SSOAR

Open Access Repository
www.ssoar.info

Die Verknüpfung von digitalen Verhaltensdaten und Umfragedaten

Weiß, Julia; Stier, Sebastian

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:
Weiß, J., & Stier, S. (2023). Die Verknüpfung von digitalen Verhaltensdaten und Umfragedaten. *easy_social_sciences*, 68, 31-38. https://doi.org/10.15464/easy.2023.04

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

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

doi: 10.15464/easy.2023.04

Die Verknüpfung von digitalen Verhaltensdaten und Umfragedaten

Julia Weiß & Sebastian Stier

Während die Sozialwissenschaften seit Jahrzehnten Daten aus Umfragen verwenden, stehen in den letzten Jahren zunehmend auch digitale Verhaltensdaten, z.B. Textdaten aus sozialen Medien oder Informationen über das Nutzungsverhalten von Smartphones, zur Verfügung. Beide Datenformen bringen dabei spezifische Potentiale und Limitationen mit sich, welche durch deren Verknüpfung adressiert werden können. Dennoch findet diese Verknüpfung bisher nur in verhältnismäßig wenigen Studien statt. Der vorliegende Beitrag erörtert, warum es sinnvoll sein kann Befragungen und digitale Verhaltensdaten miteinander zu verknüpfen, wie dadurch Forschung zu neuen Themen ermöglicht wird, welche Wege es für die Verknüpfung gibt und welche Herausforderungen damit einhergehen.

While the social sciences have been using survey data for decades, digital behavioral data, e.g., text data from social media or information about behavior on smartphones, have become increasingly available in recent years. Both data forms have specific potentials and limitations, which can be addressed by linking both with each other. Although the combination of these two forms of data opens up new potentials for social science research, this has so far only been done in comparatively few studies. This article discusses why it can make sense to link surveys and digital behavioral data, how this enables research on new topics, what ways there are for linking them, and what challenges arise.

Keywords: Data linking, Umfragedaten, digitale Verhaltensdaten

Die Digitalisierung bringt im Privaten, wie auch auf gesellschaftlicher Ebene einen Strukturwandel mit sich. Wir alle agieren tagtäglich mit soziotechnischen Systemen, beispielsweise lesen oder schreiben wir etwas in sozialen Medien, tragen eine smarte Uhr, die unseren Standort aufzeichnet, oder kaufen online etwas ein. Die dadurch entstehenden Interaktionen zwischen Menschen und digitaler Technologie können dabei soziale Strukturen und Prozesse wie z.B. die Verstärkung von Meinungsunterschieden (Polarisierung) beeinflussen oder sogar neue Phänomene wie die kollaborative Wissensproduktion und -verbreitung über Wikipedia entstehen lassen. Ob die Digitalisierung unter dem Strich insgesamt wünschenswert oder problematisch ist, soll an anderer Stelle beurteilt werden, klar ist jedoch, dass man für eine solche Beurteilung diese Interaktionen und Einflüsse zwischen Menschen und digitaler Technologie näher betrachten muss. Aus Sicht der Sozialwissenschaften eröffnet sich hier ein großes Forschungsfeld und damit einhergehend die

Frage, welche Daten man zur Beforschung dieser Phänomene zur Verfügung hat oder neu erhalten kann.

