



## Digitale Daten in der qualitativen Lehrforschung

### Handreichung zum digitalen Datenmanagement für Studierende

Die vorliegende Handreichung klärt Spezifika der qualitativen *Online*forschung in Bezug auf die Erhebung, Speicherung und Verarbeitung digitaler Daten mit besonderem Fokus auf Social Media. Sie ist damit für Studierende, die im Rahmen der Methodenausbildung am Institut für Soziologie der LMU Onlineforschung betreiben möchten, eine Ergänzung zur allgemeinen Handreichung »Datenmanagement in der qualitativen Methodenausbildung«<sup>1</sup>. Einerseits sollen praktische Wie-Fragen der Erhebung, Speicherung und Verarbeitung von Onlinedaten geklärt, andererseits ethische und datenschutzrechtliche Spezifika aufgezeigt werden.<sup>2</sup>

#### Inhalt

1. Überblick und allgemeine Hinweise .....	2
2. Interview-Daten digital erheben .....	2
3. Social Media Daten erheben .....	3
3.1 Automatisierte Datenerhebung bei Twitter & YouTube .....	4
3.2 Händische Datenerhebung z.B. bei Facebook, Instagram, TikTok oder Reddit .....	6
4. Sonstige Webinhalte erheben .....	7
5. Stolpersteine .....	8
5.1 Ethische und rechtliche Spezifika: Anonymisierung und Zitation digitaler Daten .....	8
5.2 Aufbereitung der Daten .....	9
5.3 Reflexion technischer Infrastruktur .....	9
6. Weiterführende Literatur & Links .....	11

---

<sup>1</sup> Handreichung zum Datenmanagement in der qualitativen Methodenausbildung für Studierende siehe:

<https://www.qualitative-sozialforschung.soziologie.uni-muenchen.de/aktuelles/handreichung-stud/index.html>

<sup>2</sup> Hinweise zu veralteten Informationen in der Handreichung oder neuen Möglichkeiten der Erhebung digitaler Daten immer gerne an: [unger@lmu.de](mailto:unger@lmu.de)

## 1. Überblick und allgemeine Hinweise

Für die Verwendung digitaler Daten gibt es aufgrund ihrer Vielfältigkeit nicht das eine passende Verfahren, sondern nur den jeweils auf das konkrete Forschungsvorhaben zugeschnittenen Zugang. Für verschiedene digitale Daten (u.a. Onlineinterviews, Social Media Daten, Webseiteninhalte) werden hier allgemeine Möglichkeiten der Erschließung dieser Daten und dabei zu beachtende praktische, ethische und datenschutzrechtliche Dimensionen skizziert, die in Seminaren mit den Dozierenden für das jeweilige Lehrforschungsprojekt nochmals spezifisch zugeschnitten werden müssen.

Als erstes wird im Folgenden das digital vermittelte qualitative Interview betrachtet (Kapitel 2). Daraufaufgehend werden Möglichkeiten der Datenerhebung bei den sogenannten Social Media Plattformen aufgezeigt (Kapitel 3). In Kapitel 4 wird dann noch beschrieben, wie Inhalte von Webseiten allgemein erhoben werden können. Abschließend wird unter „Stolpersteine“ (Kapitel 5) auf verschiedene allgemeine Problematiken im Umgang mit digitalen Daten hingewiesen. Was in dieser Handreichung dezidiert nicht behandelt wird, ist, wie digitale Daten im Anschluss an die Erhebung ausgewertet werden können – sei dies rein mit klassischen Ansätzen der qualitativen Sozialforschung oder in Kombination mit neueren, computergestützten Verfahren. Zudem werden sogenannte Trace Data, also unbewusst hinterlassene Spurendaten im Netz, nicht behandelt.<sup>3</sup>

Es ist darauf hinzuweisen, dass auch für Onlineforschung die allgemeinen ethischen und datenschutzrechtlichen Grundsätze der qualitativen Sozialforschung wie Vertraulichkeit, informierte Einwilligung, Anonymisierung sowie sicheres Erheben, Verwalten und Verwahren der Daten – bei uns über die LMU-Dateninfrastruktur – gelten.

Für das Schreiben von Arbeiten gilt zudem, dass wie in jeder qualitativen Forschung auch Online-Erhebungen von digitalen Daten methodisch/methodologisch angemessen reflektiert werden müssen. Diese Reflexion ist Bestandteil jeder Prüfungsleistung. Ausgewählte Literatur und weiterführende Links zu qualitativer Online-Forschung finden Sie auch am Ende der Handreichung.

Die praktischen, ethischen und datenschutzrechtlichen Hilfeleistungen und Hinweise dieser Handreichung dienen als Orientierung für eigene Lehrforschungsprojekte von Studierenden. Die Handreichung ersetzt jedoch nicht eine individuelle proaktive Absprache der konkreten Umsetzung mit Ihren Dozierenden.

