



Grundlagenbeitrag: Nutzungsdatenanalyse digitaler Medien als Instrument der evaluativen Wissenschaftskom- munikationsforschung

Armin Hempel

Zusammenfassung

Der Grundlagenbeitrag skizziert die Nutzungsdatenanalyse für den Bereich der Evaluationsforschung digitaler Wissenschaftskommunikationsprojekte und diskutiert ihre Stärken wie ihre Herausforderungen. Dazu werden verschiedene Typen von Nutzungsdaten beschrieben und exemplarisch Werkzeuge vorgestellt, diese zu erheben. Inwiefern kann es lohnenswert sein, Nutzungsdatenanalyse innerhalb von digitalen Wissenschaftskommunikationsprojekten zu betreiben und was kann bereits während der Projektplanungsphase beachtet werden, um Nutzungsdatenanalysemethoden sinnvoll einzusetzen? Es wird beschrieben, wie Erwägungen zum Datenschutz eine Herausforderung für die Nutzungsdatenanalyse darstellen, der damit verbundene Aufwand jedoch durch die Verwendung passender Werkzeuge auf ein Minimum reduziert werden kann. Zuletzt wird das Desiderat formuliert, die Nutzungsdatenanalyse als Methodenset weiter zu elaborieren – auch, um zukünftige digitale Wissenschaftskommunikationsformate untersuchen zu können.

A. Hempel (✉)

SFB 980 „Episteme in Bewegung“, Freie Universität Berlin, Berlin, Deutschland

E-Mail: armin.hempel@fu-berlin.de

© Der/die Autor(en) 2023

P. Niemann et al. (Hrsg.), *Evaluationsmethoden der*

Wissenschaftskommunikation, https://doi.org/10.1007/978-3-658-39582-7_10

155

1 Einleitung

Die Digitalisierung transformiert unsere Gesellschaften seit Mitte der 1990er Jahre grundlegend. Diese Transformation erstreckt sich über alle gesellschaftlichen Teilbereiche und verändert so auch die Wissenschaften sowie deren Beziehung zur Gesellschaft nachhaltig (Neuberger et al. 2021).

Gleichzeitig erfahren auch die Wissenschaftskommunikation und deren Erforschung eine Transformation: Jedes neu entstehende digitale Medium bietet immer auch eine neue Möglichkeit, Wissenschaft zu kommunizieren. Damit einher gehen dann immer auch neue Forschungsfragen, die es erforderlich machen, diese digitalen Wissenschaftskommunikationsformen zu untersuchen. Das umfasst Fragen nach den untersuchten Projekten an sich, den Zielen, die mit ihnen verfolgt werden, ihrer Wirkung sowie danach, wie sie zu verbessern sind (Ziegler und Fischer 2020).

Beim Nachdenken über evaluative Forschungsdesigns im Rahmen der Wissenschaftskommunikationsforschung liegt es aus unterschiedlichen Gründen nahe, Methoden der Nutzungsdatenanalyse mit einzubeziehen.

Denn – egal, ob Social-Media-Präsenzen, Blogs, Podcasts, Smartphone-Apps, Internetvideos, E-Mail-Newsletter, Messenger-Dienste oder klassische Projektwebseiten – bei so gut wie allen digitalen Medien der Wissenschaftskommunikation werden automatisch Nutzungsdaten generiert, die für die Bewertung von Projekten als unterstützende Datengrundlage herangezogen werden können.

Nutzungsdaten können dabei – je nach Medium und Auslegung – bloße Abrufzahlen, aber auch personenbezogene Daten sein. Sie erlauben Rückschlüsse auf den geografischen Aufenthaltsort und die technische Ausstattung der Nutzer:innen oder enthalten Informationen zur Verweildauer und dem eingeschlagenen Weg durch eine Website. Häufig bleiben diese Daten ungenutzt, sei es aus Unkenntnis darüber, dass sie existieren, aufgrund von Schwierigkeiten, auf sie zuzugreifen bzw. sie zu interpretieren oder auch aus Sorge vor datenschutzrechtlichen Problemen.

Dieser Beitrag richtet sich an Forschende und Praktiker:innen – explizit an Einsteiger:innen in das Gebiet – die planen, im Rahmen eines evaluativen (Forschungs)-Designs ihr eigenes Projekt unter die Lupe zu nehmen und selbst Nutzungsdatenanalyse zu betreiben. Er versteht sich als eine erste Handreichung zur Annäherung an das Themenfeld. Die Nutzungsdatenanalyse wird dabei nicht als eine etablierte wissenschaftliche Methode vorgestellt, sondern es werden verschiedene Perspektiven auf die Nutzungsdatenanalyse als Instrument skizziert und ein Überblick über ihre Stärken und Schwächen gegeben.

Der detailliertere Umgang mit den durch Nutzungsdatenanalyse erlangten Daten, ihre Bereinigung, Verarbeitung, Aufbereitung, Visualisierung, Archivierung und eventuelle Weitergabe können nicht Gegenstand dieses Beitrags sein und müssen unberücksichtigt bleiben. Es existiert aber eine Vielzahl von disziplinspezifischen Leitlinien zum Forschungsdatenmanagement (Deutsche Forschungsgemeinschaft [DFG] 2021), die hier weiterhelfen können.

2 Nutzungsdatenanalyse – Eine Kurzdefinition

Nutzungsdatenanalyse ist die Sammlung, systematische Erhebung und Auswertung von selbst oder durch Dritte erhobenen Daten, die bei der Nutzung digitaler Angebote anfallen (McFadden 2005). Ihre Verfahren werden auch mit den Begriffen „Logfile-Analyse“, „Web Mining“, „User Tracking“ oder „Social-Media-Analyse“ umschrieben (Priemer 2004). Während sie in der Marktforschung als etabliertes Werkzeug gilt, um bspw. zu erfahren, ob eine Werbekampagne erfolgreich ist, wie gut eine Website funktioniert, welche Zielgruppen intensiver umworben oder fallen gelassen werden sollten oder in welcher Art ein eCommerce-Angebot gestaltet sein muss, um letztendlich viele Verkäufe zu realisieren, ist der Einsatz der Nutzungsdatenanalyse in anderen Feldern nicht unumstritten. Im Online-Journalismus beispielsweise wird die unreflektierte Anwendung von Nutzungsdatenanalysen dafür kritisiert, dass redaktionelle Entscheidungen darauf fußen, wie und mit welchen Inhalten die meisten Klicks generiert werden können, um Werbeumsätze in die Höhe zu treiben. So könne z. B. eines der Ziele des Journalismus, das Befördern des demokratischen Prozesses, durch eine rein auf Klickzahlen ausgelegte Publikationsstrategie in Gefahr gebracht werden (Tandoc und Thomas 2015).

