

ETHISCHE FRAGEN IN DER FORSCHUNG MIT DIGITALEN SPURDATEN

Dr. Johannes Breuer

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

gesis

Leibniz-Institut
für Sozialwissenschaften



RESEARCH
FOR THE
DIGITAL AGE



Programm

1. Was sind digitale Spurdaten?
2. Forschungsethische Fragen zur Nutzung digitaler Spurdaten
 1. Einleitung
 2. Datenerhebung
 3. Informed Consent
 4. Datenaufbereitung & -analyse
 5. Veröffentlichung und Archivierung
 6. Hilfestellungen & Ressourcen
3. Rückblick & Abschluss

WAS SIND DIGITALE SPURDATEN?

Definition digitaler Spurdaten

“digital trace data can be roughly defined as ‘**records of activity** (trace data) undertaken through an **online information system** (thus, digital)’ (Howison, Wiggins, & Crowston, 2011) and can be collected from a multitude of technical systems, such as **websites, social media platforms, smartphone apps, or sensors**” (Stier et al., 2020)

“**large digital datasets that describe human behavior** (e.g., social media sites, internet search data, blogs, administrative records, historical archives, audio-visual, or geospatial data)”
([Bail, 2020](#))

Entstehung der Daten

- Found vs. designed data
- Custom-made vs. ready-made data ([Bail, 2020](#))
- Unterschiede in den datengenerierenden Prozessen

- Digitale Spurdaten sind i.d.R. found bzw. ready-made data
 - nicht (originär) für die Forschung produziert
 - Entstehung der Daten und wissenschaftliche Nutzung sind voneinander getrennt

- Designed bzw. Custom-made data
 - genuin für die Forschung produziert (→ Nutzungszweck vor Generierung der Daten klar)
 - z.B. Befragungen, Interviews, Experimente
 - Digitale Spurdaten können auch „designed“ sein: z.B. Online-Experimente oder Interventionsstudien

Quellen digitaler Spurdaten



Arten von digitalen Verhaltensdaten

- intentional (z.B. Posts/Kommentare in sozialen Medien) vs. unintentional (z.B. Location oder Log Data von Smartphones) traces (Hox, 2017)
- participation traces (z.B. Kommentare/Posts) vs. transactional data (z.B. Logins) (Menchen-Trevino, 2013)
- horizontal trace data (z.B. alle Tweets zu einem bestimmten Hashtag) vs. vertical trace data (umfassende Daten zur Nutzung für begrenzte Gruppe von Nutzer*innen) (Menchen-Trevino, 2013)

Herausforderungen in der Forschung mit digitalen Spurdaten

- Praktisch
 - Datenzugang
 - Datenumfang & -format
 - Datenaufbereitung & -analyse (Bsp. Textdaten)
- Ethisch
 - Was ist öffentlich?
 - Persönliche & möglicherweise sensible Daten
 - Informed Consent
- Rechtlich
 - Datenschutz
 - Copyright
 - Terms of Service (ToS) von Plattformen

➤ Implikationen für Open Science (v.a. Data Sharing)

FORSCHUNGSETHISCHE FRAGEN

Forschungsethik

„Unter dem Stichwort ‚Forschungsethik‘ werden in den Sozialwissenschaften im Allgemeinen all jene ethischen Prinzipien und Regeln zusammengefasst, in denen mehr oder minder verbindlich und mehr oder minder konsensuell bestimmt wird, in welcher Weise die Beziehungen zwischen den Forschenden auf der einen Seite und den in sozialwissenschaftliche Untersuchungen einbezogenen Personen auf der anderen Seite zu gestalten sind.“

(Hopf, 2016, S. 195)

Zentrale Forschungsethische Prinzipien

- **Wissenschaftliche Güte und Integrität der Forschenden**
 - Allgemeine wissenschaftliche Standards
 - Integrität der Forschenden
 - Dokumentation, Archivierung, Sekundärnutzung
- **Vermeiden von Schaden**
 - Schutz der Studienteilnehmenden
 - Schutz der Forschenden
- **Informierte Einwilligung**
 - Information
 - Freiwilligkeit der Teilnahme
 - Entscheidungskompetenz und Einwilligungsfähigkeit
 - Form der Einwilligung

Quelle: RatSWD (2019). Forschungsethik in den Sozial- und Wirtschaftswissenschaften.
<https://www.konsortswd.de/aktuelles/publikation/forschungsethik-in-den-sozial-und-wirtschaftswissenschaften/>

Forschungsethik

- Ziele: Schutz von Rechten, Maximierung des Nutzens, Minimierung von Schäden/Gefahren/Risiken
 - “respect for persons, justice, and beneficence”
- Trotz alltagssprachlichem Gebrauch (ethisch vs. unethisch) keine Binarität
- Häufig keine eindeutigen bzw. allgemeingültigen Antworten
 - abhängig von Gewichtung verschiedener Werte und Ziele, die u.U. miteinander in Konflikt stehen können
 - unterschiedliche Perspektiven auf Forschungsethik in verschiedenen Disziplinen
 - z.B. Informatik vs. Medizin vs. Sozialwissenschaften: Wie gängig ist Human Subjects Research und was bedeutet dies?
- Forschungsethische Fragen ergeben sich für alle Phasen des Forschungsdatenzklus (Erhebung, Aufbereitung, Analyse, Veröffentlichung)
- Wichtig und hilfreich, früh im Forschungsprozess (zu Beginn der Studienplanung oder Projektkonzeption) forschungsethische Fragen zu adressieren