## 2. Interview-Daten digital erheben

Eine Alternative zu einem klassischerweise in Präsenz geführten qualitativen Interview sind digital vermittelte Interviews. Bei geltenden Kontaktbeschränkungen (z.B. im Zug der COVID-Pandemie) oder aus forschungspragmatischen Gründen wie der Erreichbarkeit der Zielgruppe oder räumlicher Distanz zwischen Befragten und Interviewenden kann dies je nach Forschungsinteresse ein passendes Vorgehen sein. Möglichkeiten der digitalen Vermittlung sind dabei E-Mail Interviews, Telefon- oder Video-Interviews (z.B. via Zoom).

---

<sup>3</sup> Bei näherem Interesse dazu siehe z.B. Florian Keusch & Frauke Kreuter 2021: Digital trace data. Modes of data collection, applications, and errors at a glance. In: Uwe Engel et al.: Handbook of Computational Social Science, Volume 1. London/New York 2021, S. 100-118. <https://doi.org/10.4324/9781003024583-8>.

Bei E-Mail Interviews nutzen Sie – wie auch für die sonstige Kommunikation im ganzen Forschungsprozesses – bitte nur Ihre Campus-E-Mail-Adresse<sup>4</sup> und verzichten Sie auf eine Weiterleitung an eine private E-Mail-Adresse. Bei besonders sensiblen Daten sollten Sie die E-Mails zudem verschlüsseln.<sup>5</sup>

Bei Telefon- und Video-Interviews nutzen Sie bitte so weit wie möglich – wie bei Interviews in Präsenz – nur professionelle Aufnahmegeräte des Instituts für Soziologie.<sup>6</sup> Sollte eine Ausleihe eines professionellen Aufnahmegeräts ausnahmsweise nicht möglich sein, satteln Sie, falls nicht bereits geplant, bitte auf ein Zoom-Interview um und nutzen die integrierte Aufnahmefunktion von Zoom – aber VORSICHT: unbedingt die Option „Aufzeichnung auf dem lokalen Computer speichern“ wählen (nicht in der Zoom-Cloud). Bei Zoom-Interviews verwenden Sie zudem Ihren LMU-Account<sup>7</sup> und führen die Interviews nur über geschützte Zoom-Meetings mit Meeting-ID und Kenncode. Sollten sich im Raum, in dem Sie selbst das Telefon- oder Video-Interview führen, sogenannte ‚always on‘ Geräte wie z.B. Smart Speaker befinden, schalten Sie diese zur Sicherstellung des Datenschutzes Ihrer Interviewperson aus.

In allen Fällen der digitalen Vermittlung von Interviews unterstützen Sie die Interviewpersonen bei technischen Fragen und stellen für den Fall von technischen Störungen jeweils alternative Kommunikationsmöglichkeiten zur Verfügung.

### 3. Social Media Daten erheben

Social Media Plattformen sind für digitale Gesellschaften der Gegenwart je nach wissenschaftlicher Fragestellung relevant. Im Folgenden wird für einige ausgewählte<sup>8</sup> Plattformen beschrieben, wie mehr oder weniger automatisiert mit Tools oder aber auch händisch Daten von diesen Social Media Plattformen erhoben werden können.

Bei Social Media Daten ist die Frage der Zugänglichkeit grundlegend. Sind Plattforminhalte überhaupt für uns als Forschende anders zugänglich als in der individuellen Ansicht als Benutzende? Je nach Dateninfrastruktur und AGBs der Plattformanbieter fällt die Zugänglichkeit von Daten sehr unterschiedlich aus und muss für eine Datenerhebung jeweils individuell in Bezug auf die jeweilige Social Media Plattform geklärt werden. Eine öffentliche Speicherung und Nachnutzung von Social Media Daten ist von allen großen Anbietern verboten, weshalb für Studierende keine Möglichkeit besteht, auf vorhandene öffentliche Social Media Datensätze zurückzugreifen. Für gewisse Plattformen ist eine automatisierte Erhebung möglich (siehe Unterkapitel 3.1), bei anderen muss auf einen händischen Zugang zurückgegriffen werden (siehe Unterkapitel 3.2).

---

<sup>4</sup> <https://webmail.lrz.de>

<sup>5</sup> Die genaue Anleitung findet sich unter den FAQs des IT-Supports des Instituts für Soziologie: <https://dienste.sozioologie.uni-muenchen.de/faq-pmwiki/pmwiki.php?n=Main.MailverschlSselung>

<sup>6</sup> Für den Geräteverleihe am Institut für Soziologie siehe: [https://www.sozioologie.uni-muenchen.de/institut/it-service1/software\\_geraete\\_verleih1/index.html](https://www.sozioologie.uni-muenchen.de/institut/it-service1/software_geraete_verleih1/index.html)

<sup>7</sup> <https://lmu-munich.zoom.us/>

<sup>8</sup> Die vorliegende Auswahl wurde willkürlich nach Relevanz der Plattformen im deutschsprachigen und europäischen Raum getroffen. Damit werden gewisse Plattformen und damit Bereiche westlicher Gesellschaften wie auch kulturelle Bereiche nicht westlicher Gesellschaften ausgeschlossen, die aber auch Forschungsgegenstand sein könnten (z.B. die chinesische Microblogging Webseite weibo oder die alternative Streaming-Plattform Nebula). Zur Reflexion eigener Datenzugänge siehe auch Kapitel 5 ‚Stolpersteine‘.