Ähnlich gelagerte, grundlegende Divergenzen zwischen den beiden Forschungskulturen der akademischen Sozialforschung und den Methoden der Marktforschung beschreibt Ziegler (2014) detailliert.

Aus der Betrachtung dieser Vorbehalte ergibt sich, dass die Techniken der Nutzungsdatenanalyse nur insoweit für die evaluative Wissenschaftskommunikationsforschung Anwendung finden sollten, als sie ethische Grundsätze, insbesondere die des Datenschutzes, die Standards der Evaluationspraxis im Hinblick auf Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit (siehe auch van den Bogaert in diesem Band), und – ganz allgemein – die Grundsätze der guten wissenschaftlichen Praxis nicht verletzen (DFG 2019).

Wenn in diesem Artikel von Nutzungsdatenanalyse die Rede ist, ist damit ausdrücklich nicht das *Data Scraping* gemeint, wie es z. B. von Batrinca und

Trelaeven (2015) beschrieben wird: Ein Verfahren, das Webseiten von Dritten oder ganze soziale Netzwerke in den Blick nimmt und das häufig eine Rolle für die Inhaltsanalyse dieser Seiten spielt. Nutzungsdatenanalyse im Sinne dieses Artikels schließt auch andere, sogenannte Offsite-Tools aus, die ein Wissenschaftskommunikationsprojekt aus einer Außenperspektive heraus untersuchen.

Außerdem soll es nicht um das sogenannte *Web Tracking* gehen, das versucht, mit Techniken wie dem *Browser Fingerprinting* (Boda et al. 2012) ein möglichst detailliertes Bild einer einzelnen Person zu zeichnen, indem ihre Aktivitäten beim Besuch vieler verschiedener Webseiten und über einen möglichst langen Zeitraum hinweg aufgezeichnet und ausgewertet werden, um z. B. ihr Konsumverhalten möglichst präzise vorhersagen zu können. Insbesondere aufgrund solcher – datenschutzrechtlich eher zweifelhafter – Praktiken steht die Nutzungsdatenanalyse häufig in der Kritik.

Der vorliegende Beitrag begrenzt sich auf Instrumente der sogenannten *Reichweitenanalyse* (genauer beschrieben durch Fritz 2004), die versuchen, aus der Perspektive eines einzelnen digitalen Angebots heraus Informationen über dessen Interaktionen mit Besucher:innen und das technische Funktionieren dieses Angebots zu erlangen.

Eine Reichweitenanalyse kann Aufschluss darüber geben, wie viele Besucher:innen ein Angebot pro Tag hat, darüber, was die populärsten Episoden, Unterseiten oder Beiträge eines Projekts der digitalen Wissenschaftskommunikation sind, wie viel Zeit Besucher:innen mit dem Angebot verbringen, wie sie zum Projekt gefunden haben, welche Sprache sie sprechen und vieles mehr (Clifton 2012). Die Reichweitenanalyse bietet die Basis für evaluative Fragestellungen darüber, welche Teile eines Angebots gut funktionieren oder welche das meiste Verbesserungspotenzial bieten, und spielt ihre Stärken insbesondere dann aus, wenn sie mit anderen Methoden kombiniert wird, wie bereits durch Welker und Wunsch (2010) vorgeschlagen. So wären beispielsweise Fragestellungen zum Zusammenhang der sprachlichen Komplexität einzelner Textbeiträge mit ihrer Nutzungshäufigkeit in spezifischen Zielgruppen denkbar.

Die folgenden Unterkapitel beschreiben, wo und wie Nutzungsdaten anfallen, was für und was gegen eine Anwendung von Werkzeugen der Nutzungsdatenanalyse in der evaluativen Wissenschaftskommunikationsforschung spricht und was es bei der Planung eines digitalen Wissenschaftskommunikationsprojekts zu beachten gilt, wenn die Nutzungsdatenanalyse eine Rolle bei seiner evaluativen Betrachtung spielen soll.

3 Was sind Nutzungsdaten?

Nutzungsdaten fallen durch die Nutzung eines digitalen Angebots, z. B. beim Besuch einer Internetseite, an. Dies können zum einen *Zielgruppendaten* sein, also solche, die Eigenschaften der Nutzer:innen einer digitalen Wissenschaftskommunikationsmaßnahme beschreiben, zweitens *Aktivitätsdaten*, also solche, die die Aktivitäten der Nutzer:innen während ihres Besuchs betreffen oder drittens *technische Daten*, die Aufschluss über technische Aspekte eines digitalen Angebots geben (Guba und Gebert 1998).

Beispiele für *Zielgruppendaten* im Rahmen einer Reichweitenanalyse eines Internetangebots sind Angaben über die geografische Region, aus der ein Zugriff erfolgt, die im Webbrowser eingestellte Sprache und darüber, mit welchem Endgerät, unter Verwendung welchen Betriebssystems, welchen Browsers und in welcher Bildschirmauflösung auf ein Angebot zugegriffen wurde.

Aktivitätsdaten beinhalten die Dauer eines Besuchs, nach welchen Begriffen auf einer Seite gesucht wurde, welche Wege die Besucher:innen durch die Webseite genommen haben (Funnel Analytics), über welchen Weg und zu welcher Uhrzeit Besucher:innen zum Angebot fanden oder an welcher Stelle ein Angebot wieder verlassen wurde.