Recht & Forschungsethik

- Rechtliche Aspekte behandeln wir in diesem Workshop nicht
- Relevante Rechtsbereiche: v.a. Datenschutzrecht, Urheberrecht, Vertragsrecht
 - Für Datenschutzrecht in der EU maßgeblich: Datenschutzgrundverordnung (DSGVO; siehe z.B. <https://gdpr.eu/>) + Bundesdatenschutzgesetz für Deutschland (siehe: https://www.gesetze-im-internet.de/bdsg_2018/index.html)
- Ressourcen:
 - Bayer, S., Breuer, J., Lösch, T., & Göbel, J. W. (2021). *Nutzung von Social-Media-Daten in der Bildungsforschung*. Forschungsdaten Bildung informiert 9. https://www.forschungsdaten-bildung.de/files/fdb-informiert-nr-9_version1.1.pdf
 - Social Media Research Assessment Template for Ethical Scholarship – Section 4 on Legal Requirements: <https://leibniz-hbi.github.io/socrates/#4-legal-requirements>
 - 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>
 - 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.
 - Podcast „Datenliebe“ mit Folgen zu Datenschutz & Urheberrecht: <https://open.spotify.com/show/2oLLWXTRoR9zv9kEtYi5cx>

Recht & Forschungsethik

- Rechtliche Vorgaben regeln, was man (nicht) machen darf, ethische Richtlinien zeigen auf, was man (nicht) machen sollte
- Rechtliche Anforderungen als notwendige, ethische als hinreichende Bedingungen
 - „Ethik beginnt dort, wo die Verkehrszeichen enden“
(<https://taz.de/Solidaritaet-in-der-Pandemie/!5815713/>)
- Dennoch viele Bereiche/Aspekte, die sowohl rechtlich als auch ethisch relevant sind (z.B. Datenschutz oder Informed Consent)

Relevanz forschungsethischer Fragen

- Neue & komplexere Daten & Studiendesigns
 - Digitale Spurdaten (kombiniert mit anderen Daten)
 - Experimental-/Interventionsstudien online (ggf. auch längsschnittliche Designs)
- Entwicklungen im Bereich Open Science (z.B. Data Sharing)
- Anforderungen an Forschende: z.B. bei Drittmittelanträgen oder Publikationen

Diskussion – Ethische Fragen: Eigene Erfahrungen

- Welche ethischen Fragen haben sich in Ihrer Forschung mit digitalen Spurdaten ergeben? Wie sind Sie diese angegangen?
- Falls Sie selbst noch nicht mit digitalen Verhaltensdaten geforscht haben: Welche ethischen Aspekte/Fragen/Herausforderungen sind Ihnen beim Lesen von Publikationen aufgefallen, in denen digitale Spurdaten genutzt wurden?