Zur Erhebung von Daten auf Social Media Plattformen, sei dies automatisiert oder händisch, ist in der Regel eine Anmeldung über einen Account notwendig. Zur Trennung eigener privater Accounts vom Forschungsvorhaben wie auch zur Verminderung datenschutzrechtlicher und ethischer Risiken wird dringend empfohlen, sich hierzu einen eigens erstellen Account für das Forschungsvorhaben zu erstellen. In diesem Account sollte zur Transparenz das eigene Forschungsinteresse offengelegt werden, etwa in der Profilbeschreibung.

Social Media Plattformen können nicht nur zur Erhebung von darauf enthaltenen Inhalten, sondern bei passender Zielgruppe (z.B. Mitglieder eines Forums, in dem es um ein interessantes Thema geht) auch gut zur Rekrutierung von Interviewpartner\*innen genutzt werden. Dabei kann es je nach Zielgruppe hier Sinn ergeben, doch den eigenen privaten Account zu nutzen, um z.B. auf ein bereits vorhandenes Netzwerk an Personen zurückzugreifen. Die Kommunikation mit einzelnen Personen sollte dann aber so schnell wie möglich auf eine sichere Infrastruktur, d.h. die Campus-E-Mail, verlagert werden.

### 3.1 Automatisierte Datenerhebung bei Twitter & YouTube

Zur automatisierten Datenerhebung besteht bei gewissen Social Media Plattformen die Möglichkeit, direkt über die Programmierschnittstelle einer Plattform (die sogenannte API<sup>9</sup>) automatisiert Zugang zu den verfügbaren Daten zu erhalten. Wer selbst über Programmierkenntnisse z.B. in Python verfügt, kann sich somit zielgerichtet einen auf das eigene Forschungsvorhaben ausgerichteten Zugang über die entsprechende API programmieren.<sup>10</sup> Aber auch ohne Programmierkenntnisse ist teils eine automatisierte Erhebung über bestehende, auf den APIs aufbauende Tools möglich. Das konkrete Vorgehen sollte dabei immer in einem Forschungstagebuch oder in Feldnotizen festgehalten werden. Im Folgenden werden Möglichkeiten der automatisierten Datenerhebung für exemplarische Plattformen, bei denen ein solcher Zugang existiert, dargestellt.

#### Twitter

Eine automatisierte Datenerhebung der primären Plattforminhalte (z.B. bei Twitter Tweets, bei Facebook Posts, bei YouTube Videos) ist aktuell nur bei Twitter über deren API möglich – und dies auch nur beschränkt für eine Woche zurückführend und einem Maximum von 18.000 Tweets pro Erhebung.<sup>11</sup> Das bedeutet, dass entsprechende Erhebungen geplant werden müssen, da eine Datenerfassung nur „in Echtzeit“ und nicht für die Vergangenheit möglich ist. So könnte etwa eine Erhebung von wöchentlich mittels Tools gespeicherten Tweets stattfinden, die dann im Anschluss

---

<sup>9</sup> Eine API (englisch application programming interface) ist allgemein eine Programmierschnittstelle zur Anbindung anderer Programme an ein primäres Programm. Im Falle von Social Media dient eine API hier dazu, entsprechende Daten der Plattformen über dafür programmierte Tools automatisch zu erheben und zu speichern.

<sup>10</sup> Siehe dazu etwa <https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/>. Ein Tool, das mit Programmierkenntnissen nutzbar ist und die APIs bereits strukturiert ausliefert, ist Tweepy, vgl. <https://www.tweepy.org/>.

<sup>11</sup> Für Forschungsteams besteht die Möglichkeit der Beantragung eines Accounts als Forschende\*r (Academic Research Access, vgl. <https://developer.twitter.com/en/products/twitter-api/academic-research>), mit welchem zeitlich zurückliegende Tweets gespeichert werden können. Dies ist durch Twitter jedoch erst ab Masterarbeiten möglich. Außerdem ist die Bezahlung von durch Twitter zusammengestellten Datensätzen zu Forschungszwecken möglich, für ein Lehrforschungsprojekt o.ä. ist dies jedoch weder möglich noch sinnvoll.

analysiert werden. Iterative Anpassungen der Datengrundlage, wie etwa im Zuge der Grounded Theory üblich, sind dabei nicht für die Vergangenheit möglich.