Seit Jahrzehnten basiert ein erheblicher Teil der empirischen Sozialforschung auf Umfragedaten. Täglich finden Umfragen zu allerlei Themen statt: Sei es im Bereich der kommerziellen Markt- und Meinungsforschung, zur Wahrnehmung von Themen und Beurteilung von Produkten oder in der akademischen Forschung, etwa im Kontext von größeren Ereignissen wie etwa Wahlen und den Fragen rund um das Wahlverhalten. Die mit den resultierenden Daten stattfindende Forschung ist dabei vielfältig. Sie stößt aber insbesondere im Bereich der Beforschung des Zusammenspiels von Menschen und digitaler Welt an ihre Grenzen. Mit der immer größer werdenden Verfügbarkeit von digitalen Verhaltensdaten eröffnen sich neue Möglichkeiten zur Lösung dieser Herausforderung in der Forschung, und gleichzeitig führt der neue Datentyp zur Etablierung eines eigenständigen Forschungsfeldes. Digitale Verhaltensdaten kann man als Aufzeichnungen von Aktivitäten, die über ein digitales Informationssystem durchgeführt wurden, definieren (Howison et al., 2011). Dabei handelt es sich um eine Vielzahl von digitalen Informationssystemen, wie etwa Webseiten, Social-Media-Plattformen, Smartphone-Apps oder Sensoren (z.B. GPS oder Bluetooth). Hierbei können beispielsweise Meinungsäußerungen in Form von Textdaten auf Plattformen, das Interaktionsverhalten mit Freund*innen oder auch die Klickzahlen für Nachrichtenseiten von Interesse sein. Wie das Zusammenspiel aus Umfragedaten und digitalen Verhaltensdaten erfolgen kann, warum es überhaupt sinnvoll sein kann die beiden Datenformen miteinander zu verbinden und welche zukünftigen Wege denkbar und wünschenswert sind, werden wir im Folgenden darlegen.

Warum kann es sinnvoll sein, digitale Verhaltensdaten mit Umfragedaten zu verknüpfen?

Sowohl digitale Verhaltensdaten als auch Umfragedaten bringen jeweils Vor- und Nachteile mit sich. Die Vorteile digitaler Verhaltensdaten liegen dabei insbesondere in der Möglichkeit der aktiven und passiven Messung von „natürlichem“ menschlichen Verhalten in einem zuvor nicht dagewesenen Maßstab und Detailgrad. Digitale Verhaltensdaten können dabei in vielen Formen und aus vielen Quellen für wissenschaftliche Zwecke relevant sein. Als grobe Gruppierungsmöglichkeit hat sich die Unterscheidung zwischen „found data“, also z.B. öffentliche Kommunikation in sozialen Netzwerken, und spezifisch für wissenschaftliche Zwecke erhobenen Daten („designed data“), z.B. das Aufzeichnen des Internetnutzungsverhalten von Personen, die im Rahmen eines Projektes dem zugestimmt haben, bewährt. Gefundene digitale Verhaltensdaten („found data“) sind vielfältig, sie umfassen Texte, Bilder oder Videos, die allesamt für sozialwissenschaftliche Fragestellungen relevant sein können. Sie werden in großer Menge erzeugt und weisen oft eine genauere räumliche und zeitliche Auflösung als Umfragedaten auf, zum Beispiel bis hin zu einem sekundengenauen Zeitstempel, wann ein Inhalt erzeugt wurde. Zudem können Wissenschaftler*innen gefundene digitale Verhaltensdaten in Echtzeit erheben, da diese ständig und unabhängig von einzelnen Forschungsprojekten durch die Nutzung digitaler Angebote entstehen. Nachteilig ist jedoch, dass gefundene digitale Verhaltensdaten, z.B. die wissenschaftlich viel genutzten Twitter-Daten, meist nur begrenzt Informationen über die Produzierenden der Daten umfassen, die selten über den pseudonymisierten Twitter-Accountnamen und äußerst lückenhaft berichtete Informationen (wie z.B. dem Wohnort, Beruf oder Geschlecht in der freiwilligen Profilbeschreibung) hinaus gehen.

Eigens für wissenschaftliche Zwecke generierte Daten („designed data“) können beispielsweise Erhebungen durch eine Smartphone-App zur mobilen Befragung und Erhebung von Sensordaten sein oder aber die Erfassung der Webseiten-Besuche durch ein Browser-Plugin. Diese Arten von Datenerhebungen geben Wissenschaftler*innen mehr Transparenz und Kontrolle zur Sicherung wissenschaftlicher Standards, als wenn man beispielsweise Datensätze von Firmen erhält (z.B. Kaufverhalten auf digitalen Marktplätzen oder Suchanfragen in digitalen Suchmaschinen), die diese kommerziell erheben und kaum bis keine Informationen über das Entstehen der Datensätze preisgeben. Gleichzeitig erfordern Erhebungen von „designed data“ einen erheblich höheren Ressourceneinsatz der Wissenschaftler*innen, da sowohl mit der Erhebung solcher Daten selbst Kosten einhergehen als auch spezifische Kenntnisse zur technisch korrekten Erhebung häufig erst einmal erlernt werden müssen.

