org/0000-0002-4689-3049>

Johannes Breuer

GESIS Leibniz-Institut für Sozialwissenschaften & Center for Advanced Internet Studies (CAIS)

E-Mail johannes.breuer@gesis.org

Johannes Breuer ist Senior Researcher im Team Digital Society Observatory bei GESIS, Köln und Leiter des Teams Research Data & Methods am Center for Advanced Internet Studies (CAIS). Seine Forschungsinteressen sind die Nutzung und Wirkung digitaler Medien, Computational Methods, digitale Spurdaten, Datenmanagement und Open Science.

<https://www.johannesbreuer.com>