Das zugänglichste Tool für Twitter ist die Software MaxQDA, welche neben ihrer Hauptfunktion der computergestützten qualitativen Analyse auch zur automatisierten Erhebung von Tweets verwendet werden kann.<sup>12</sup> Dem Vorteil der einfachen Bedienbarkeit von MaxQDA steht jedoch der Nachteil gegenüber, dass MaxQDA eine kostenpflichtige Software ist. Für Studierende wird vom Institut für Soziologie bei Bachelor- oder Masterarbeiten auf Antrag eine dreimonatige Lizenz zur Verfügung gestellt.<sup>13</sup> Nach Absprache mit den Dozierenden können aber auch in anderen Fällen Lizenzen zur Verfügung gestellt werden.

Kostenfreie Alternativen zur automatisierten Erhebung von Tweets stellen die beiden Tools ‚TAGS‘<sup>14</sup> und ‚Social Feed Manager‘<sup>15</sup> dar, die in etwa die gleiche Funktionalität besitzen, als Open Source bzw. Freeware Software jedoch frei zugänglich sind. Leider sind sie gleichzeitig auch weniger intuitiv zu bedienen als MaxQDA. Beide Tools haben jedoch ausführliche Schritt für Schritt Erklärungen auf ihren Webseiten, die eine Verwendung nach kurzer Einarbeitung möglich machen. Im Netz finden sich weitere Tools zur Speicherung von Tweets, aktuell (Stand Februar 2022) sind diese beiden jedoch besonders zu empfehlen.

Darüber hinaus bestehen verschiedene Tools zur Analyse von Tweets. Beispielsweise kann mit der Webseite ‚Mention Map‘<sup>16</sup> in Bezug auf ein gewähltes Twitter-Profil oder ein Hashtag ein Netzwerkgraph ausgegeben wird. Je nach Forschungsinteresse kann dies ergänzend zu einer Erhebung von Tweets sinnvoll eingesetzt werden.

## YouTube

Bei YouTube können die primären Inhalte, also die YouTube Videos, nicht über die API automatisiert heruntergeladen werden. Jedoch können mit Tools Kommentare und auch die Untertitel (z.B. als Transkripte) von Videos automatisiert erhoben werden. Dies ist wie bei Twitter ebenfalls über MaxQDA möglich (siehe oben).<sup>17</sup> Eine kostenfreie Alternative zur Erhebung der Untertitel von Videos ist das Tool ‚4K Video Downloader‘, dass zwar sehr nutzerfreundlich, jedoch auf 30 Downloads pro Tag beschränkt ist.<sup>18</sup> Eine kostenfreie Alternative zur Erhebung von Kommentaren wie auch von Untertiteln ist das Tool ‚FacePager‘, welches unbegrenzte Downloads ermöglicht, jedoch in der Anwendung eine gewisse Einarbeitung in das nicht sehr intuitive Nutzerinterface voraussetzt.<sup>19</sup>

Sollen YouTube Videos erhoben werden, ist dies ebenfalls über das Tool ‚4K Video Downloader‘ möglich. Dabei muss aber beachtet werden, dass dies nicht über die API von YouTube geschieht,

---

<sup>12</sup> <https://www.maxqda.de/hilfe-mx20/dokumente/twitter>

<sup>13</sup> Zum Verleih von MaxQDA-Lizenzen am Institut für Soziologie siehe: [https://www.sozioogie.uni-muenchen.de/institut/it-service1/software\\_geraete\\_verleih1/maxqda/index.html](https://www.sozioogie.uni-muenchen.de/institut/it-service1/software_geraete_verleih1/maxqda/index.html)

<sup>14</sup> <https://tags.hawksey.info/>

<sup>15</sup> <https://gwu-libraries.github.io/sfm-ui/>; der ‚Social Feed Manager‘ ist auch für weibo, Flickr und Tumblr verwendbar.

<sup>16</sup> <https://mentionmapp.com/>

<sup>17</sup> <https://www.maxqda.de/hilfe-max18/dokumente/youtube-daten>

<sup>18</sup> <https://www.4kdownload.com/>

<sup>19</sup> <https://github.com/strohne/Facepager>

also nicht alles technisch Mögliche auch rechtlich erlaubt ist. Für eine wissenschaftliche Analyse dürfen jegliche Videos gespeichert werden. Sollen Videos (bzw. Screenshots davon) jedoch darüber hinaus z.B. in Publikationen oder Präsentationen verwendet werden, dürfen urheberrechtlich nur solche verwendet werden, welche unter der Creative Commons Lizenz hochgeladen wurden. Bei der YouTube Suche können Videos in der erweiterten Suche nach dieser Creative Commons Lizenz gefiltert werden. Zudem ist in den Beschreibungen der Videos vermerkt, falls diese unter einer solchen Lizenz publiziert sind.