Technische Daten umfassen die Anzahl der Besuche einer Internetseite innerhalb eines bestimmten Zeitraums und die Information darüber, welche Unterseiten wie häufig besucht wurden, die Zeit, die eine Webseite im Schnitt benötigt, um sich aufzubauen, die Häufigkeit und Typen bestimmter Server-Fehlermeldungen (z. B. „404 – Seite nicht gefunden“). Außerdem zählen Daten, die über die Auslastung des Servers in Bezug auf dessen benötigte Rechenleistung oder Serverkapazität Aufschluss gewähren, zu den technischen Daten.

Nutzungsdaten fallen entweder über die Anwendung sogenannter *Page Tags* – meist gestützt durch die Verwendung von Cookies – im Browser der Nutzer:innen oder als Log-Dateien direkt auf einem Webserver an. Ein Eintrag in solch einem *Server Log*, genauer spezifiziert in einem Working Draft des World Wide Web Consortiums (Hallam-Baker und Behlendorf 1996), kann folgendermaßen aussehen:

```
94.130.145.107 - - [20/Jan/2022:15:37:21 +0100] "GET
/transkripte/Beispiel.pdf HTTP/1.1" 206 16384
"https://www.wissenschaftskommunikation.de/"
"Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7)
AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.2
Safari/605.1.15"
"wissenschaftskommunikationsevaluationsforschung.de"
```

Der Eintrag besteht aus den Komponenten *IP-Adresse des Clients* (des/der Besucher:in) [94.130.145.107]; *Identität*, findet in der Praxis meist keine Anwendung [-]; *Benutzername*, falls die abgerufene Datei durch ein Passwort geschützt ist [-]; *Zeitstempel* [20/Jan/2022:15:37:21+0100]; *Anfrage des Clients*, untergliedert in Methode, Dateiname und verwendetes Protokoll [GET/transkripte/Beispiel.pdf HTTP/1.1]; *HTTP-Statuscode* der Server-Antwort [206]; *Dateigröße* der ausgelieferten Datei (in Bytes) [16384]; *HTTP-Referrer* (die Website, über die ein:e Besucher:in zur aktuellen Seite gekommen ist); *User Agent*, also Informationen über den für den Abruf verwendeten Browser und das Betriebssystem [Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/605.1.15 (KHTML, like Gecko) Version/15.2 Safari/605.1.15]; *Abgerufenes Objekt* [wissenschaftskommunikationsevaluationsforschung.de].

Log-Einträge in dieser oder ähnlicher Form legt jede Webserver-Software automatisch an. Die Einträge können entweder direkt verarbeitet oder durch eine Analyse-Software mit Page Tags oder anderen Daten kombiniert, aufbereitet und miteinander verknüpft werden¹ und bieten so einen schnellen Einblick in komplexere Zusammenhänge – denkbar ist z. B. ein Vergleich der Anzahl der Abrufe mehrerer Podcast-Episoden in den ersten zwei Monaten nach ihrer jeweiligen Veröffentlichung. Je nach genutztem Medium können aber auch vollkommen andere Datentypen und Visualisierungsarten zur Verfügung stehen, z. B. so genannte Heatmaps, die Aufschluss über häufig geklickte Bereiche auf einer Website geben oder Daten darüber, welche Web-Videos zu besonders vielen Abonnements des eigenen Kanals geführt haben. Eine genauere Analyse der Vor- und Nachteile des Zusammenwirkens von Page Tags, Cookies, Server Logs sowie anderer, weniger verbreiteter Verfahren liefert Clifton (2012).

Nachfolgend wird ein Überblick über die wichtigsten Typen von Nutzungsdaten (auch: Metriken) gegeben, die in den unterschiedlichen Medien der digitalen Wissenschaftskommunikation anfallen können (Tab. 1):

Abhängig von der Fragestellung oder dem Forschungsinteresse kann es entweder sinnvoll sein, einige dieser Werte miteinander zu verknüpfen oder sich auf nur auf einzelne Werte zu konzentrieren. Üblicherweise werden in den gängigen Analyseinstrumenten Besuchszeiten, Wege durch die Webseite oder Akquisitionsdaten, die Auskunft über die Website geben, von der ein:e Besucher:in kommt, visualisiert.

¹Üblich ist z. B. eine Standortdatenbank, die Geolocations zu IP-Adressen bereithält.

Tab. 1 Typen von Nutzungsdaten mit Relevanz für digitale Wissenschaftskommunikationsprojekte**Wissenschaftskommunikationsprojekte auf Web-Basis** (Webseiten, Web-Apps, Blogs, Podcasts)

Anzahl der Seitenbesuche, Downloads bzw. Hörintentionen, aufgerufene URL, Seitentitel, Aufenthaltsdauer bzw. geschätzte Wiedergabezeit, Zugriffsquellen (der Weg, über den Besucher:innen zum Inhalt gelangen, ob über andere Websites, durch Suchen, durch Social Media etc.), Geolocations, Datum und Uhrzeit des Aufrufs, Gerätetyp, -modell, -hersteller, Bildschirmauflösung und verwendetes Betriebssystem, Absprungrate, Aktionen pro Besuch, Anzahl und Inhalte der Suchen, Anzahl, Inhalte und Dateigrößen von Downloads, Ort der Wiedergabe (Download, RSS-Feed oder Podcast-Web-Player)

Daraus abgeleitet: Der Weg durch die Webseite;

Bedingt (im Falle von externem Hosting): Andere Datentypen, z. B. Blogs oder Podcasts, die die Benutzer:innen sehen, lesen oder hören

E-Mail-Newsletter:

Anzahl Empfänger:innen, Öffnungsrate, Datum und Uhrzeit der Öffnung der E-Mail, Klicks auf Links, Rückläufer, Abmeldungen, Geolocations, bedingt: Gerätetyp

Webvideos:

Anzahl der Aufrufe, Gesamtwiedergabezeit, Zugriffsquellen, Geolocations, Datum und Uhrzeit des Aufrufs, Gerätetyp und verwendetes Betriebssystem, Ort der Wiedergabe (direkt oder eingebettet), bedingt: Untertiteltyp, bedingt (nur bei externem Hosting): Alter und Geschlecht der Zuschauer:innen, andere Kanäle und Videos, die sich die Zuschauer:innen angesehen haben

Twitter (bezogen jeweils auf ein einzelnes Posting):