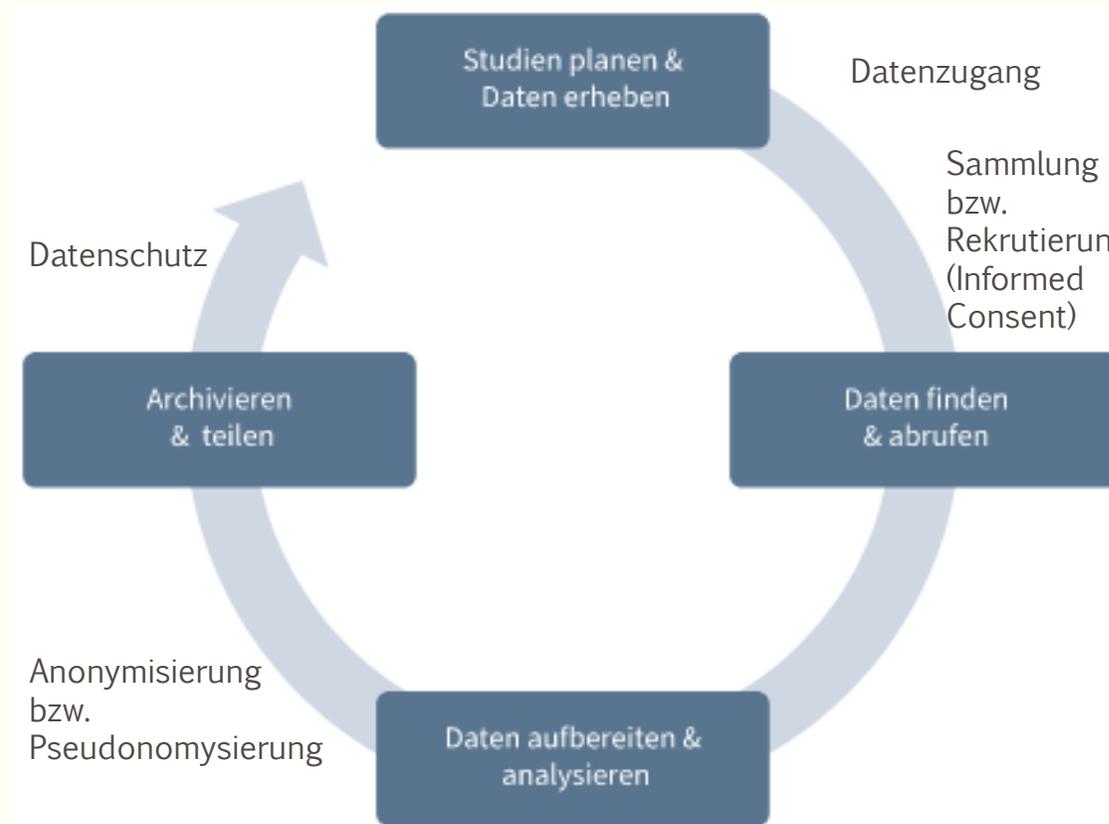
Diskussion – Cambridge Analytica

- Was wissen Sie über den Cambridge-Analytica-Fall?
- Welche (forschungs-)ethischen Dimensionen und Aspekte sehen Sie durch diesen Fall berührt?
- An welchen Stellen bzw. für und durch wen ergeben sich in diesem Fall (forschungs-)ethische Probleme?

How it started: Grassegger, H., & Krogerus M. (2018, March 20). „Ich habe nur gezeigt, dass es die Bombe gibt“. Psychologe Michal Kosinski weiß, wie man Menschen anhand ihres Facebook-Verhaltens analysiert. *Tagesanzeiger*.
<https://www.tagesanzeiger.ch/ausland/europa/diese-firma-weiss-was-sie-denken/story/17474918>

How it ended (?): Connor, P. (Host) (2021, April 04). Facebook-Gate! Inside the 2018 Cambridge Analytica Scandal (with Alex Kogan) (Season 2, Episode 07) [Audio podcast episode]. In More of a Comment Than a Question.
<https://moreofacomment.buzzsprout.com/1207223/8271165-facebook-gate-inside-the-2018-cambridge-analytica-scandal-with-alex-kogan>

Herausforderungen in der Forschung mit digitalen Spurdaten



Quelle der Abbildung des Forschungsdatenzyklus: <https://www.gesis.org/angebot>

Datenerhebung

- Verschiedene Möglichkeiten des Zugangs zu digitalen Verhaltensdaten (siehe Breuer et al., 2020):
 1. Daten selbst erheben
 - a) Application Programming Interface (API)
 - b) Web Scraping
 2. Direkte Kooperationen mit Anbietern (z.B. als „embedded researcher“)
 3. Erwerb von Drittanbietern (Data Reseller oder Marktforschungsfirmen)
 4. Kooperation mit Nutzer*innen (siehe Halavais, 2019), z.B. über Datenspende (siehe Boeschoten et al., 2020)
 5. Nutzung bereits erhobener und veröffentlichter Daten (z.B. [TweetsKB](#), [Geotagged Tweets from the US](#), [TweetSets](#), [Documenting the Now](#))

- Zugangswege unterscheiden sich auf verschiedenen Dimensionen, (z.B. benötigte Ressourcen, Art & Umfang der Daten), haben dadurch jeweils eigene Vor- und Nachteile und bringen auch spezifische ethische Fragen mit sich

Datenerhebung

- Verhältnis zu den Beforschten
 - direkter Kontakt (möglich)? → wichtig z.B. für die Einholung von Informed Consent
 - Bsp.: Bei Datenerhebung über APIs kann man nicht wirklich von „Teilnehmer*innen“ sprechen
- Definition/Auswahl der Stichprobe
 - Minderjährige oder vulnerable Gruppen?
 - Personen des öffentlichen Lebens (und/oder Institutionen)?
- Was steckt in den Daten?
 - z.B. Texte, Bilder
 - Daten Dritter (z.B. Social-Media-Kontakte)
 - sensible Themen
 - Struktur und Inhalt der Daten beeinflussen bspw. Anforderungen im Hinblick auf Anonymisierung/Pseudonymisierung
- Prinzip der Data Minimization: Möglichst nur erheben, was man tatsächlich benötigt

Datenerhebung

- Sind die Daten „öffentlich“?
 - Was bedeutet „öffentlich“ hier bzw. was wird darunter verstanden?
 - keine binären Kategorien privat vs. öffentlich für digitale Spurdaten möglich
- Selbst wenn Inhalte (z.B. Nutzer*innenprofile oder Posts) rechtlich und/oder technisch gesehen öffentlich sind, deckt sich dies nicht zwangsläufig mit der Perspektive der Nutzenden
 - Bsp.: Forschende i.d.R. nicht Teil der „imagined audience“ von Twitter-Nutzer*innen (Marwick & boyd, 2011)
- zudem: Nutzung von Daten durch Dritte zumeist durch Terms of Service (ToS) von Plattformen abgedeckt, jedoch ist Nutzenden dies oft nicht bewusst bzw. unklar, was dies bedeutet
- ”Concerns over consent, privacy and anonymity do not disappear simply because subjects participate in online social networks; rather, they become even more important” (Zimmer, 2010, S. 324)

Informed Consent

- Informierte Einwilligung sowohl aus rechtlicher als auch aus ethischer Perspektive relevant
- Informierte Einwilligung ist nach DSGVO eine der möglichen rechtlichen Grundlagen für die Erhebung und Verarbeitung personenbezogener Daten
 - eine andere ist z.B. die Durchführung einer Aufgabe im öffentlichen Interesse (muss klar begründbar sein); rechtliche Grundlage muss dokumentiert werden
 - Informed Consent auch bei Vorliegen anderer rechtlicher Grundlage aus ethischen Gründen sinnvoll
- Informierte Einwilligung v.a. bei der Erhebung personenbezogener Daten wichtig; sinnvoll und empfehlenswert jedoch auch wenn keine personenbezogenen Daten erhoben werden