Schlussendlich gibt es noch die Webseite ‚YouTube Data Tools‘, mit deren Tools zwar nicht Inhalte wie Videos, Untertitel oder Kommentare, jedoch Informationen über Kanäle, Video-Playlists oder auch Video-Netzwerke erhoben werden können, die allenfalls ergänzend zu einer qualitativen Analyse von Inhalten verwendet werden.<sup>20</sup> Ähnliche Funktionalitäten sind beim bereits erwähnten Tool ‚Face Pager‘ zu finden.

### **3.2 Händische Datenerhebung z.B. bei Facebook, Instagram, TikTok oder Reddit**

Bei vielen Social Media Plattformen wie z.B. Facebook, Instagram oder Reddit ist eine in Kapitel 3.1 beschriebene automatisierte Datenerhebung aufgrund fehlender API gar nicht oder aufgrund fehlender Tools nur sehr eingeschränkt möglich<sup>21</sup>, weshalb für eine Erhebung auf einen händischen Zugang zurückgegriffen werden muss.<sup>22</sup> Dieser wird hier allgemein beschrieben und kann somit grundsätzlich auf jegliche Art von Social Media Plattform angewandt werden.

Der entscheidende Unterschied zur automatisierten Erhebung über eine API ist, dass die Datenerhebung über einen Zugang zu den Plattformen geschieht, der sich nicht vom Zugang der Nutzenden der Plattform selbst unterscheidet. Sie betrachten also über einen eigenen Account und die Webseite oder App der Plattform die Inhalte auf der Plattform. Wird per Webbrowser auf die Social Media Plattform zugegriffen, können die Seiten mit den relevanten Inhalten als HTML-Dateien gespeichert werden. Dies geschieht bei den meisten Webbrowsers über die Menüleiste und einer Funktion wie etwa ‚Seite speichern unter...‘. Wird per App auf die Social Media Plattform zugegriffen, können die interessierten Inhalte per Screenshot gespeichert werden. Wie sie mit ihrem mobilen Gerät Screenshots machen können, finden Sie mit der Suchmaschine Ihrer Wahl schnell heraus. Screenshots sind auch beim Zugriff über Webbrowser möglich, wie dies in Ihrem Betriebssystem funktioniert, zeigt ebenfalls eine Recherche per Suchmaschine.

Handelt es sich bei interessierenden Daten um audiovisuelle Inhalte wie TikTok Videos oder videoförmige Instagram-Stories, die nicht automatisiert oder händisch heruntergeladen werden können, können diese per sogenanntem Screencast aufgenommen und gespeichert werden. Bei den

---

<sup>20</sup> <https://tools.digitalmethods.net/netvizz/youtube/>

<sup>21</sup> Mit Programmierkenntnissen siehe für Reddit, 4Chan, Telegram und andere das Tool 4CAT<sup>2</sup> (<https://wiki.digitalmethods.net/Dmi/Tool4CAT>) und für TikTok und Instagram in Kombination mit ‚4CAT‘ das Tool ‚Zeeschuimer‘ (<https://github.com/digitalmethodsinitiative/zeeschuimer>).

<sup>22</sup> Die händische Erhebung von Inhalten aus Social Media Plattformen wird dabei als Werkzeug aus dem Baukasten einer digitalen Ethnografie verstanden. Sie kann aber wie hier beschrieben auch davon losgelöst für Forschungsvorhaben genutzt werden, die nicht eine vollumfängliche ethnografische Studie zum Ziel haben. Weiterführende Literatur zur digitalen Ethnografie siehe unter Kapitel 6.



gängigen Betriebssystemen von Smartphones und Computern sind entsprechende Funktionen vorhanden.<sup>23</sup>

Die Dokumentation des Vorgehens sollte dabei im Forschungstagebuch im Sinne eines Logbuches festgehalten werden, in dem die Art der Erhebung, die Webseiten-URL bzw. der Fundort in der App<sup>24</sup> sowie der Zeitpunkt der Erhebung der gespeicherten Dateien festgehalten wird. Bei solchen Einträgen ist es zielführend, den Kontext der Erhebung ebenfalls festzuhalten (Wie ist man auf den Inhalt gestoßen? In welchem Zusammenhang erscheint der interessierte Inhalt auf der Plattform?). Einträge im Forschungstagebuch können zudem auch auf der Plattform gemachte Beobachtungen umfassen, die über die gespeicherten Daten und deren Kontext hinaus gehen.