» *Verbindet man digitale Verhaltensdaten mit Umfragedaten, so können sich die jeweiligen Nachteile ausgleichen und produktive Synergien entstehen.* «

Beiden Typen von digitalen Verhaltensdaten ist jedoch gemein, dass sie für sich stehend beschränkte oder gar keine Informationen über die datengenerierenden Personen enthalten. Basierend auf digitalen Verhaltensdaten ist so zum Beispiel generell keine direkte Messung von Einstellungen möglich, wir kennen also z.B. die Meinungen und Überzeugungen der Personen hinter einem Beitrag in den sozialen Medien nicht. Darüber hinaus enthalten digitale Verhaltensdaten keine Informationen zu den Offlineverhaltensweisen der Individuen. All dies mündet in der Frage, inwiefern mit der Analyse dieser Form der Daten überhaupt Aussagen über die gesamte Bevölkerung oder zumindest bestimmte Gruppen getroffen werden können. Im Falle der für wissenschaftliche Zwecke generierten („designed“) Daten ist es möglich die Stichprobe, also eine Auswahl an Personen, die stellvertretend für eine Grundgesamtheit (z.B. die deutsche Bevölkerung oder alle Twitter-Nutzenden) ausgewählt werden, zu bestimmen, um so später Aussagen über die Gesamtbevölkerung treffen zu können. Bei den gefundenen Daten ist es dagegen nicht möglich festzulegen, wie sich eine Stichprobe zusammensetzt. So sind später auch keine Aussagen über die Grundgesamtheit möglich.

Der Vorteil von Umfragedaten liegt darin, dass sie die Erhebung verschiedener Konzepte ermöglichen, die Teilnehmende am besten selbst einschätzen können, wie z.B. persönliche Werte, Einstellungen und Überzeugungen. Es kann aber auch Offline- und Onlineverhalten erfragt werden. Darüber hinaus ist es hier möglich, eine für die Grundgesamtheit stellvertretende Auswahl an Befragten festzulegen. Nachteilig ist jedoch, dass Selbstauskünfte der Befragten zu ihrem Onlineverhalten nur bedingt valide sind (Parry et al., 2021). So kann es sein, dass Befragte sich schlicht nicht erinnern oder nur schwer einschätzen können, wie viele Stunden sie beispielsweise täglich oder auf welchen Websites sie Nachrichten konsumieren. Darüber hinaus kann soziale Erwünschtheit die Ergebnisse einer Befragung beeinflussen, wenn sich also Befragte bei ihrer Antwort an sozialen Normen und nicht an ihrer tatsächlichen Einstellung oder ihrem tatsächlichen Verhalten orientieren. Schließlich haben Umfragen mit sinkender Teilnahmebereitschaft, also mit einem zunehmend geringeren Anteil an tatsächlich realisierten Interviews an der ursprünglich ausgewählten Stichprobe, zu kämpfen.

Verbindet man digitale Verhaltensdaten mit Umfragedaten so können sich die jeweiligen Nachteile ausgleichen und produktive Synergien entstehen (Stier et al., 2020). Beispielsweise liegen durch die Umfragedaten

Informationen über die Produzierenden der Daten vor, wir wissen also etwa wie alt die befragte Person ist und welche politischen Einstellungen sie hat. Mit Hilfe der digitalen Verhaltensdaten werden detaillierte Informationen beispielsweise zum Onlinenutzungsverhalten erhoben, welche nun nicht mehr durch Schätzungen der Befragten erhoben werden müssen. Das Verbinden beider Datenformen bringt neben der Lösung zuvor vorhandener Probleme, wie etwa dass Befragte sich nun nicht mehr zu ihrem Onlineverhalten erinnern müssen, auch spezifische eigene Vorteile mit sich. So eröffnen sich für die Wissenschaft neue Forschungsfelder und für die Befragten kann sich der Aufwand reduzieren, da die Belastung durch die Teilnahme an einer Umfrage reduziert wird, wenn Informationen stattdessen als digitale Verhaltensdaten erhoben werden.