Informed Consent

- Allgemeine Anforderungen:
 - eindeutige Einverständniserklärung für die Erhebung, Verarbeitung und Nutzung der (personenbezogenen) Daten
 - Einwilligung muss auf freier Willensentscheidung beruhen
 - Einwilligung kann verweigert und jederzeit widerrufen werden
 - Einwilligung muss dokumentiert werden
- Teilnehmende sollten u.a. informiert werden über Ziele & Methoden der Studie, ihre Rechte, mögliche Risiken, Nutzung der Daten (z.B. durch wen?), Maßnahmen zum Datenschutz
- Informationen im Informed Consent sollten Teilnehmende verständlich sein
- idealiter sollte Informed Consent vor der Datenerhebung eingeholt werden

Informed Consent & digitale Spurdaten

- Möglichkeit, Informed Consent einzuholen, abhängig von der Art der Erhebung digitaler Spurdaten
 - z.B. schwierig bis unmöglich bei umfangreicher Datensammlung über APIs
- Kombination mit Befragung/Interview erleichtert auch das Einholen von Informed Consent
 - insbesondere wenn Befragung/Interview zuerst durchgeführt werden
- Einholung von Informed Consent auch über Modell der Data Donation möglich
 - dies kann zudem Transparenz erhöhen und aktive(re) Einbindung der Teilnehmenden erlauben (z.B. über Möglichkeit, eigene Daten zu explorieren und zu entscheiden, was davon geteilt werden soll)

Informed Consent

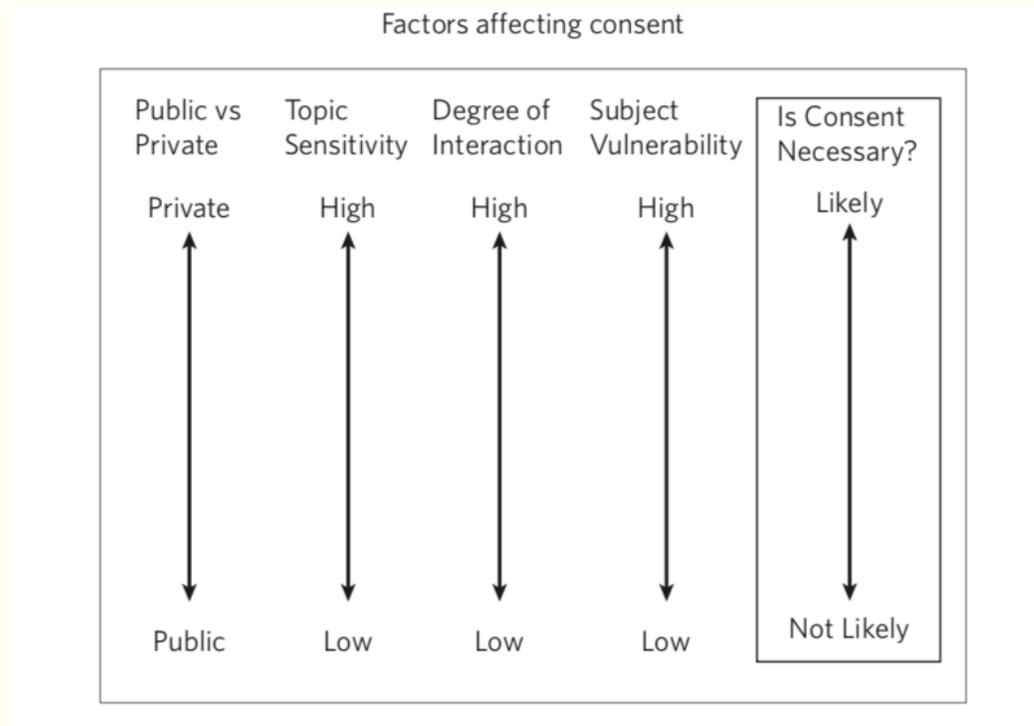


Abbildung aus McKee & Porter, 2009, S. 88

Formulierung von Informed Consent

- Spezifisch für die Erhebung von Social-Media-Daten (in Kombination mit Befragungsdaten): Breuer et al., 2021, Appendices; Sloan et al., 2020, Appendix
- Informed Consent allgemein (insb. wenn Daten geteilt/veröffentlicht werden sollen):
 - Verbund Forschungsdaten Bildung (2019): Checkliste zur Erstellung rechtskonformer Einwilligungserklärungen mit besonderer Berücksichtigung von Erhebungen an Schulen. Version 2.0, fdbinfo Nr. 1. https://www.forschungsdaten-bildung.de/files/fdbinfo_1.pdf
 - Verbund Forschungsdaten Bildung (2018): Formulierungsbeispiele für „informierte Einwilligungen“. Version 2.1. fdbinfo Nr. 4. https://www.forschungsdaten-bildung.de/files/fdbinfo_4.pdf
 - UK Data Service „Consent for Sharing“: <https://ukdataservice.ac.uk/learning-hub/research-data-management/ethical-issues/consent-for-data-sharing/>
 - Utrecht University „Informed consent for data sharing“: <https://www.uu.nl/en/research/research-data-management/guides/informed-consent-for-data-sharing>
 - Abschnitt zu Informed Consent im CESSDA Data Management Expert Guide: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Informed-consent>

