

KonsortSWD - Research Data Grants - Abschlussbericht

für die kooperative Aufbereitung, Archivierung und Publikation der Forschungsmaterialien des Projekts „Paying Attention to Attention: Media Exposure and Opinion Formation in an Age of Information Overload“ (MEOF) im Rahmen der KonsortSWD Research Data Grants

Berichtsjahr 2022

Antragsteller: Prof. Dr. Simon Munzert, Hertie School (Berlin)

Beteiligte(s) FDZ: Abt. Data Services for the Social Sciences (DSS), Leitung: Dr. Holger, Döring und FDZ Wahlen, GESIS, Leitung Dr. Christina Eder.

Datensatzzitation

Titel der Studie:	<i>Media Exposure and Opinion Formation in an Age of Information Overload (MEOF)</i> https://simonmunzert.github.io/meof/
Projektleiterin/Datenhalterin:	<i>Prof. Dr. Simon Munzert</i>
Institution:	<i>Hertie School, Berlin</i>
Kurzcharakteristik:	<i>Panelbefragung “Politics and Media”, USA, Juli 2018 und März 2019; Panelbefragung “Politik und Medien”, Deutschland, Juli und Oktober 2017 und Oktober 2018 und April 2019; passive Trackingdaten für einen Großteil der Befragten in beiden Panels</i>
Förderungszeitraum:	<i>01.08.2021 bis 31.08.2022</i>
Aufbereitung in Kooperation mit FDZ:	GESIS https://www.gesis.org/datenservices/home; https://www.gesis.org/institut/forschungsdatenzentren/fdz-wahlen
Aspekte der FDM-Förderung (Call):	<i>Gefördert wurden durch KonsortSWD folgende Services: Archivierung PLUS von MEOF (d.h. Eingangsprüfung PLUS; Langezeitarchivierung von Daten und Dokumenten, standardisierte Beschreibung dt. u. engl., Datenbereitstellung (sofern möglich) zum Download); Datenangebot on-site im Secure Data Center von GESIS in Köln für sensitive Daten (Details sind über Links in Abs. 2 unten zugänglich)</i>

Zitation der drei derzeit zur Studie gehörenden Datensätze	<p><i>Munzert, Simon; Barberá, Pablo; Guess, Andrew M.; Yang, JungHwan (2022): Media Exposure and Opinion Formation in an Age of Information Overload (MEOF) – Survey Germany. GESIS Datenarchiv, Köln. ZA7894 Datenfile Version 1.0.0, doi:10.4232/1.13979</i></p> <p><i>Guess, Andrew M.; Barberá, Pablo; Yang, JungHwan; Munzert, Simon (2022): Media Exposure and Opinion Formation in an Age of Information Overload (MEOF) – Survey U.S.A.. GESIS Datenarchiv, Köln. ZA7895 Datenfile Version 1.0.0, doi:10.4232/1.13980</i></p> <p><i>Munzert, Simon; Barberá, Pablo; Guess, Andrew M.; Yang, JungHwan (2022): Media Exposure and Opinion Formation in an Age of Information Overload (MEOF) – Webtracking on-site. GESIS Datenarchiv, Köln. ZA7896 Datenfile Version 1.0.0, doi:10.4232/1.13981</i></p>
Weitere Informationen zu den Daten:	<p><i>Weitere Versionen der Webtracking-Daten mit niederschwelligeren Zugangsbedingungen befinden sich in Vorbereitung (siehe White Paper, 2022, Abschnitt “Application”).</i></p>

1) Allgemeine Angaben

MEOF-Projekt

Das Vorhaben „Paying Attention to Attention: Media Exposure and Opinion Formation in an Age of Information Overload“ (im Folgenden MEOF) wurde zwischen Januar 2017 und Juni 2022 durch Drittmittel der VolkswagenStiftung Computational Social Science Initiative, Referenz 92 143 gefördert (Leitung: Simon Munzert, Hertie School; Andrew Guess, Princeton University; Pablo Barberá, University of Southern California; JungHwan Yang, University of Illinois at Urbana-Champaign). Die Daten wurden in überlappenden Zeiträumen in den USA und der Bundesrepublik Deutschland erhoben und sollten die Auswirkungen der Nutzung von Online-Medien auf die Meinungsbildung und andere Einstellungs- und Verhaltensergebnisse untersuchen (Munzert et al., 2017).