Veröffentlichungszeitpunkt, Impressions (Sichtkontakte), Interaktionen, Interaktionsrate, Retweets, Antworten, „Gefällt mir“-Angaben, Nutzerprofilklicks, URL-Klicks, Hashtag-Klicks, Detailerweiterungen, Permalink-Klicks, Follows, Medienanzeigen, Medieninteraktionen

Facebook (bezogen jeweils auf ein einzelnes Posting):

Veröffentlichungszeitpunkt, Reichweite, „Gefällt mir“-Angaben und Reaktionen, Antworten, Link-Klicks, Kommentare, Geteilte Inhalte, grobe Altersverteilung und Wohnorte der Zielgruppe

Instagram (bezogen jeweils auf ein einzelnes Posting):

Profilaufrufe, Reichweite, Interaktionen, neue Abonnent:innen, Impressions, „Gefällt mir“-Angaben, Anzahl Kommentare, Geteilte Inhalte, Favorisierte Inhalte

4 Warum Nutzungsdatenanalyse (in der evaluativen Wissenschaftskommunikationsforschung) betreiben?

Bereits in der Planungsphase eines Wissenschaftskommunikationsprojekts stellt sich häufig die Frage nach den Zielen und den Zielgruppen eines Projekts, und es kommen Fragen danach auf, welche Evaluationsmethode sich eignet, um das Erreichen dieser Ziele zu überprüfen. Eine Übersicht über mögliche Motive und Ziele in der Wissenschaftskommunikation und konkret in Hinblick auf deren Evaluation geben Ziegler und Fischer (2020).

Bei *digitalen* Wissenschaftskommunikationsprojekten kann bereits ein regelmäßiger Blick auf die Nutzungsdatenauswertung dabei helfen, das eigene Projekt zu reflektieren sowie ein Gespür für seine Reichweite, insbesondere für seinen Output zu bekommen. Eine Frage, die sich an jedes Wissenschaftskommunikationsprojekt stellen lässt, ist die danach, ob das Projekt für die eigene Zielgruppe optimal erreichbar ist. Hier kann z. B. ein Web-Analysetool wie *Matomo* (ehemals *Piwik*) dabei helfen, Schwachstellen und Fehler in der Webseite des Projekts zu identifizieren, oder einzelne Unterseiten ausfindig zu machen, die langsam oder überhaupt nicht geladen werden können. Zugangsbarrieren bei der Erreichbarkeit der Seite müssen dabei nicht zwangsläufig technischer Natur sein, auch ein zu komplexes oder verwirrendes User-Interface kann durch Nutzungsdatenanalyse als Fehlerquelle entlarvt werden. Eine kurzweilige und leicht verständliche Einführung ins User-Interface-Design bietet Steve Krug in „Don’t make me think!“ (2014). Je nachdem, welche Beweggründe einem Evaluationsvorhaben zugrunde liegen, liefert die Nutzungsdatenanalyse einen Beitrag zum Erkenntnisgewinn. Mögliche Beweggründe sind der Wunsch danach, mehr Wissen über das Wissenschaftskommunikationsprojekt zu erlangen, Optimierungspotenziale aufzudecken oder Ergebnisse und Wirkungen des Projekts zu prüfen (siehe Impact Unit 2021). Je nach Evaluationsdesign kann die Nutzungsdatenanalyse sowohl als Komponente einer formativen Evaluation oder einer summativen Evaluation dienen (siehe auch Volk in diesem Band). Insbesondere für Evaluationsdesigns, deren Ziel es ist, ein Projekt noch während seiner Laufzeit im Prozess zu verbessern, bietet sich die Nutzungsdatenanalyse an.

Stellen wir uns folgendes Beispiel vor: Die anvisierte Zielgruppe einer naturwissenschaftlich ausgerichteten YouTube-Video-Reihe ist Schüler:innen zwischen 13 und 18 Jahren. Schon jeweils kurz nach Veröffentlichung der ersten drei

Videos stellt sich beim Blick in die Statistiken heraus, dass fast ausschließlich Zuschauer:innen über 65 Jahre erreicht werden. Dadurch bietet sich die Gelegenheit, im Produktionsprozess weiterer Videos frühzeitig gegenzusteuern – oder das Projekt und seine Zielsetzung zu überdenken.

Viele Logfile-Analysen haben zum Ziel, Typen von Nutzer:innen und verschiedene Strategien, mit denen diese Nutzer:innen versuchen, bestimmte Informationen zu erlangen, zu identifizieren und sie mit weiteren Variablen oder bestimmten Konstrukten wie Lernerfolgen, die in Befragungen erhoben werden, in Zusammenhang zu bringen (Priemer 2004). Innerhalb der Wirkungsforschung sind beispielsweise Fragestellungen dazu vorstellbar, inwiefern sich das Bild von Wissenschaft im Allgemeinen oder von einer Disziplin im Speziellen für Nutzer:innen im Laufe der Zeit ändert, wenn diese regelmäßig bestimmte Podcasts hören oder einen Blog lesen. Denkbar wäre zur Annäherung an diese Fragestellung eine Kombination von Nutzungsdatenanalyse mit qualitativen und quantitativen Befragungsinstrumenten. Auch Fragen nach dem spezifischen Interesse von Besucher:innen als Grund für den Besuch eines Angebots lassen sich unter Zuhilfenahme der Nutzungsdatenanalyse stellen – indem man z. B. verwendete Suchbegriffe betrachtet.

Fagan (2013) beschreibt, wie die Adaption von bereits etablierten Leistungskennzahlen (*Key Performance Indicators, KPIs*) aus dem Marketing für das akademische Bibliothekswesen fruchtbar gemacht werden kann. Besonders im Hinblick auf jene *KPIs*, die das *User Engagement* betreffen, könnte eine solche Adaption auch auf Fragestellungen der Wissenschaftskommunikationsforschung anwendbar sein, z. B. auf solche, anhand derer die Einbindung und die Beteiligung von Bürger:innen an Forschungsprozessen untersucht werden sollen.

