

# Beim Datentracking gleich noch die Nutzerstatistiken von Bibliotheken ausgewertet

Stephan Holländer

Ein großes Echo fand im vergangenen Jahr der in Heft 1/2022 von o-bib veröffentlichte Beitrag „Das Lesen der Anderen“ von Renke Siems<sup>1</sup> in unterschiedlichen Stellungnahmen von verschiedenen Seiten. Die Auswirkungen von User Tracking auf Bibliotheken wurden so breiteren Kreisen bekannt gemacht. Nun ist Datentracking schon lange ein lukratives Geschäftsmodell verschiedener Konzerne in vielen Branchen. Dass es auch in der Wissenschaft stattfindet, war dagegen bis letztes Jahr noch nicht sehr bekannt. Die Gefahren für Datenschutz und die Freiheit von Wissenschaft und Forschung durch Tracking sind erheblich. Auch Bibliotheken als Akteurinnen im wissenschaftlichen Ökosystem sind davon nicht ausgenommen, vor allem dann nicht, wenn sie in irgendeiner Form Verträge mit profitorientierten Informationsanbietern wie Hosts und Verlagen schließen, bei denen auch die Daten von Forschenden Teil des Datentransfers sein können.

Das ist problematisch, weil wissenschaftliche Bibliotheken heute mehr als nur Informationen und Medien vermitteln. Dank steigender Lizenzkosten für wissenschaftliche Fachartikel gelingt es Konzernen, für wissenschaftliche Information eine kommerzielle Parallelstruktur zu den Angeboten wissenschaftlicher Information aufzuziehen, die sich Bibliotheken aufgrund begrenzter finanzieller Budgets in diesem Umfang nicht leisten können.

So werden Forschungsergebnisse, die mit öffentlichen Mitteln finanziert wurden, dem öffentlichen Eigentum und den Universitäten entzogen und ein zweites Mal mit Gewinn über Lizenzen an die gleichen Zielgruppen verkauft. Damit wird seitens der kommerziellen Informationsanbieter eine gewinnorientierte Parallelstruktur geschaffen und wissenschaftliche Bibliotheken an der Wahrnehmung ihrer Kernaufgabe, der Vermittlung wissenschaftlicher Information, aufgrund ungleicher finanzieller Wettbewerbsbedingungen gehindert.

Digitale Medien nehmen einen immer größeren Anteil am Informationsangebot von Bibliotheken ein. Während der Pandemie hat die Nutzung digitaler Informationen eine weitere Zunahme erfahren. Dazu gehören auch die Umstellung auf digitale Fernlehre und die Entwicklung von Software zu digital unterstütztem Lesen und Schreiben sowie neu entwickelte Arbeitsmethoden wie kollaborati-



ves Schreiben und Textmining, um nur zwei zu nennen. Diese Entwicklungen werden sich auch danach nicht mehr ändern und ermöglichen den Informationsanbietern die Auswahl von wissenschaftlichen Fachartikeln und wissenschaftlichen E-Books, die Lesedauer, die Lesegeschwindigkeit, sowie das Markieren von Abschnitten zu steuern. Was mit den dabei erfassten Nutzerdaten durch die Informationsanbieter geschieht, ist hingegen nicht annähernd transparent.

So beschäftigen sich beispielsweise Clarivate und Elsevier schon länger nicht mehr nur mit dem Erstellen von wissenschaftlichen Datenbanken und der Veröffentlichung wissenschaftlicher Zeitschriften. Beide Anbieter bieten Zitationsdatenbanken an und bewerben Software-Werkzeuge zur Unterstützung des kompletten Forschungszyklus. Clarivate wurde für seine wettbewerbsfeindlichen Praktiken auf dem stark oligopolen Markt für akademische Datenbanken kritisiert, weil es die Preise für die von ihm angebotenen Produkte erhöht hat. Außerdem sind beide Konzerne Anbieter für Analysetools zur Forschungsbewertung und -evaluation für Wissenschaftler.

Der Europäische Gerichtshof (EuGH) hat am 16. Juli 2020 das EU-US Privacy Shield mit sofortiger Wirkung für ungültig erklärt (Urteil in der Rechtssache C-311/18 Data Protection Commissioner / Maximilian Schrems und Facebook

<sup>1</sup> Siems, Renke: Das Lesen der Anderen. Die Auswirkungen von User Tracking auf Bibliotheken, in: o-bib 9 (1), 2022, S. 1-25. Online: <<https://doi.org/10.5282/o-bib/5797>>.

Ireland<sup>2</sup>). Der internationale Datenverkehr personenbezogener Daten mit den USA darf sich daher nicht mehr auf das Privacy Shield stützen.