Ziel des hier beschriebenen Projektes von war es, die im Rahmen des MEOF-Projekts gesammelten Daten für die wissenschaftliche Nachnutzung bereitzustellen. Der zu bearbeitende Datenkorpus besteht aus folgenden Teilen:

- (1) einer Panelbefragung unter dem Titel “Politics and Media”, die in den USA zwischen Juli 2018 und März 2019 in insgesamt 8 Befragungswellen unter anfänglich 1551 Befragten durchgeführt wurde (Feldinstitut: YouGov USA),
- (2) einer Panelbefragung unter dem Titel “Politik und Medien”, die in Deutschland zwischen Juli und Oktober 2017 und mit einer Teilauffrischung des Samples nochmals zwischen Oktober 2018 und April 2019 (und einer Nacherhebung im Juni 2020) in insgesamt 9 Wellen unter anfänglich 1503 Befragten durchgeführt wurde (Feldinstitut: YouGov Deutschland), und
- (3) passiven Trackingdaten für einen Großteil der Befragten in beiden Panels, die anfänglich über die Tracking-Software Wakoopa und später über RealityMine die Browserhistorie auf allen registrierten Geräten (Desktop und mobil) umfasst.

Von den Befragten wurden informierte Einwilligungen entsprechend der zum Zeitpunkt der jeweiligen Befragung geltenden rechtlichen Regelungen eingeholt. Neben der Aufbereitung, Dokumentation und Bereitstellung der eigentlichen Daten in mehreren, sich aus Datenschutzgründen inhaltlich unterscheidenden Varianten der Tracking-Daten, diente das Vorhaben auch der Erarbeitung eines Konzeptes für die Weitergabe der Trackingdaten an Dritte (Munzert et al. 2022).

Zum Abschluss des Projektes wurde neben den Survey-Daten eine erste Version der Tracking-Daten für die Nutzung in den Räumlichkeiten von GESIS (Secure Data Center, on-site) sowie ein White Paper (Munzert et al. 2022) mit Details zu Erhebung, Aufbereitung und Veröffentlichung der Daten publiziert.

GESIS (DSS, FDZ Wahlen)

GESIS erbringt grundlegende, überregional und international bedeutsame forschungsbasierte Dienstleistungen für die Sozialwissenschaften. Als die größte europäische Infrastruktureinrichtung für die Sozialwissenschaften steht das Institut Wissenschaftlerinnen und Wissenschaftlern auf allen Ebenen ihres Forschungsprojekts mit Expertise und Dienstleistungen beratend zur Seite. Mit dieser Unterstützung lassen sich gesellschaftlich relevante Fragen auf der Basis neuester wissenschaftlicher Methoden, qualitativ hochwertiger Daten und Forschungsinformationen beantworten. GESIS bietet u. a. Services zur Sicherung, Archivierung, Aufbereitung, Dokumentation und Veröffentlichung von sozial- und wirtschaftswissenschaftlichen Forschungsdaten an.

2) Arbeits- und Ergebnisbericht

MEOF-Projekt

Das MEOF-Projekt zielte auf die Untersuchung der Folgen der Nutzung digitaler Medien für politisches Verhalten. Im Mittelpunkt des Projekts standen drei Fragen und damit verbundene methodische Herausforderungen. Erstens: Wer ist welcher Information ausgesetzt? Dieses

wichtige Messproblem ist durch die zunehmend fragmentierte Online-Landschaft noch größer geworden. Zweitens: Welche Faktoren bestimmen Medienkonsumverhalten? Die meisten Ansätze zur Beantwortung dieser Frage leiden unter Problemen umgekehrter Kausalität, die empirisch schwer zu lösen sind. Schließlich drittens: Wie verhalten sich unterschiedliche Menschen in diesem Umfeld? Trotz der massiven Zunahme von individuellen Verhaltensdaten z.B. in Social-Media-Kanälen besteht bisher die Tendenz zur Aggregatdatenanalyse, um Rückschlüsse auf Individualverhalten zu ziehen.

In Zusammenarbeit mit dem Online-Umfrageinstitut YouGov wurden insgesamt 1.551 Befragte aus den USA und 1.516 Befragte aus den deutschen Pulse-Panels rekrutiert. Die Befragten erklärten sich bereit, an einer Studie zum Thema "Politik und Medien" (in der deutschen Stichprobe als "Politik & Medien" bezeichnet) mit mehreren Befragungswellen teilzunehmen. Die Teilnahme wurde durch ein YouGov-eigenes Punktesystem belohnt und beinhaltete einen Bonus für die Teilnahme an allen Wellen, um einen Anreiz zum Verbleib in der Studie zu schaffen. Pro durchschnittlich etwa 10- bis 20-minütiger Befragung (je nach Welle) wurden die Befragten mit umgerechnet etwa 2-5 Euro entschädigt, was dem Standard der Vergütung zum damaligen Zeitpunkt auf dieser Plattform entsprach. Für die Bereitstellung der Trackingdaten wurden Teilnehmende laut Provider mit umgerechnet etwa 10-15 Euro pro Monat vergütet. Die Teilnahme war freiwillig und die Befragten konnten jederzeit aus dem passiven Messteil der Studie aussteigen.