Datenaufbereitung & -analyse

- Anonymisierung: Keine Identifikation von Individuen mehr möglich (z.B. durch Aggregation der Daten)
- Pseudonymisierung: Entfernung direkter Identifier (z.B. User Names) & Bearbeitung der Daten (z.B. durch Ausschluss bestimmter Variablen oder Top- und Bottom-Coding für extreme Werte/Outlier), so dass eine Identifikation von Individuen nicht ohne erheblichen zusätzlichen Aufwand möglich ist (z.B. nur durch Verknüpfung mit anderen/externen Daten)
- Anonymisierung und Pseudonymisierung von digitalen Spurdaten schwieriger als z.B. bei Befragungsdaten
 - Bsp. Textdaten aus sozialen Medien → möglicherweise auffindbar über Suchmaschinen
 - Bsp. Webtrackingdaten: Besuch eigener Profile (bei Full URLs) oder der eigenen Webseite (auch bereits Domain-Ebene), URL-Parameter
 - Bsp. Twitter-Daten: zahlreiche (Meta-)Daten pro Tweet/User, die auch ohne den Tweet-Text (indirekte) Identifier sein können (siehe dazu Sloan et al., 2020)

Datenaufbereitung & -analyse

- Direct identifier so bald wie möglich aus den Daten entfernen (z.B. User Names oder User IDs)

- Weitere Möglichkeiten:
 - Rohdaten nicht mehr nutzen und unter besonderen Schutzmaßnahmen (Zugangsbeschränkungen, Verschlüsselung) speichern, sobald derived variables für die Analyse erstellt sind (z.B. word counts oder sentiment scores für Textdaten)
 - Data reduction: Nicht benötigte Variablen löschen bzw. nicht ins „working data set“ aufnehmen (v.a. wenn diese identifier sind/sein können)
 - Daten getrennt halten (z.B. digitale Spurdaten und Befragungsdaten) und nur verknüpfen, was für Analysen unmittelbar benötigt wird
 - mit Lookup Tables für die Verknüpfung arbeiten

Datenaufbereitung & -analyse

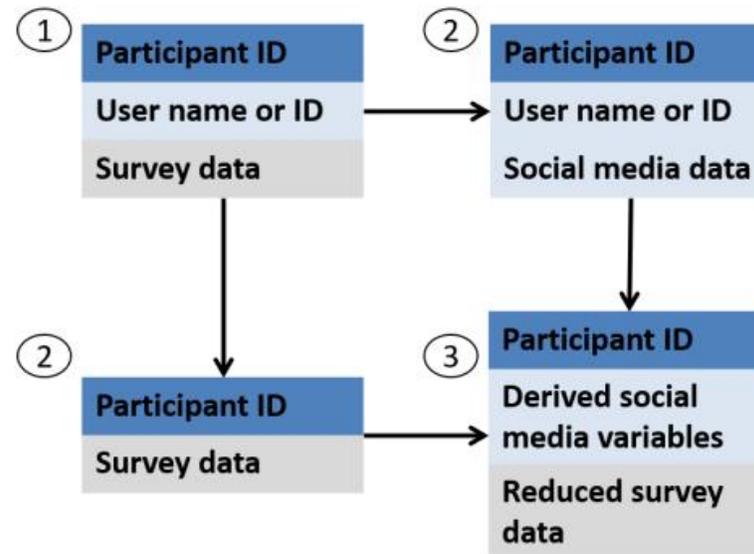


Figure 3: Proposed workflow for linking survey and social media data based on Sloan et al. (2020)

Abbildung aus Beuthner et al., 2021, S. 7

Veröffentlichung & Archivierung

- “Publishing will refer to the presentation of data extracts, aggregations, or summaries, while sharing refers to the practice of making the underlying data available postpublication for others to use” (Bishop & Gray, 2017)
- Daten teilen, um Transparenz, Reproduzierbarkeit und Replizierbarkeit zu erhöhen und Nachnutzung (z.B. für weitergehende oder Meta-Analysen) zu ermöglichen
- Forschungsdaten sollten FAIR sein (Findable, Accessible, Interoperable, Reusable)
 - Datenformat(e), Dokumentation, Speicherort & Zugriff (dauerhafte und z.T. auch aktiv kuratierte Archivierung über Anbieter wie GESIS: siehe <https://www.gesis.org/datenservices/daten-teilen> & <https://data.gesis.org/sharing/#!Home>)
 - siehe <https://www.go-fair.org/fair-principles/>
- Erweiterung um explizite forschungsethische Dimensionen: CARE (Collective Benefit, Authority to Control, Responsibility, Ethics; siehe <https://www.gida-global.org/care>)

Veröffentlichung & Archivierung

- Risiko der Gefährdung von Teilnehmenden durch Veröffentlichung von Daten (insb. Rohdaten)
 - Balance zwischen Offenheit/Transparenz und Nachnutzungswert auf der einen und Datenschutz bzw. Schutz der Teilnehmenden auf der anderen Seite muss gefunden werden
 - abhängig von Art der Daten, Forschungsfrage/-interesse, betrachteter Population...
- “As open as possible, as closed as necessary”
([Guidelines on FAIR Data Management in Horizon 2020](#))