#### 4. Sonstige Webinhalte erheben

Neben digitalen Daten aus Social Media Plattformen finden sich auf verschiedensten Webseiten zahlreiche andere Inhalte wie beispielsweise Onlinenews, Blogbeiträge oder Wikipediaartikel, die für ein Forschungsvorhaben von Interesse sein können. Auch öffentliche Archive, Datenbanken zum kulturellen Erbe wie auch zu statistischen Daten, Dokumente der staatlichen Verwaltung, von Bundestags- oder Stadtratsprotokollen bis zu Gesetzen und Verordnungen sind weitestgehend öffentlich über das Web zugänglich. Wenn Sie bereits wissen, welche Webinhalte Sie wo finden bzw. diese über eine Suchmaschine einfach auffinden können, können Sie diese gezielt händisch erheben. Gewisse Webinhalte sind bereits durch die Anbieter zum Download aufbereitet, wie z.B. Bundestagsprotokolle als PDF Downloads. Interessiert ein Webinhalt, der nicht in diesem Sinne explizit herunterladbar ist, können Webseiten händisch als HTML-Dateien oder per Screenshot gespeichert werden (genauer dazu siehe Kapitel 3.2).

Auch bei Webinhalten über Social Media hinaus gibt es Möglichkeiten der automatisierten Erhebung, die einerseits sinnvoll sein können, wenn zwar bekannt ist, was genau erhoben werden soll (z.B. Bundestagsprotokolle), aber für die gewünschte Anzahl (alle online verfügbaren Bundestagsprotokolle) eine händische Speicherung deutlich zu lange dauern würde. Andererseits sind solche automatisierten Erhebungen auch sinnvoll, wenn mindestens eine, aber noch nicht alle relevanten Webseiten bekannt und eine händische Suche zu ineffizient wäre. Die entsprechenden automatisierten Verfahren für diese Vorgehensweisen nennt man ‚Crawling‘ – dass nach gewissen Regeln automatisierte Auffinden von Webseiten – und ‚Scraping‘ – das nach gewissen Regeln automatisierte Speichern von Informationen von zuvor ‚gescrawlt‘ Webseiten. Zurzeit gibt es aber leider keine für sozialwissenschaftliche Zwecke geeigneten Tools, mit denen auch ohne Programmierkenntnisse Crawling und Scraping möglich wäre.<sup>25</sup>

---

<sup>23</sup> Bspw. Android: <https://support.google.com/android/answer/9075928?hl=de>; iOS: <https://support.apple.com/de-de/HT207935>; Windows: [https://support.microsoft.com/de-de/windows/aufzeichnen-eines-spielclips-auf-dem-pc-mit-xbox-game-bar-2f477001-54d4-1276-9144-b0416a307f3c#ID0EDD=Windows\\_10](https://support.microsoft.com/de-de/windows/aufzeichnen-eines-spielclips-auf-dem-pc-mit-xbox-game-bar-2f477001-54d4-1276-9144-b0416a307f3c#ID0EDD=Windows_10); MacOS: <https://support.apple.com/de-de/HT208721>

<sup>24</sup> Wenn in der App keine URL für den entsprechenden Inhalt generiert werden kann, kann im Forschungstagebuch auch ein nachvollziehbarer Fundort angegeben werden, wie z.B. Post des Profils @Beispielprofil vom 02.02.2022.

<sup>25</sup> Bei näherem Interesse dazu siehe Seppe vanden Broucke & Bart Baesens 2018: Practical Web Scraping for Data Science. Best Practices and Examples with Python. Leuven 2018. <https://doi.org/10.1007/978-1-4842-3582-9>

## 5. Stolpersteine

Zuletzt sollen nun noch einige Stolpersteine besprochen werden, die bei der Erhebung, Speicherung und Verarbeitung digitaler Daten im Weg liegen können. Allgemein sei nochmals darauf hingewiesen, dass die methodischen Zugänge zu digitalen Daten noch in ihren Kinderschuhen stecken. Gerade bei den verschiedenen Tools zur Erhebung digitaler Daten muss viel ausprobiert werden, bis gewünschte Ergebnisse erzielt werden können. Mit der vorliegenden Handreichung sowie etwas Geduld, Muße und Spaß am Ausprobieren steht einem erfolgreichen Lehrforschungsprojekt oder einer Abschlussarbeit, die sich an neuere digitale Daten heranwagt, aber nichts mehr im Wege, wenn zudem einige Spezifika beim Umgang mit digitalen Daten beachtet werden.

### 5.1 Ethische und rechtliche Spezifika: Anonymisierung und Zitation digitaler Daten

Bei der Anonymisierung des Datenmaterials gilt auch bei online erhobenen digitalen Daten wie sonst in der qualitativen Sozialforschung die Grundregel: so wenig wie möglich, so stark wie nötig. Bei der Erhebung von Daten aus dem Internet, also Social Media Daten oder sonstigen Webinhalten, müssen jedoch einige spezielle Herausforderungen beachtet werden. Einerseits sind Daten aus dem Internet durch eine Rückwärtssuche relativ einfach zu ihrem Ursprung rückverfolgbar. Entsprechend muss bei der Anonymisierung und Zitation etwas anders abgewogen werden als bei sonstigen qualitativen Daten. Bei Social Media Plattformen haben Produzent\*innen der Daten in der Regel zwar allgemeinen Geschäftsbedingungen zugestimmt, die die Rechte an den Daten den Plattformen übertragen. Das ist eine rechtliche Absicherung, aber längst noch keine ethische. Auch wenn die Urheberrechte damit an die Plattform abgegeben sind und die Daten über die offizielle API der Plattform bezogen werden, sind bei Publikationen rechtliche Feinheiten zu beachten. Besonders wichtig sind in allen Fällen der automatisierten Speicherung Fragen der Anonymisierung, da wir in der Regel kein informiertes Einverständnis von einzelnen Personen einholen können. Aufgrund dieser rechtlichen und ethischen Spezifika sollte sich bei Fragen der informierten Einwilligung, Anonymisierung sowie Zitation an folgendem Schema orientiert werden:

Sind die Daten (1.) komplett öffentlich zugänglich, muss nicht zwingend eine informierte Einwilligung eingeholt werden, zitierte Textstellen sollten jedoch stark anonymisiert und bei ethisch besonders heiklen Themen bzw. einer im Falle einer Rückverfolgung hohen Gefährdung der Personen, deren Daten verwendet werden, zudem sinngemäß paraphrasiert werden. Sind die Daten (2.) teil-öffentlich, wie z.B. bei Foren mit benötigtem Login, sollte über sogenannte Gate Keeper, also Forenleiter, Webseitenbetreiber etc., eine informierte Einwilligung eingeholt werden und für Anonymisierung und Zitation der gleiche Maßstab wie bei öffentlich zugänglichen Daten angewandt werden. Sind die Daten (3.) privat, also z.B. nur bei einer gegenseitigen Freundschaft auf Facebook sichtbar, muss unbedingt eine informierte Einwilligung eingeholt werden. Da solche Daten jedoch fast nicht rückverfolgbar sind, kann die Anonymisierung weniger strikt ausfallen und im Normalfall direkt zitiert werden. Es empfiehlt sich, trotzdem wie im Falle von Interviews nicht den Usernamen zu verwenden und ggf. auch andere Aspekte zur Person zu verändern.

Bei nicht-textbasierten Daten, also gespeicherten Bildern oder Videos, ist sich am selben Anonymisierungsschema zu orientieren, wobei da teils nicht nur Namen und Inhalte, sondern auch audiovisuelle Elemente wie Gesichter, Umgebungen oder Stimmen anonymisiert werden müssen.



Die Abwägung der Notwendigkeit einer informierten Einwilligung, des Anonymisierungsgrades wie auch der Zitationsart bleibt schlussendlich immer ein am Einzelfall zu entscheidendes Dilemma zwischen Personenschutz und Nachvollziehbarkeit der eigenen Interpretation von Daten. Insbesondere bei Forschungsvorhaben über/mit gesellschaftlichen Minderheiten und Diskriminierungen fällt auf der Seite der Nachvollziehbarkeit zusätzlich die Frage der Sichtbarkeit und Repräsentation ins Gewicht.

Der Quellenbeleg bei Zitaten aus Social Media oder sonstigen Webdaten erfolgt so, dass das Zitat eindeutig auf das entsprechende Quellenmaterial verweist. Das bedeutet beispielsweise bei Tweets, dass diese in der Datenaufbereitung durchnummeriert werden und die entsprechende Nummer bei einem Zitat angegeben wird (siehe dazu auch Kapitel 5.2). Zitate sollten zudem – gleich wie bei Interviewzitaten – zur intersubjektiven Nachvollziehbarkeit immer mit genügend Kontext ausgestattet sein.

Schlussendlich sei noch darauf verwiesen, dass es außerdem in unserer ethischen Verantwortung liegt, Inhalte zu melden, die gegen Rechtsvorgaben verstoßen, also zum Beispiel Hassrede oder Pornografie verbreiten. Für solche Bereiche ist zudem relevant, dass wir, wenn wir Webinhalte suchen und speichern, ebenfalls als Nutzer mitgezählt werden.

## 5.2 Aufbereitung der Daten

Wie bei sonstigen Daten der qualitativen Sozialforschung auch, sind sowohl für Social Media Daten als auch für sonstige Webdaten Aufbereitungen und Sortierungen notwendig, um die Datenmengen zu organisieren und für die Analyse zugänglich zu gestalten. Eine eigene Datenbank oder doch Datenablage aufzusetzen und zu strukturieren ist ein essenzieller Teil der Datenaufbereitung. Es ist hilfreich, sich hierfür entweder an den Datenformaten der genutzten API oder dem genutzten Tool zu orientieren. Gerade, wenn unterschiedliche Datensätze kombiniert werden, bedeutet das Mapping und Anpassen der Datensätze aneinander einen wichtigen Schritt.