Wie lassen sich digitale Verhaltensdaten und Umfragedaten verknüpfen?

Grundsätzlich sind für die Verknüpfung von digitalen Verhaltensdaten und Umfragedaten mindestens zwei Wege denkbar. Zum einen kann man ausgehend von einer Umfrage die Teilnehmenden einladen, ihre digitalen Verhaltensdaten zu teilen. Dieses Teilen der digitalen Verhaltensdaten kann dabei auf unterschiedliche Arten erfolgen: Die Abfrage der Profilnamen der Nutzenden in sozialen Medien ermöglicht Forschenden, im Anschluss die Daten der entsprechenden Profile in den sozialen Medien zu erheben (Al Baghal et al., 2020). Durch Studienteilnehmende installierte Apps (Keusch et al., 2022) oder Erweiterungen (Plugins) für den Webbrowser (Christner et al., 2022) erheben über einen gewissen Zeitraum hinweg in Echtzeit deren Verhalten. Schließlich besteht die Option der sogenannten Datenspenden, bei denen die Teilnehmenden einer Umfrage ihre digitalen Verhaltensdaten selbst herunterladen und an die Forschenden übermitteln (Boeschoten et al., 2022). Zum anderen können, ausgehend von digitalen Verhaltensdaten, Teilnehmende für eine Umfrage gewonnen werden. So ist es möglich, über Werbung auf sozialen Medien oder die direkte Ansprache von Nutzenden digitaler Anwendungen Menschen für die Beantwortung einer Umfrage zu rekrutieren.

Insgesamt kann das Verknüpfen der Daten dabei in unterschiedlichen Dimensionen variieren. In Bezug auf die zeitliche Abfolge der Verknüpfung ist es sowohl möglich Umfragedaten und digitale Verhaltensdaten gemeinsam, etwa innerhalb eines Projekts, zu erheben (ex-ante Linking), als auch einen bereits bestehenden Datensatz mit einem nachträglich erhobenen Datensatz zu verbinden (ex-post Linking) (Stier et al., 2020). Darüber hinaus kann die Verknüpfung von Umfragedaten und digitalen Verhaltensdaten auf verschiedenen Ebenen stattfinden. Also sowohl auf individueller Ebene, wenn etwa eine Person an einer Umfrage teilnimmt und von genau dieser Person auch digitale Verhaltensdaten erhoben werden, als auch auf Aggregatsebene, beispielsweise durch die Verknüpfung von Durchschnittsergebnissen aus Umfragen und digitalen Verhaltensdaten über verschiedene Zeitpunkte oder geografische Regionen.

Welche Herausforderungen gehen mit der Verknüpfung einher?

Die Herausforderungen, welche mit der Verknüpfung von digitalen Verhaltensdaten und Umfragedaten einhergehen, umfassen praktische Hürden sowie rechtliche und ethische Aspekte.

Die praktischen Hürden bei der Verknüpfung können vielfältig sein (Breuer et al., 2022). Geht es um die Erhebung einer bestimmten Nutzungsart, z.B. die Smartphone-Nutzung, so können diejenigen, die kein Smartphone

> ***Verknüpfung von digitalen Verhaltensdaten und Umfragedaten auf individueller Ebene erst nach Aufklärung und Einwilligung der Teilnehmenden.***

besitzen oder von den Forschenden zur Verfügung gestellt bekommen, nicht an der Studie teilnehmen. Auf Seiten der Teilnehmenden einer Umfrage, welche gebeten werden ihre digitalen Verhaltensdaten zu teilen, können fehlende technische Kenntnisse zu Problemen führen. Vermeintlich einfache Dinge, wie das Wiederaufrufen von Zugangsdaten oder Nutzendennamen können die Erhebung von digitalen Verhaltensdaten erschweren. Dies kann dazu führen, dass bestimmte Gruppen mit dieser Erhebungsform nicht erreicht werden können. So werden beispielsweise bei Onlineumfragen Menschen ohne Internetzugang oder bei Erhebungen mittels einer Erweiterung (Plugin) für den Webbrowser, Personen mit mangelnden technischen Kenntnissen von der Teilnahme ausgeschlossen. Gleichzeitig ist es Aufgabe der Forschenden, Erhebungen möglichst so zu gestalten, dass praktische Hürden auf Seiten der Befragten minimiert werden. Dafür sind auch auf Seiten der Forschenden entsprechende technische Kenntnisse erforderlich.