Nutzungsdatenanalyse kann für die Evaluation von Wissenschaftskommunikation aber nicht nur eine Methode zur Generierung von Forschungsdaten sein und somit zur Verbesserung einer Wissenschaftskommunikationsmaßnahme beitragen. Bereits das bloße Nachdenken über die Möglichkeiten der Nutzungsdatenanalyse kann bei der Generierung neuer Forschungsfragen unterstützen.

Im Praxisbeitrag zur Nutzungsdatenanalyse in diesem Band (siehe auch Buckermann und Greving in diesem Band) wird die Untersuchung von Beteiligungsmustern an einem Bürgerwissenschaftsprojekt beschrieben, mit dem Ziel, Aussagen über die Motivation der Teilnehmenden treffen zu können. Voraussetzung für das Stellen einer solchen Forschungsfrage ist die Kenntnis über die Möglichkeiten der Reichweitenanalyse, speziell der Auswertung der Aktivitätsdaten von Webseiten-Besucher:innen.

Eines der gewichtigsten Argumente für die Erwägung von Nutzungsdatenanalyse im Rahmen der Evaluation von Wissenschaftskommunikationsmaßnahmen ist aber der geringe Aufwand, der mit ihrer Durchführung verbunden ist. Bei beinahe jedem Vorhaben der digitalen Wissenschaftskommunikation stehen Nutzungsdaten mehr oder weniger automatisch zur Verfügung – ohne aufseiten der Forscher:innen oder der Nutzer:innen einen zusätzlichen Zeitaufwand zu erzeugen. Die Erhebung von Nutzungsdaten ist meist leicht einzurichten, wenig aufwendig während der Durchführung und außerdem kostengünstig zu betreiben. Nutzungsdatenanalyse erfordert kein Experimentaldesign, keinen Entwurf von Fragebögen und kann – einmal eingerichtet – automatisiert objektive, eindeutige, vollständige und detailgenaue Daten (Priemer 2004) über die komplette Projektlaufzeit hinweg liefern.

Darüber hinaus ist sie nicht-invasiv, sie erfordert vonseiten der Nutzer:innen eines Angebots der digitalen Wissenschaftskommunikation keinerlei Ressourcen, weder deren Zeit noch ihre besondere Aufmerksamkeit – Nutzungsdaten werden „nebenbei“ erhoben, während die Besucher:innen eines Angebots dieses nutzen – ohne, dass sie gestört werden.

Eine der Herausforderungen bei der Anwendung der Nutzungsdatenanalyse als Instrument der evaluativen Wissenschaftskommunikationsforschung ist, dass es in manchen Fällen sehr schwierig sein kann, die erforderlichen Daten zu erlangen. Dies gilt insbesondere für Daten, die aus den proprietären Analysewerkzeugen sozialer Netzwerke gewonnen werden müssen. Außerdem wird ein nicht unerheblicher Teil von Server-Log-Dateien durch automatisierten Traffic (Suchmaschinen-Bots etc.) erzeugt und muss häufig herausgefiltert werden, da mittels dieser Daten keine Erkenntnisse über das Verhalten von Nutzer:innen gewonnen werden können. Ohne eine Kombination mit anderen Methoden, wie z. B. mit Befragungen oder Inhaltsanalysen, sind die Ergebnisse der Nutzungsdatenanalyse nur bedingt aussagekräftig (Priemer 2004). Nicht zuletzt können auch datenschutzrechtliche Erwägungen, insbesondere jene mit Blick auf die 2016 eingeführte europäische Datenschutzgrundverordnung (DSGVO)², die Möglichkeiten und die Aussagekraft von Nutzungsdatenanalyse einschränken.

²Eine ausführlichere Aufschlüsselung der Herausforderungen datenschutzrechtlicher Erwägungen im konkreten Projektkontext in Abschn. 5.3.

5 Was ist bei der Gestaltung eines Wissenschaftskommunikationsprojekts zu beachten?

Grundsätzlich bieten alle digitalen Wissenschaftskommunikationsprojekte die Möglichkeit, Nutzungsdatenanalyse zu betreiben. Das gewählte Medium und die konkrete Ausgestaltung eines Projekts bestimmen dabei die Herangehensweise, die Daten zu erheben und beeinflussen damit auch die Forschungsfragen, die gestellt werden können.

5.1 Einfluss des gewählten Mediums auf Möglichkeiten der Nutzungsdatenanalyse

Das gewählte Medium hat einen großen Einfluss darauf, welche Typen von Daten auf welche Art und Weise erhoben werden können. Während für Websites, Web-Apps, Blogs, E-Mail-Newsletter-Systeme und Podcasts, also für Medien, die grundsätzlich durch einen eigenen Webserver bereitgestellt werden können, eine breite Palette von Analysewerkzeugen und -techniken verfügbar ist, ist der Zugriff auf die Nutzerdaten plattformbasierter oder sozialer Medien wie Twitter, Instagram, Facebook, Whatsapp oder YouTube sowie bei nativen Smartphone-Apps immer auf die Analysewerkzeuge begrenzt, die die Anbieter:innen dieser Plattformen und App-Stores zur Verfügung stellen.

Die qualitativen Unterschiede zwischen diesen Werkzeugen sind hoch. Twitter Analytics beispielsweise ist ein leicht zugängliches und klar strukturiertes Analysewerkzeug, das immerhin einen direkten Datendownload im csv-Format erlaubt. Ähnlich – wenn auch eingeschränkter – verhält es sich bei Facebook Insights. Bei Instagram hingegen ist der Insights-Bereich nur aus der mobilen App heraus und nur für spezielle Business- oder Creator-Accounts zugänglich, auch ein direkter Datendownload ist nicht möglich. Die Nutzungsdaten manch anderer sozialer Netzwerke sind sogar gar nicht oder nur dann zugänglich, wenn eine spezielle Drittanbieter-Software für die Datenauswertung genutzt wird. Nichtsdestotrotz können sich auch Daten aus schwer zugänglichen Quellen für eine Nutzungsdatenanalyse in der evaluativen Wissenschaftskommunikationsforschung eignen: Beispielsweise stützen Essig et al. (2020) ihre Analyse eines Instagram-Accounts als geeignetes Mittel, um histologisch Befunde für Medizinstudent:innen zugänglich zu machen einerseits auf Befragungen, andererseits aber auch auf erhobene Nutzungsdaten.