Die erste Welle der deutschen Umfrage wurde am 13. Juli 2017 gestartet, etwa zehn Wochen vor der Bundestagswahl 2017 am 24. September 2017. Drei weitere Wellen wurden vor der Wahl durchgeführt, die fünfte Welle nach der Wahl im Oktober 2017. Nach einer längeren Pause wurde das Panel mit der sechsten Welle am 29. Oktober 2018 wieder aufgenommen und mit 1.009 neuen Teilnehmern aufgefrischt. Die Tracking-Daten deckten die Zeiträume von Juli bis Dezember 2017 und Oktober 2018 bis Februar 2018 ab und umfassten insgesamt 123.572.973 getrackte URLs.

In den USA bestehen die Daten aus 8 Erhebungswellen. Die erste Welle wurde am 23. April 2018 gestartet. Insgesamt fanden vier Wellen vor der US-Midterm-Wahl 2018 und vier danach statt. Die sechste Welle wurde am 24. Januar 2019 durchgeführt. Für das US-Panel wurde der Webtracking-Client von RealityMine eingesetzt. Die Tracking-Daten umfassen den Zeitraum von August 2018 bis März 2019 mit insgesamt 77.709.911 Logs.

Obwohl es bereits Beispiele für die Publikation von Webtracking-Daten gibt, mussten für das vorliegende Projekt diverse ethische, rechtliche, sowie technische und organisatorische Herausforderungen in den Blick genommen werden. So sind die Tracking-Daten durch ihren unmittelbaren Erhebungskontext denjenigen Befragten direkt zuordenbar, die einer Erhebung zugestimmt hatten. Die Tracking-Daten, deren unbearbeiteter Kern vollständige Uniform Resource Locator (URLs) sind, umfassen Hinweise auf sensitive Inhalte, wie beispielsweise gesundheitsbezogene Suchen, detaillierte Standortinformationen über Kartendienste, sowie direkte und indirekte Hinweise auf die Betroffenen (z.B. Namen und Emailadressen).

Das im Rahmen dieses Projekts erarbeitete Framework für die Publikation verlinkter Survey- und Tracking-Daten basiert auf existierenden Ansätzen (z.B. Beuthner et al. 2021) sowie eigener Forschungsarbeit im Projekt. Die wesentlichen Bausteine des Vorgehens sind die zum einen die Reduzierung der URLs auf verschiedene Niveaus und zum anderen die Beschränkung des Zugangs zu den verlinkten Daten. Wurde im Zuge des Projektes zunächst nur eine schwache Anonymisierung realisiert, die lediglich on-site im Secure Data Center genutzt werden kann, sind weitere Varianten des Webtracking-Daten in Vorbereitung. So soll es folgende Varianten geben:

1. einen Public Use File mit aggregierten Informationen,
2. einen „Academic“ Use File mit Daten auf individueller Ebene, gefiltert auf Domänenebene, die mit den Survey-Daten verknüpft werden können, und
3. einen „Academic“ Use File mit vorverarbeiteten vollständigen URLs auf individueller Ebene, die auf eine vordefinierte Gruppe von Nachrichten-Websites beschränkt sind.

Die Ausarbeitung dieser Niveaus sowie deren Umsetzung stellen die größten Herausforderungen des Projektes dar. Alle Details sowie Best-Practice-Empfehlungen finden sich im White Paper (Munzert et al. 2022).

GESIS

Die Publikation der oben genannten Daten erfolgte im Rahmen des Servicepakets Archivierung PLUS (<https://www.gesis.org/datenservices/daten-teilen/archivierung-plus>), welches durch wissenschaftliche Beratungsleistungen für die Unterstützung der Erstellung eines Anonymisierungskonzeptes für die Browser-Trackingdaten ergänzt wurde.

Das Angebot der ersten Version der Webtracking-Daten erfolgt on-site über das Secure Data Center bei GESIS (<https://www.gesis.org/angebot/daten-aufbereiten-und-analysieren/forschungsaufenthalte/secure-data-center-sdc>).