Die konkrete Verknüpfung von digitalen Verhaltensdaten und Umfragedaten birgt selbst einige Herausausforderungen. Die Frage nach der Teilnahme an einer Erhebung von digitalen Verhaltensdaten kann von den Befragten in einer Umfrage unbeantwortet bleiben oder sogar zum Abbruch der Befragung führen. Für die Forschenden bringt das nicht nur höhere Kosten für die Erzielung ihrer benötigten Fallzahlen mit sich, sondern auch durch potenziell verzerrte Daten und wissenschaftliche Fehlschlüsse, wenn beispielsweise online-affine Gruppen mit schwach ausgeprägten Einstellungen zum Datenschutz in Studiensamples überrepräsentiert sind. Bisherige Forschung zeigt, dass in Panelbefragungen, wenn man also mit zeitlichem Abstand immer wieder dieselben Personen befragt, mehr Befragte bei der Frage, ob sie bereit wären ihre digitalen Verhaltensdaten zur Verfügung zu stellen, abbrechen bzw. diese Frage unbeantwortet übergehen als bei anderen Anfragen. Die Wirkung lässt jedoch in den folgenden Wellen nach. Das erscheint hinnehmbar angesichts des Mehrwerts, der durch die Verbindung von Umfrage- und Verhaltensdaten erreicht werden kann (Trappmann et al., 2022).

Aus rechtlicher und ethischer Perspektive kommen weitere Herausforderungen für die Verknüpfung von digitalen Verhaltensdaten und Umfragedaten hinzu. So sind beispielsweise im Falle der Erhebung von Daten aus sozialen Medien stets die Nutzungsbedingungen der Plattform zu beachten. Diese untersagen zum Teil die automatisierte Erhebung von öffentlich einsehbaren Informationen oder setzen sogar technische Hürden ein, um dies zu verhindern. Darüber hinaus muss der Datenschutz gewahrt und die seit 2018 anwendbare Datenschutzgrundverordnung (DSGVO) eingehalten werden. Dies bedeutet für die Forschenden nun umfangreicher die Einhaltung des Datenschutzes dokumentieren zu müssen, eröffnet aber gleichzeitig eine neue Form des Zugangs zu digitalen Verhaltensdaten. Konkret wahrt die DSGVO für die Nutzenden das Recht auf Auskunft, also das Recht darauf, von den Betreibenden sozialer Medien die über sie gespeicherten Daten zu erhalten (Boeschoten et al., 2022). Diese können dann im Rahmen von Datenspenden für die Forschung nutzbar gemacht werden. Generell kann die Erhebung sowie die Verknüpfung von digitalen Verhaltensdaten und Umfragedaten auf individueller Ebene erst nach Aufklärung und Einwilligung der Teilnehmenden stattfinden. Die Aufgabe der Forschenden ist es dabei, den Befragten alle nötigen Informationen zu geben, damit diese informiert entscheiden können, ob

sie einer Datenerhebung zustimmen oder nicht. Dies muss in derart strukturierter und überschaubarer Form stattfinden, dass die Befragten nicht mit unnötigen Informationen überfordert werden und somit wohlmöglich ihre Zustimmung geben, ohne die Folgen ihrer Zustimmung verstanden zu haben (Beuthner et al., 2022; Walzenbach et al., 2022).