5.2 Einfluss des Hosting-Providers auf die Möglichkeiten der Nutzungsdatenanalyse

Während der Planung eines digitalen Wissenschaftskommunikationsprojektes stellt sich häufig die Frage danach, wie und durch wen die produzierten Inhalte zur Verfügung gestellt werden sollen; ob es beispielsweise sinnvoller ist, sie über einen eigenen Webserver oder durch einen externen Blog-, Webvideo- oder Podcast-Hosting-Anbieter bereitzustellen. Bei der Entscheidung darüber können technische, datenschutzrechtliche oder Erwägungen zur Barrierefreiheit und Zugänglichkeit eine Rolle spielen. Ebenso stellen sich Fragen danach, wie aufwendig ein eigenes Hosting einzurichten und zu pflegen ist, oder welche Kosten bei der Nutzung einer externen Plattform entstehen.

Auch Fragen nach der Verfügbarkeit, der jeweiligen Aufbereitung und den Downloadmöglichkeiten von Nutzungsdaten müssen abgewogen werden. Wird ein eigener Server für das Hosting des Wissenschaftskommunikationsprojektes genutzt, bringt das einerseits den Vorteil mit sich, dass der unbeschränkte und wiederholte Zugriff auf die Server-Logdateien möglich ist. Darüber hinaus besteht die Möglichkeit, unter vielen verschiedenen Datenanalyse-Tools zu wählen. Auch diese Tools lassen sich dann entweder auf dem eigenen Webserver installieren – beispielsweise *AWStats* oder *Webalizer* – oder als Software-as-a-Service bei einem Drittanbieter einkaufen, wie es z. B. für den Einsatz von *Google Analytics* zutrifft. Eine selbstgehostete Analyse-Software läuft meist auf demselben Webserver wie die zu analysierende Website, sei es als Plugin im verwendeten Content-Management-System oder als separate, eigenständige Software. Die freie Analyse-Software *Matomo* beispielsweise bietet mehr als 100 Plugins, die zur Vereinfachung ihrer Einbindung in verschiedenste Content-Management-Systeme dienen. Ansätze systematischer Reviews der verbreitetsten Webanalyse-Tools finden sich in Bekavac und Garbin Praničević (2015) sowie in Kumar und Ogunmola (2020).

Wird dagegen ein digitales Wissenschaftskommunikationsprojekt von einem externen Anbieter gehostet, geht dadurch die Entscheidungshoheit darüber verloren, welche Nutzungsdaten erhoben werden und in welcher Art diese Daten dargestellt werden. Die proprietären Analyse-Werkzeuge der Hosting-Anbieter sind häufig in ihrem Datenzugriff, in ihren Analysefähigkeiten sowie in der Weise, in der die Daten visualisiert werden, limitiert. Sie können aber – bei allen Einschränkungen – auch Einblicke bieten, die mit einer selbstgehosteten Analyse-Software nicht zu erlangen sind. Einige Podcast-Hosts stellen z. B. Informationen darüber zur Verfügung, welche anderen Podcasts bzw. Podcast-Kategorien

Hörer:innen des eigenen Podcasts hören oder liefern eine minutengenaue Übersicht über die Anzahl der Hörer:innen pro Episode. Der Video-Host *YouTube* bietet eine Übersicht über Alter und Geschlecht der Zuschauer:innen oder die Uhrzeit, zu der sie üblicherweise Videos schauen. Das sind Informationen, die sich weder aus einer Auswertung der Server-Log-Dateien noch über JavaScript-Tracking ergeben können. Diese Daten sind nur dann zu erlangen, wenn ein Hosting-Dienst auf ergänzende Daten zurückgreift, über die ausschließlich dieser selbst verfügt. Dies können persönliche Daten wie z. B. Alter und Geschlecht oder Nutzungsdaten der Endgeräte sein, mit denen Podcasts gehört oder Videos gesehen werden. Im Podcast-Bereich erscheint das sogenannte *Client-Side-Tracking*, also das Einbeziehen von Nutzungsdaten, die im Podcatcher (also direkt auf dem jeweiligen Abspielgerät), erhoben werden, präzisere Daten zu versprechen, es ist allerdings noch nicht sehr verbreitet (Podigee 2019).

Zusammenfassend lässt sich also sagen, dass es sich aus der Perspektive der evaluativen Wissenschaftskommunikationsforschung eher empfiehlt, ein digitales Wissenschaftskommunikationsprojekt unter Verzicht auf externe Anbieter:innen zu hosten, um in der Wahl der Analysewerkzeuge flexibel zu bleiben und einen direkten Zugriff auf Server-Log-Dateien zu haben. Nur in Ausnahmefällen sind Nutzungsdaten, die ausschließlich durch externe Hosting-Dienste angeboten werden, so interessant, dass ein Fremdhosting dem eigenen Server vorzuziehen ist. Sollte trotzdem ein solches Fremdhosting in Betracht gezogen werden, kann es hilfreich sein, den Hosting-Anbieter um die regelmäßige Bereitstellung der Server-Log-Dateien für das eigene Angebot zu bitten.

5.3 Erwägungen zum Datenschutz

Nutzungsdaten sind häufig personenbezogene Daten, die zur Identifikation oder Standortbestimmung einzelner natürlicher Personen genutzt werden können. Auch, wenn die Nutzungsdatenanalyse zum Zwecke der wissenschaftlichen Evaluation oder der Verbesserung einer Wissenschaftskommunikationsmaßnahme durchaus häufig als „wissenschaftlicher Forschungszweck“ nach Artikel 9 Abs. 2 der DSGVO beurteilt werden kann, ist es erforderlich, nur so viele personenbezogene Daten wie unbedingt nötig zu erheben. Weiterhin muss sichergestellt sein, dass nicht mehr Personen als unbedingt nötig Zugriff auf die erhobenen Daten erhalten und dass diese Daten nicht länger als nötig gespeichert werden. Darüber hinaus ist es ratsam, in der eigenen Datenschutzerklärung so transparent wie möglich darüber zu sein, welche Daten zu welchen Zwecken gesammelt werden und wie lange diese Daten gespeichert werden. Auch die Frage danach,

ob eine Datenschutz-Folgenabschätzung oder ein Datenschutzkonzept erstellt werden muss, sollte in der Planungsphase eines Projekts immer mit dem/der zuständigen Datenschutzbeauftragten besprochen werden.