Veröffentlichung & Archivierung

- Digitale Spurdaten enthalten i.d.R. personenbezogene Daten
 - u.U. auch von Dritten (z.B. Social-Media-Kontakte)
 - häufig auch sog. „besondere Kategorien“ personenbezogener Daten (gemäß DSGVO), wie z.B. Informationen über politische Meinungen, religiöse Überzeugungen oder Gesundheit
- Digitale Spurdaten sind oft besonders sensibel (z.B. Location Data, politische Äußerungen etc.)
- Besonders in Kombination mit anderen Daten hohes Identification Risk (auch wenn direkte Identifier entfernt werden)
- Terms of Services (ToS) von digitalen Plattformen beinhalten zumeist Vorgaben dazu, ob bzw. wie/in welcher Form die Daten archiviert & veröffentlicht werden können

Veröffentlichung & Archivierung

- Mögliche Lösungsansätze für Archivierung und Veröffentlichung digitaler Spurdaten
 - Nur aufbereitete (und reduzierte) „replication data“ (King, 1995) teilen; ggf. auch nur Metadaten frei zugänglich veröffentlichen
 - Restricted secure access für vollständige Daten
 - Synthetic data oder differential privacy approaches
 - Neue technische und organisationale Lösungen für den Datenzugang (van Atteveldt, Althaus et al., 2020; van Atteveldt, Strycharz et al., 2019)
 - z.B. non-consumptive use/remote execution, secure remote access

Beispiel: Twitter Developer Policy

“We limit the circumstances under which you may match a person on Twitter to information obtained or stored off-Twitter. Off-Twitter matching involves associating Twitter Content, including a Twitter @handle or user ID, with a person, household, device, browser, or other off-Twitter identifier. **You may only do this if you have express opt-in consent from the person** before making the association, or as described below. In situations in which you don't have a person's express, opt-in consent to link their Twitter identity to an off-Twitter identifier, we require that any connection you draw be based only on information that someone would reasonably expect to be used for that purpose.”

“If you provide Twitter Content to third parties, including downloadable datasets or via an API, **you may only distribute Tweet IDs, Direct Message IDs, and/or User IDs (...). Academic researchers are permitted to distribute an unlimited number of Tweet IDs and/or User IDs** if they are doing so on behalf of an academic institution and for the sole purpose of non-commercial research.”

<https://developer.twitter.com/en/developer-terms/policy>

Siehe auch: Littman, J. (2019). *Twitter's Developer Policies for Researchers, Archivists, and Librarians*. <https://medium.com/on-archivy/twitters-developer-policies-for-researchers-archivists-and-librarians-63e9ba0433b2>

Diskussion: Veröffentlichung & Archivierung

- Welche negativen/unerwünschten Folgen kann die Veröffentlichung von Social-Media- oder anderen digitalen Spurdaten haben, wenn Personen in diesen identifizierbar sind?
- Durch die Verknüpfung mit welchen Informationen/Daten können sich Identification/Privacy Risks ergeben bzw. erhöhen?
- Sind Beispiele/Situationen denkbar, in denen Einschränkungen zum Archivieren und Teilen von Daten aus ToS von Plattformen mit forschungsethischen Prinzipien in Konflikt geraten können?

Veröffentlichung & Archivierung

- Breuer, J., Borschewski, K., Bishop, L., Vávra, M., Štebe, J. Strapcova, K., & Hegedús, P. (2021). *Archiving Social Media Data: A guide for archivists and researchers*. <https://doi.org/10.5281/ZENODO.5041072>
- Bishop, L., & Gray, D. (2017). Chapter 7: Ethical Challenges of Publishing and Sharing Social Media Research Data. In K. Woodfield (Ed.), *Advances in Research Ethics and Integrity* (Vol. 2, pp. 159–187). Emerald Publishing Limited. <https://doi.org/10.1108/S2398-601820180000002007>
- 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), 205395171773633. <https://doi.org/10.1177/2053951717736336>
- Weller, K., & Kinder-Kurlanda, K. E. (2016). A manifesto for data sharing in social media research. *Proceedings of the 8th ACM Conference on Web Science*, 166–172. <https://doi.org/10.1145/2908131.2908172>
- 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>

Ethikkommissionen

- Begutachtung durch Ethikkommissionen in aller Regel sinnvoll (und z.T. auch von Förderern und Publishern gefordert)
- „Übersicht sozial- und wirtschaftswissenschaftlicher Ethikkommissionen“ vom RatSWD:
<https://www.konsortswd.de/ratswd/themen/forschungsethik/ethikkommissionen/>
- Ethikkommission der DGPs:
<https://www.dgps.de/serviceangebote/ethikkommission/>
 - „Ein Wissenschaftler oder eine Wissenschaftlerin stellt einen Ethikantrag bei der Ethikkommission der DGPs, wenn nicht eine lokale oder eine medizinische Ethikkommission zuständig ist.“
 - kostenpflichtig (für DGPs-Mitglieder vergünstigt)
- Falls keine Ethikkommission vorhanden/zuständig: Eigene Ethikkommission gründen
- Beteiligung an Ethikkommissionen generell empfehlenswert (Einbringung der Perspektive des Fachs/Forschungsbereichs: Bsp. digitale Spurdaten)

Leitfäden?