> ***Öffentliche Infrastrukturen, die Forschenden den Zugang zu digitalen Verhaltensdaten erleichtern, sind deshalb unverzichtbar.***

Sind die Daten einmal erhoben und miteinander verbunden, ergeben sich neue Herausforderungen im Bereich der Analyse, Archivierung und Wiederverwendbarkeit des entstandenen Datensatzes. In den Sozialwissenschaften gängige statistische Analyseverfahren sind nicht ohne Weiteres auf derart vielschichtige Datensätze anwendbar (Stier et al., 2020). Stattdessen bedarf es der Verwendung und Kombination unterschiedlicher Methoden und für individuelle Forschende womöglich auch die Erlernung neuer Verfahren zur Auswertung der Daten.

Handelt es sich bei dem entstandenen Datensatz um Daten, die nicht öffentlich verfügbar sind, stellt sich die Frage wie der Datensatz an andere Forschende weitergegeben werden kann. Es ist ein wichtiges Grundprinzip einer „offenen Wissenschaft“ (*Open Science*), die Überprüfung von Forschungsergebnissen und Sekundärnutzung von Daten zu ermöglichen. Zudem geben Drittmittelgeber oder auch Verlage und Fachzeitschriften regelmäßig vor, dass Daten für die Öffentlichkeit verfügbar gemacht werden müssen (Breuer et al., 2020). Dabei entstehen Zielkonflikte mit den Nutzungsrichtlinien von Online-Plattformen und der Notwendigkeit, dem Datenschutz genüge zu tragen.

Was kann die Zukunft bringen?

Eine eigene Forschungsinfrastruktur aufzubauen, ist für Individuen und viele Forschungsgruppen technisch herausfordernd und kostenintensiv. Kommerzielle Anbieter zur Erhebung von Umfrage- und digitalen Verhaltensdaten sind teuer, arbeiten zum Teil intransparent und sind meist grundsätzlich auf die Marktforschung ausgerichtet. Für Forschende verschließt all dies häufig den Zugang zu einem möglichen Forschungsfeld. Institutionalisierte öffentliche Infrastrukturen, die Forschenden den Zugang zur Erhebung und Auswertung von digitalen Verhaltensdaten erleichtern, sind deshalb unverzichtbar.

Darüber hinaus bedarf es vielfältiger Forschung zu diesem immer noch relativ „neuen“ Datentyp. Dies beinhaltet theoretische Arbeiten, welche die außerordentliche Informationsvielfalt von verknüpften Datensätzen aufgreifen und methodische Zugriffe, mit denen sich diese Datenkombinationen bestmöglich analysieren lassen. Während die Auswertung von strukturierten Befragungsdaten wissenschaftlich standardisiert ist, stellt sich die Frage, wie man aus digitalen Daten sinnvolle Messinstrumente ableiten kann. Es ist eine erhebliche Herausforderung, Konstrukte wie Populismus oder *Hate Speech* (dt.: Hassrede) mit computergestützten Verfahren in großen Datensätzen valide zu messen. Zudem bedarf es weitergehender Forschung dazu, wie individuelle Kosten-Nutzen-Strukturen, also etwa finanzielle Abhängigkeiten der Befragten von einer monetären Entlohnung für die Teilnahme, sowie allgemeine Einstellungen von Befragten die Bereitschaft zum Teilen von digitalen Verhaltensdaten beeinflussen und wie sich dies zwischen unterschiedlichen Formen digitaler Verhaltensdaten unterscheidet (Silber et al., 2022). Daran schließen sich Fragen soziodemografischer Verzerrungen an sowie weitere experimentelle Ansätze, die

Erhebung von digitalen Verhaltensdaten für Befragte möglichst einfach zu gestalten. Teilen beispielsweise Befragte von Online-Befragungen häufiger ihre digitalen Verhaltensdaten als Befragte von Telefoninterviews und sollte dies der Fall sein, liegt das an der höheren Affinität zu digitalen Inhalten oder an der Art der Anfrage digitale Daten zu teilen?