Proprietäre (nichtsdestotrotz populäre) Webanalyse-Tools von Anbietern wie Google oder Adobe werden in Bezug auf den Datenschutz häufig kritisch gesehen – unter anderem, weil bei ihrer Nutzung nicht zweifelsfrei geklärt werden kann, auf welchen Servern die Daten verarbeitet werden (Gamalielsson et al. 2021).

Um solche Konflikte aus dem Weg zu gehen, empfiehlt sich die Verwendung eines quelloffenen und datenschutzkonformen Webanalysewerkzeugs wie beispielsweise *Matomo Analytics*. In der Studie, die dem Praxisbeitrag zur Nutzungsdatenanalyse (siehe auch Buckermann und Greving in diesem Band) zugrunde liegt, wurde diese Software verwendet. *Matomo* gilt nicht nur als datenschutzkonforme Open-Source-Alternative zu proprietärer Analyse-Software. Auch das *Data Ownership* ohne Ausnahme, das insbesondere bei Citizen-Science-Projekten zunehmend an Bedeutung gewinnt, ist eine der Kernfunktionalitäten von *Matomo*. Auch deswegen eignet sich diese Software als Instrument für die evaluative Wissenschaftskommunikationsforschung.

Die DSGVO und die dadurch begrenzte Verwendbarkeit von Analyse-Werkzeugen, die auf Browser-Cookies basieren, schränken die Anwendungsfelder von Nutzungsdatenanalysemethoden ein. Grundlegende Abrufzahlen sowie durch Webseiten-Besucher:innen verwendete Suchbegriffe sind nach wie vor problemlos zu erhalten; jedoch die Information darüber, ob ein:e Besucher:in ein Angebot mehrfach besucht hat, ist nicht mehr ohne das Vorschalten eines sogenannten *Cookie-Banners* möglich, durch das das Einverständnis ein:e:r Besucher:in zur Erfassung dieser Daten eingeholt werden muss. Verzichtet man jedoch auf die Erfassung von Informationen, für deren Erhebung die Installation von Browser-Cookies notwendig ist, und anonymisiert andere personenbezogene Daten wie Besucher:innen-IP-Adressen, ist die Vorschaltung eines solchen Banners unter Umständen nicht nötig (Matomo 2021).

6 Fazit

Nutzungsdatenanalyse kann dabei unterstützen, mehr über die eigenen Wissenschaftskommunikationsprojekte zu erfahren, sie zu hinterfragen und sie zu verbessern. Sie hilft dabei, technische Fehlerquellen zu identifizieren und Barrierefreiheit zu gewährleisten. Ihre Werkzeuge können darüber hinaus vor allem für Fragestellungen, die auf das Engagement der Benutzer:innen in partizipativen Formaten abzielen, interessant sein. Dafür bietet sich häufig an,

Methoden der Nutzungsdatenanalyse mit anderen quantitativen oder qualitativen Methoden und Data-Scraping-Techniken zu kombinieren.

In der institutionellen Wissenschaftskommunikation ist es üblich, dass einzelne Institutionen mehrere Wissenschaftskommunikationsprojekte durchführen. In diesen Fällen empfiehlt es sich, mittels Web Analytics im Auge zu behalten, wie viele Besucher:innen durch welche Projekte der Wissenschaftskommunikation auf die Seite der Institution finden. So können z. B. Entscheidungen darüber, welche Ressourcen zukünftig auf welches Projekt verwendet werden, informierter getroffen werden. Hier ergibt sich jedoch zugleich eine wichtige Einschränkung: Kennzahlen der Nutzungsdatenanalyse allein lassen keine Aussagen über den quantitativen (Mehr-) Wert einzelner Maßnahmen oder gar ganzer Projekte zu. Ganz generell erscheint es sinnvoll, sich vor oder während der Einrichtung eines Wissenschaftskommunikationsprojekts einen Auszug der verfügbaren Analysedaten anzusehen, um sich mit den verschiedenen Metriken, die potenziell zur Verfügung stehen, vertrauter machen zu können. Es gilt daher basierend auf dem Gebot der Datensparsamkeit, nur diejenigen Daten, die für die Beantwortung konkreter Forschungs- oder Evaluationsfragen notwendig sind, zu erheben. Nach Formulierung konkreter Ziele und Forschungsfragen für das Projekt sollten überflüssige Metriken verworfen werden.

Trotz der Einführung der DSGVO und generell strengerer Richtlinien im Umgang mit Forschungs- wie persönlichen Daten können Nutzungsdatenanalyse-Methoden mit verhältnismäßig geringem Personal- oder Sachaufwand durchgeführt werden. Wie Nutzungsdatenanalyse datenschutzkonform und datensparsam durchgeführt werden kann, wird im Einzelnen bei Karg und Thomsen (2011) und Matomo (2021) beschrieben. Im Zweifel sollte das Vorgehen gemeinsam mit dem/der zuständigen Datenschutzbeauftragten koordiniert werden.

Für die evaluative Forschung an Wissenschaftskommunikationsprojekten gilt es, die Nutzungsdatenanalyse als Methodenset weiter zu elaborieren. Neben einer Schärfung der Instrumente bedarf es der weiteren Auseinandersetzung mit ihren Stärken und Schwächen, um methodischen, ethischen und praktischen Herausforderungen der Ansätze zu begegnen. Angesichts des rasanten Wachstums sowie des steten Wandels digitaler Kommunikationsformate sollten sich Praktiker:innen und Theoretiker:innen gleichermaßen mit der Nutzungsdatenanalyse befassen – um so auch zukünftig neue, digitale Wissenschaftskommunikation zu ihrem Betrachtungsgegenstand machen zu können.