Referenzen im Text

Beuthner et al. (2021): C. Beuthner, J. Breuer, J. and S. Jünger, Data Linking — Linking survey data with geospatial, social media, and sensor data. GESIS Survey Guidelines. (https://doi.org/10.15465/GESIS-SG_EN_039; access date: 30.08.2022).

Munzert et al. (2017): S. Munzert, P. Barberá, A. Guess and J. Yang, Media Exposure and Opinion Formation in an Age of Information Overload (<https://simonmunzert.github.io/meof/material/design-paper-apsa17.pdf>; access date: 30.08.2022).

3) Zusammenfassung und Nachnutzungspotenzial

Das vorliegende von KonsortSWD geförderte Projekt dient der Archivierung und Publikation der Daten des Projektes „Paying Attention to Attention: Media Exposure and Opinion Formation in an Age of Information Overload“ (im Folgenden MEOF), vertreten durch Professor Dr. Simon Munzert. MEOF wurde zwischen Januar 2017 und Juni 2022 durch Drittmittel der VolkswagenStiftung gefördert und zielte auf die Untersuchung der Folgen der Nutzung digitaler Medien für politisches Verhalten mit Blick die Art der Information, die das Medienkonsumverhalten bestimmenden Faktoren, sowie die Frage, wie sich unterschiedliche Menschen in diesem Umfeld verhalten.

Archiviert und publiziert wurden im von KonsortSWD geförderten Vorhaben sowohl die Panelbefragungen als auch die Trackingdaten. Aus Datenschutzgründen wurde zunächst eine erste, schwach anonymisierte variante veröffentlicht, die on-site im Secure Data Center von GESIS genutzt werden kann. Weitere Varianten der Trackingdaten befinden sich in Vorbereitung. Da es bisher nur wenige Vorhaben gibt, die derartige Daten zur Nachnutzung anbieten, wurde das Vorhaben auch zur Erarbeitung eines Konzeptes für die Weitergabe der Trackingdaten an Dritte genutzt. Dieses geht ausführlich auf die Herausforderungen ein (Munzert et al. 2022).

Nachnutzungspotentiale von sozialwissenschaftlichen Forschungsdaten lassen sich nur schwer vorhersagen. Ausschlaggebende Faktoren sind nach Ansicht von GESIS a) die Vielfalt der wissenschaftlichen Fragestellungen, die sich mit Hilfe der Daten beantworten lassen, b) die Größe des oder der Samples von Individuen, die für die Datenerhebung genutzt wurden, und z.B. regionale Vergleiche ermöglichen, sowie c) die Güte der Daten, die statistisch valide Analysen zulassen. Wir gehen daher davon aus, dass die verschiedenen Daten von MEOF unterschiedliche Nachnutzungsniveau erfahren werden. Während der internationale Vergleich zwischen den USA und Deutschland auch für Forscher*innen interessant sein dürfte, die keinen Schwerpunkt auf der Kommunikationsforschung haben, könnten die kombinierten Survey- und Webtracking-Daten aufgrund der hohen Erhebungskosten für derartige Vorhaben für eine Vielzahl von Nutzer*innen von Bedeutung sein. Interessant wird hier die Veröffentlichung stärker anonymisierter Trackingdaten, die mit den Survey-Daten kombiniert werden können, und weniger strikten Zugangsbedingungen unterliegen.

4) Publikationen

Munzert et al. (2022): S. Munzert, S. Ramirez-Ruiz, O. Watteler, J. Breuer, V. Batzdorfer, Ch. Eder, D. Wiltshire, A. Guess, P. Barberá, J. Yang, Publishing Combined Web Tracking and Survey Data, Berlin / Köln (URL tba.).

Datenpublikation

Munzert, Simon; Barberá, Pablo; Guess, Andrew M.; Yang, JungHwan (2022): Media Exposure and Opinion Formation in an Age of Information Overload (MEOF) – Survey Germany. GESIS Datenarchiv, Köln. ZA7894 Datenfile Version 1.0.0, doi:10.4232/1.13979

Guess, Andrew M.; Barberá, Pablo; Yang, JungHwan; Munzert, Simon (2022): Media Exposure and Opinion Formation in an Age of Information Overload (MEOF) – Survey U.S.A.. GESIS Datenarchiv, Köln. ZA7895 Datenfile Version 1.0.0, doi:10.4232/1.13980

Munzert, Simon; Barberá, Pablo; Guess, Andrew M.; Yang, JungHwan (2022): Media Exposure and Opinion Formation in an Age of Information Overload (MEOF) – Webtracking on-site. GESIS Datenarchiv, Köln. ZA7896 Datenfile Version 1.0.0, doi:10.4232/1.13981