- Es gibt keine „Patentlösungen“
 - Abhängig von Studiendesign, Datentyp, Stichprobe, Thema...
- Bestehende Beispiele/Vorlagen (z.B. für Informed Consent) müssen angepasst werden
- Entscheidungen liegen letztlich immer bei den Forschenden selbst
- Hilfestellung zur Identifizierung und Beantwortung forschungsethischer Fragen bei der Nutzung digitaler Verhaltensdaten: Fragenkataloge in *SOCRATES: Social Media Research Assessment Template for Ethical Scholarship* (Rau et al., 2021) sowie im *Report on legal and ethical framework and strategies related to access, use, re-use, dissemination and preservation of social media data* aus dem SERISS-Projekt (Hagen et al., 2019) → Nutzung ähnlich wie (oder auch in Kombination mit) Forschungsdatenmanagementplänen (SOCRATES bietet auch ein “GDPR Documentation Template”)
- Manche Publikationen bieten auch Entscheidungshilfen für konkretere Fragen

Beispiel Veröffentlichung von digitalen Spurdaten

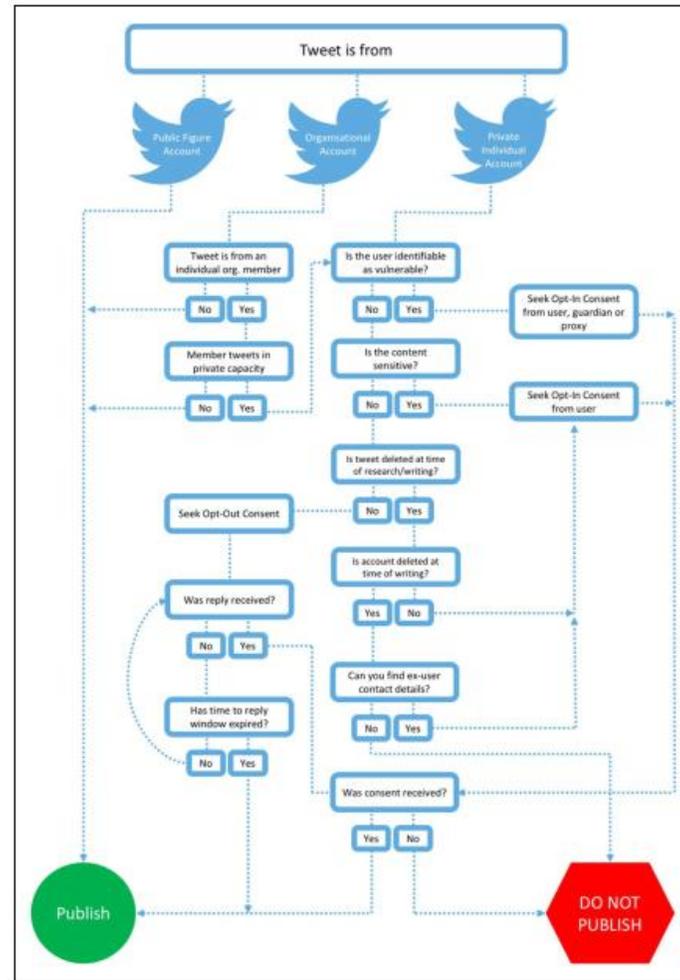


Abbildung aus Williams et al., 2017, S. 1163

Figure 1. Decision flow chart for publication of Twitter communications.

Online-Ressourcen

- SOCRATES: Social Media Research Assessment Template for Ethical Scholarship vom Hans-Bredow-Institut: <https://leibniz-hbi.github.io/socrates/>
 - inkl. “GDPR Documentation Template” (https://s.unhb.de/gdpr_doc)
- Ethics Section der Webseite der Association of Internet Researchers (AoIR): <https://aoir.org/ethics/>
 - Dort finden Sie u.a. auch das Dokument *Internet Research: Ethical Guidelines 3.0* (<https://aoir.org/reports/ethics3.pdf>)
- Legal and Ethical Issues of Social Media Collecting: Annotated Bibliography: <https://gwu-libraries.github.io/sfm-ui/resources/annotated-bibliography>
- Webseite des BMBF-Projekts „Forschungsethik in der Kommunikations- und Medienwissenschaft“ (FeKoM): <https://www.forschungsethik-kmw.de/startseite>
 - Bietet z.B. sehr gut aufbereitete Literatur- und Linksammlungen
- Web-Angebot des RatSWD zum Thema Forschungsethik: <https://www.konsortswd.de/ratswd/themen/forschungsethik/>
 - inkl. „Lehr- und Übungsmaterialien für die akademische Lehre. Grundlagen der Forschungsethik in den Sozial- und Wirtschaftswissenschaften.“

Publikationen

- Zimmer, M., & Kinder-Kurlanda, K. (Eds.). (2017). *Internet Research Ethics for the Social Age*. Peter Lang US. <https://doi.org/10.3726/b11077>
- Woodfield, K. (Ed.). (2017). *The Ethics of Online Research* (Vol. 2). Emerald Publishing Limited. <https://doi.org/10.1108/S2398-6018201802>
- Heise, N., & Schmidt, J.-H. (2014). Ethik der Online-Forschung. In M. Welker, M. Taddicken, J.-H. Schmidt, N. Jakob (Hrsg), *Handbuch Online-Forschung*, S. 519-539.
- Iphofen, R. (Ed.). (2020). *Handbook of Research Ethics and Scientific Integrity*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-16759-2>