Literatur

- Batrinca B, Treleaven PC (2015) Social media analytics: a survey of techniques, tools and platforms. *AI & Soc* 30:89–116. <https://doi.org/10.1007/s00146-014-0549-4>
- Bekavac I, Garbin Praničević D (2015) Web analytics tools and web metrics tools: an overview and comparative analysis. *Croat Oper Res Rev* 6 (2), 373–386. <https://doi.org/10.17535/crorr.2015.0029>
- Boda K, Földes ÁM, Gulyás GG, Imre S (2012) User tracking on the web via cross-browser fingerprinting. In: Laud P (Hrsg) *Information security technology for applications*. NordSec 2011. Lecture notes in computer science 7161. Springer, Berlin. https://doi.org/10.1007/978-3-642-29615-4_4
- Clifton B (2012) *Advanced web metrics with google analytics*, 3. Aufl. Sybex
- Deutsche Forschungsgemeinschaft [DFG] (2019) Guidelines for safeguarding good research practice. Code of conduct. <https://doi.org/10.5281/zenodo.3923602>
- Deutsche Forschungsgemeinschaft [DFG] (2021) Fachspezifische Empfehlungen zum Umgang mit Forschungsdaten. https://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/forschungsdaten/empfehlungen/index.html. Zugegriffen: 30. Jan. 2022
- Essig J, Watts M, Beck Dallaghan GL, Gilliland KO (2020) InstaHisto: utilizing Instagram as a medium for disseminating visual educational resources. *Med Sci Educ* 30(3):1035–1042. <https://doi.org/10.1007/s40670-020-01010-2>
- Fagan JC (2013) The suitability of web analytics key performance indicators in the academic library environment. *J Acad Librariansh* 40(1):25–34. <https://doi.org/10.1016/j.acalib.2013.06.005>
- Fritz W (2004) *Internet-Marketing und Electronic Commerce. Grundlagen – Rahmenbedingungen – Instrumente*. Gabler, Springer Fachmedien Wiesbaden GmbH, Wiesbaden. <https://doi.org/10.1007/978-3-663-06001-7>
- Gamalielsson J, Lundell B, Butler S, Brax C, Persson T, Mattsson A, Gustavsson T, Feist J, Lönroth E (2021) Towards open government through open source software for web analytics: the case of Matomo. *JeDEM EJ EDemocr Open Govt* 13(2):133–153. <https://doi.org/10.29379/jedem.v13i2.650>
- Guba A, Gebert O (1998) Online-Monitoring: Gewinnung und Verwertung von Online-Daten. In: *Arbeitspapiere WI*, Nr. 8/1998. Lehrstuhl für Allg BWL und Wirtschaftsinformatik (Hrsg) Johannes-Gutenberg-Universität, Mainz
- Hallam-Baker PM, Behlendorf Brian (1996) Extended log file format. <https://www.w3.org/TR/WD-logfile.html>. Zugegriffen: 30. Jan. 2022
- Karg M, Thomsen S (2011) Einsatz von Piwik bei der Reichweitenanalyse. *DuD* 35(489). <https://doi.org/10.1007/s11623-011-0120-0>
- Kumar V, Ogunmola GA (2020) Web analytics for knowledge creation: a systematic review of tools, techniques, and practices. *Int J Cyber Behav Psychol Learn* 10(1):1–14. <https://doi.org/10.4018/IJCBPL.2020010101>
- Krug S (2014) *Don't make me think, revisited: a common sense approach to Web usability*. New Riders Publishing, Thousand Oaks. <https://dl.acm.org/doi/10.5555/1051204>
- McFadden C (2005) *Optimizing the online business channel with web analytics*. <https://web.archive.org/web/20051221192230/> <http://www.webanalyticsassociation.org/80/en/art/?9>. Zugegriffen: 13. Febr. 2022

- Matomo (2021) How do I use Matomo Analytics without consent or cookie banner? <https://web.archive.org/web/20210824213215/https://matomo.org/faq/new-to-piwik/how-do-i-use-matomo-analytics-without-consent-or-cookie-banner/>. Zugegriffen: 13. Febr. 2022
- Neuberger C, Weingart P, Fähnrich B, Fecher B, Schäfer MS, Schmid-Petri H, Wagner GG (2021) Der digitale Wandel der Wissenschaftskommunikation. In: Wissenschaftspolitik im Dialog: eine Schriftenreihe der Berlin-Brandenburgischen Akademie der Wissenschaften (Hrsg) Berlin-Brandenburgische Akademie der Wissenschaften, Berlin
- Podigee (2019) Unified podcast analytics whitepaper. https://web.archive.org/web/20220116121720/https://docs.google.com/document/d/e/2PACX-1vQQcfYvUUsaR2UBPudmCpe5yNDs5HczB2ST9M6jEBw13eHZyn5TcFX_HiEVSBE9y7yTBdZzF8HbRDdn/pub. Zugegriffen: 30. Jan. 2022
- Priemer B (2004) Logfile-Analysen: Möglichkeiten und Grenzen ihrer Nutzung bei Untersuchungen zur Mensch-Maschine-Interaktion. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung. Occasional Papers, S 1–23. <https://doi.org/10.21240/mpaed/00/2004.06.02.X>
- Tandoc EC Jr, Thomas RJ (2015) The ethics of web analytics. Digit J 3(2):243–258. <https://doi.org/10.1080/21670811.2014.909122>
- Welker M, Wünsch C (2010) Methoden der Online-Forschung. In: Schweiger W, Beck K (Hrsg) Handbuch Online-Kommunikation. VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92437-3_20
- Ziegler M (2014) Marktforschung. In: Baur N, Blasius J (Hrsg) Handbuch Methoden der empirischen Sozialforschung, S 183–193
- Ziegler R, Fischer L (2020) Ziele von Wissenschaftskommunikation – Eine Analyse der strategischen Ziele relevanter Akteure für die institutionelle Wissenschaftskommunikation in Deutschland, 2014–2020. Wissenschaft im Dialog, Berlin

Armin Hempel ist wissenschaftlicher Mitarbeiter im Teilprojekt Öffentlichkeitsarbeit des DFG-SFBs 980 „Episteme in Bewegung“ an der Freien Universität Berlin. Er befasst sich mit digitalen Formaten, vorrangig mit Podcasts als Mittel der Wissenschaftskommunikation.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