Die in dieser Handreichung aufgeführten Tools geben allesamt CSV-Dateiformate<sup>26</sup> aus. Werden die Daten nicht weiter computationell ausgewertet, lohnt es sich, zur einfacheren Handhabbarkeit die CSV-Dateien in Excel-Dateien umzuformatieren.<sup>27</sup> Dateien sollten einheitlich und nachvollziehbar (was? von wann?) beschriftet werden.

## 5.3 Reflexion technischer Infrastruktur

Bei jeglicher (qualitativer) Sozialforschung müssen die methodischen und methodologischen Zugänge reflektiert werden. In Bezug auf digitale Daten fällt dabei schnell auf, dass die technische Vermittlung einen besonderen Stellenwert innehat: Die durch die Plattformen und Webseiten stark strukturierten technischen Möglichkeitsräume formen das Handeln der Menschen im digitalen Raum. Die von Plattformen angebotenen Zugänge und die verfügbaren Erhebungstools ermöglichen spezifische Zugriffe von Forschenden, die auch anders aussehen könnten. Schlussendlich stellt sich auch die Frage, wer wie in digitaler Kommunikation partizipieren kann? Wir belassen es

---

<sup>26</sup> [https://de.wikipedia.org/wiki/CSV\\_\(Dateiformat\)](https://de.wikipedia.org/wiki/CSV_(Dateiformat))

<sup>27</sup> Siehe dafür: <https://docs.microsoft.com/de-de/power-automate/desktop-flows/how-to/convert-csv-excel>

hier bei diesen Andeutungen, die Sie bei ihrem Forschungsvorhaben in ihren Konsequenzen auf allen Stufen des Forschungsprozesses mitreflektieren sollten.

## 6. Weiterführende Literatur & Links

- Bampton, Roberta; Cowton, Christopher; Downs, Yvonne (2013): The E-Interview in Qualitative Research. In: *Advancing Social and Business Research Methods With New Media Technology*. Hershey, PA: IGI Global, S. 329-343.
- Deakin, Hannah; Wakefield, Kelly (2014): Skype Interviewing: Reflections of Two PhD Researchers. In: *Qualitative Research* 14 (5), S. 603-616.
- Gnambs, Tino; Batinic, Bernad (2020): Qualitative Online-Forschung. In: Mey, Günter; Mruck, Katja (Hrsg.): *Qualitative Forschung in der Psychologie*. Wiesbaden: Verlag für Sozialwissenschaften, S. 97-112.
- Gray, Lisa; Wong-Wylie, Gina, Rempel, Gawn; Cook, Karen (2020): Expanding Qualitative Research Interviewing Strategies: Zoom Video Communications. In: *The Qualitative Report* 25 (5), S. 1292-1301.
- Hallett, Ronald; Barber, Kirsten (2014): Ethnographic Research in a Cyber Era. In: *Journal of Contemporary Ethnography* 43 (3), S. 306-330.
- Heise, Nele; Schmidt, Jan-Hinrik (2014): Ethik der Online-Forschung. In: Martin Welker; Monika Taddicken; Jan-Hinrik Schmidt & Nikolaus Jakob (Hrsg.): *Handbuch Online-Forschung. Sozialwissenschaftliche Datengewinnung und -auswertung in digitalen Netzen*. Köln: Herbert v. Halem Verlag, S. 519-540.
- Hine, Christine (2015): *Ethnography for the Internet. Embedded, Embodied and Everyday*. London u.a.: Routledge.
- Knox, Hannah; Nafus, Dawn (Hg.) (2018): *Ethnography for a data-saturated world*. Manchester: MUP.
- Koch, Gertraud (2014): Ethnografieren im Internet. In: Christine Bischoff, Karoline Oehme-Jüngling und Walter Leimgruber (Hg.): *Methoden der Kulturanthropologie*. Bern: utb, S. 367–382.
- Pink, Sarah; Horst, Heather A.; Postill, John; Hjorth, Larissa; Lewis, Tania; Tacchi, Jo (2016): *Digital Ethnography. Principles and Practice*. Los Angeles u.a.: SAGE.
- Thimm C., Nehls P. (2019): Digitale Methoden im Überblick. In: Baur N., Blasius J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Verlag für Sozialwissenschaften, S. 973-990.
- Tzanetakis, Meropi (2021): Qualitative Online-Forschungsmethoden: Digitale Datenerhebung und ethische Herausforderungen, In: Prainsack, Barbara & Pot, Mirjam: *Qualitative und interpretative Methoden in der Politikwissenschaft*. Wien: Facultas, S- 130-141.

## **Weiterführende Links**

Open Source Tools for Social Science Researchers: <https://open-source-social-science.github.io/>

Tools and Services der Digital Research Infrastructure for the Arts and Humanities: <https://de.dariah.eu/en/web/guest/dienste-und-werkzeuge>

Toolsammlung der Digital Methods Initiative: <https://wiki.digitalmethods.net/Dmi/ToolDatabase>

Tutorials von Miriam Posner zu Digitalen Methoden: <https://miriamposner.com/blog/tutorials-written/>