Zusammenfassung

- Keine Patentrezepte für forschungsethische Fragen
- i.d.R. Abwägungen nötig: z.B. zwischen den Prinzipien Maximierung des Nutzens (Offenheit) und Minimierung von Risiken (z.B. im Hinblick auf mögliche negative Konsequenzen für Teilnehmer*innen)
 - Besonderheiten der Studie
 - Kontext der Forschung (z.B. regionale Unterschiede, politische Dimensionen)
 - letztlich auch grundlegende moralische Fragen (z.B. Wichtigkeit der Dimension “Care/Harm” im Sinne der Moral Foundations Theory; siehe <https://moralfoundations.org/>)
- Forschungsethische Fragen sollten möglichst früh im Forschungsprozess adressiert werden (ähnlich wie und idealiter auch zusammen mit Forschungsdatenmanagement; siehe Kinder-Kurlanda & Weller, 2020)
- Begutachtung durch Ethikkommissionen in aller Regel sinnvoll; ebenso Beteiligung an Ethikkommissionen
- Wenn möglich, ist die Einholung von Rechtsberatung (z.B. zu Informed Consent oder der Veröffentlichung von Daten) hilfreich; speziell bei besonders sensiblen Daten

Zusammenfassung

- Wenn möglich, Informed Consent einholen
 - Auf Formulierung des Informed Consent achten: Verständlichkeit, Zugang zu Daten (inkl. Nachnutzung)...
- Kombination von digitalen Spurdaten und Interviews/Befragung hilfreich, um Informed Consent einzuholen und Transparenz für Teilnehmende herzustellen/zuerhöhen
- Etwas Open Science ist besser als gar keine
 - z.B. prozessierte bzw. aggregierte Daten (replication/reproduction data), controlled access oder auch nur Metadaten veröffentlichen
- Für viele forschungsethische Fragen zur Arbeit mit digitalen Spurdaten gibt es Guidelines oder Vorlagen

Literatur

- Bail, C. (2020, May 28). What is Digital Trace Data? [Video]. YouTube. <https://www.youtube.com/watch?v=uuSWQN7uYhk>
- Beuthner, C., Breuer, J., & Jünger, S. (2021). Data Linking—Linking survey data with geospatial, social media, and sensor data. *GESIS Survey Guidelines*. https://doi.org/10.15465/GESIS-SG_EN_039
- Bishop, L., & Gray, D. (2017). Chapter 7: Ethical Challenges of Publishing and Sharing Social Media Research Data. In K. Woodfield (Ed.), *Advances in Research Ethics and Integrity* (Vol. 2, pp. 159–187). Emerald Publishing Limited. <https://doi.org/10.1108/S2398-601820180000002007>
- Boeschoten, L., Ausloos, J., Moeller, J., Araujo, T., & Oberski, D. L. (2020). Digital trace data collection through data donation. *ArXiv:2011.09851 [Cs, Stat]*. <http://arxiv.org/abs/2011.09851>
- Breuer, J., Al Baghal, T., Sloan, L., Bishop, L., Kondyli, D., & Linardis, A. (2021). Informed consent for linking survey and social media data—Differences between platforms and data types. *IASSIST Quarterly*, 45(1), 1–27. <https://doi.org/10.29173/iq988>
- 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>
- Hagen, S., Bishop, E. L., Kosciak, M., Vavra, M., Stebe, J., Ryan, L., Payne, E., Lovlie, A. G., Rod, L.-M., Straume, O., & Hogteveit Myhren, M. (2019). *Report on legal and ethical framework and strategies related to access, use, re-use, dissemination and preservation of social media data* (6.3). SERISS. https://seriss.eu/wp-content/uploads/2019/11/D6.3-Report-on-legal-and-ethical-framework-and-strategies..._FINAL.pdf
- 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>
- Hopf, C. (2016). Forschungsethik und qualitative Forschung. In C. Hopf, *Schriften zu Methodologie und Methoden qualitativer Sozialforschung* (pp. 195–205). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-11482-4_9
- 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>
- Kinder-Kurlanda, K. E., & Weller, K. (2020). Perspective: Acknowledging Data Work in the Social Media Research Lifecycle. *Frontiers in Big Data*, 3. <https://doi.org/10.3389/fdata.2020.509954>
- King, G. (1995). Replication, Replication. *PS: Political Science and Politics*, 28(3), 444. <https://doi.org/10.2307/420301>

Literatur

- 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>
- McKee, H. A., & Porter, J. E. (2009). *The ethics of internet research: A rhetorical, case-based process* (Vol. 59). Peter Lang.
- 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>
- Rau, J., Münch, F., & Asli, M. (2021): Social Media Research Assessment Template for Ethical Scholarship (SOCRATES): Your politely asking data ethics guide. (Social) Media Observatory. <https://leibniz-hbi.github.io/socrates/>
- 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>
- 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>
- 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*, 1-7. <https://doi.org/10.1080/10584609.2020.1744780>
- Van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Computational communication science – Toward open computational communication science: A practical road map for reusable data and code. *International Journal of Communication*, 13, 20.
- 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>
- Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313-325. <https://doi.org/10.1007/s10676-010-9227-5>