Fazit

Die Digitalisierung verändert die individuelle und gesellschaftliche Realität. Die Interaktion von Mensch und digitaler Technologie führt dabei zu ganz real erfahrbaren Phänomenen, wie beispielsweise Hassrede in sozialen Medien. Für die Sozialwissenschaften ergeben sich hierdurch zahlreiche neue Forschungsthemen und -fragen. Gleichzeitig eröffnen digitale Verhaltensdatendaten die Beforschung etablierter Forschungsfragen und theoretischer Fragen, beispielsweise zu menschlichen Interaktionen und ihren individuellen Folgen. Die Verbindung von Befragungen und digitalen Verhaltensdaten und ihrer jeweiligen Vorteile kann die sozialwissenschaftliche Forschung erheblich voranbringen. Gleichzeitig ist zu berücksichtigen, dass eine solche Datenverbindung selbst mit verschiedenen Herausforderungen einhergeht.

Durch die Verbindung von Umfragen und digitalen Verhaltensdaten sind an der Schnittstelle zwischen Sozialwissenschaften und Informatik neue Möglichkeiten für die Wissenschaft entstanden. Zukünftige Forschung wird nicht nur die Transparenz und die praktische Umsetzbarkeit des *Linking* beider Datentypen verbessern, sondern auch verschiedene Forschungsfelder mit innovativer inhaltlicher Forschung bereichern.

Literatur

Al Baghal, T., Sloan, L., Curtis, J., Williams, M. & Burnap, P. (2020). Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies. *Social Science Computer Review, 38*(5), 517–532. https://doi.org/10.1177/08944393198280

Beuthner, C., Keusch, F., Silber, H., Weiß, B. & Schröder, J. (2022). Consent to Data Linkage for Different Data Domains - The Role of Question Order, Question Wording, and Incentives. *SocArXiv,* 1–39. https://doi.org/10.31235/osf.io/qh93g

Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T. & Oberski, D. L. (2022). A framework for privacy preserving digital trace data collection through data donation. *Computational Communication Research, 4*(2), 388–423. https://doi.org/10.5117/CCR2022.2.002.BOES

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

Breuer, J., Kmetty, Z., Haim, M. & Stier, S. (2022). User-centric approaches for collecting Facebook data in the 'post-API age': Experiences from two studies and recommendations for future research. I*nformation, Communication & Society.* https://doi.org/10.1080/1369118X.2022.2097015

Christner, C., Urman, A., Adam, S. & Maier, M. (2022). Automated Tracking Approaches for Studying Online Media Use: A Critical Review and Recommendations. *Communication Methods and Measures, 16*(2), 79–95. https://doi.org/10.1080/19312458.2021.1907841

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

Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., Trappmann, M. & Eckman, S. (2022). Non-Participation in Smartphone Data Collection Using Research Apps. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 185*(S2), S225–S245. https://doi.org/10.1111/rssa.12827

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H. & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour, 5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Silber, H., Breuer, J., Beuthner, C., Gummer, T., Keusch, F., Siegers, P. & Stier, S. (2022). Linking surveys and digital trace data: Insights from two studies on determinants of data sharing behavior. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 185*(S2), S387–S407. https://doi.org/10.1111/rssa.12954

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

Trappmann, M., Haas, G.-C., Malich, S., Keusch, F., Bähr, S., Kreuter, F. & Schwarz, S. (2022). Augmenting survey data with digital trace data: Is there a threat to panel retention? *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smac023

Walzenbach, S., Burton, J., Couper, M. P., Crossley, T. F. & Jäckle, A. (2022). Experiments on multiple requests for consent to data linkage in surveys. *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smab053

Julia Weiß

GESIS – Leibniz-Institut für Sozialwissenschaften

E-Mail julia.weiss@gesis.org

Dr. Julia Weiß (GESIS, Mannheim) ist wissenschaftliche Mitarbeiterin bei der deutschen nationalen Wahlstudie (GLES). Ihre Forschungsinteressen liegen in den Bereichen Computational Social Science, politischer Partizipation und Arbeitsmarktforschung.

Sebastian Stier

GESIS – Leibniz-Institut für Sozialwissenschaften

E-Mail sebastian.stier@gesis.org

Dr. Sebastian Stier (GESIS, Köln) ist wissenschaftlicher Mitarbeiter und kommissarischer Abteilungsleiter der Abteilung Computational Social Science. Seine Forschungsinteressen liegen in den Bereichen Computational Social Science, digitale Medien, politische Kommunikation und vergleichende Politikwissenschaft.