Das Privacy Shield war ein Abkommen zwischen der EU und den USA, welches die Übermittlungsmöglichkeiten und deren Voraussetzungen im Hinblick auf personenbezogene Daten von der EU in die USA regelte. Mit dem Privacy Shield sollte ein Nachfolger für das vorherige Abkommen „Safe Harbor“ gefunden werden, das bereits 2015 vom EuGH für ungültig erklärt wurde („Urteil Schrems I“). Personenbezogene Daten sind zum Beispiel Informationen wie Name, Standort oder auch die IP-Adresse. Der Europäische Gerichtshof (EuGH) hatte im Sinne Schrems geurteilt, dass eine Übermittlung personenbezogener Daten in die USA „Eingriffe in die Grundrechte“ ermögliche. Diese Grundrechte sind außer im Grundgesetz auch in die im Oktober 2000 in Kraft getretene Grundrechtecharta der Europäischen Union übernommen und präzisiert. Art. 7 GRCh enthält vier Gewährleistungen, nämlich des Privatlebens, des Familienlebens, der Wohnung und der Kommunikation. Die Gewährleistungen des Privatlebens und der Kommunikation schützen wesentliche Aspekte der Selbstbestimmung über das eigene Verhalten und dessen Beobachtung durch Dritte<sup>3</sup>.

Speziell geschützt ist die Entscheidungsbefugnis des Betroffenen über seine personenbezogenen Daten gem. Art. 8 GRCh. Nach Abs. 1 hat jede Person das Recht auf Schutz der sie betreffenden personenbezogenen Daten. Soweit private Daten verarbeitet werden, erstreckt sich auch die Achtung des Privatlebens nach Art. 7 GRCh auf diesen Schutz. Da der Datenschutz allerdings weiter reicht als die Achtung des Privatlebens, wurde dafür ein eigenständiges Grundrecht begründet. Abs. 2 Satz 2 fordert die Zweckbindung der Datenverarbeitung und gewährt einen Auskunfts- und Berichtigungsanspruch. Abs. 3 bestimmt, dass eine unabhängige Stelle den Datenschutz zu überwachen hat<sup>4</sup>.

Besonders Cloud-Bibliothekssysteme sind gemäß den gemachten Ausführungen besonders exponierte Objekte, da sie detaillierte Nutzungsdaten über Nutzerinnen und Nutzer verbunden mit deren Mediennutzungsverhalten und auch andere Daten, die sich anderweitig nicht messen lassen, wie zum Beispiel Ausleihdaten sowie eventu-

ell damit verbundene Bezahltransaktionen für lizenzierte Dienste speichern.

Bibliotheksnutzende haben infolge der heutigen Cloud-technologie keinen vollständigen Überblick über die erfolgten Datentransaktionen. Renke Siems führt dazu das Beispiel des WorldShare Management Services (WMS) aus der Universität Gloucestershire an, dessen Analysetools die Verantwortlichen auf dem Campus jedenfalls zu begeistern vermochten: „Die Einführung von WMS an der University of Gloucestershire hat uns die Möglichkeit geboten, Lernanalytikfunktionen von Grund auf in ein Bibliotheksmanagementsystem zu integrieren.“<sup>5</sup>

Betroffen davon sind zwei Nutzergruppen, die Forschenden wie auch die Studierenden der Universität. OCLC reagierte auf die Veröffentlichung des Beispiels im Artikel von Renke Siems mit der Entgegnung, dass OCLC nur die Daten zur Verfügung gestellt habe, die Analyse der Daten aber bei der Bibliothek der Universität Gloucester erfolgt sei.<sup>6</sup> Betroffene Personen haben zwar theoretisch Auskunfts- und Mitwirkungsrechte, aber aufgrund der arbeitsteiligen Speicherung bei verschiedenen Dienstleistern wird das EU-Grundrechtecharta-Schutzkonzept unterlaufen.

Auf diesen Umstand haben die Wissenschaft<sup>7</sup> in Zusammenarbeit mit Bibliotheksverbänden reagiert.<sup>8</sup> Das heißt nun aber nicht, dass die Datenanalyse für wissenschaftliche Bibliotheken verboten ist. Sie ist aber von ein paar Voraussetzungen abhängig:

1. Die wissenschaftlichen Bibliotheken haben eine Fürsorge- und Schutzpflicht für ihre Nutzerinnen und Nutzer. Diese Pflicht beinhaltet die Aufklärung und auch eine Warnung über eine etwaige Gefährdung der Grundrechte bei der Recherche. Sie betrifft auch die Strukturierung und Gestaltung des Webangebots der Bibliothek. Der Rechtsdienst der Universität oder Fachhochschule muss die beabsichtigte Kooperation vor Vertragsabschluss daraufhin überprüfen, ob sie die DSGVO einhält. Eine Zusammenarbeit mit Dienstleistern und Kooperationspartnern, die diese Schutzbestimmungen unterlaufen wollen, ist ausgeschlossen.
2. Die abzuschließenden Lizenzverträge müssen in jedem Fall den Vorschriften der DSGVO und den deutschen Lizenzvertragsbestimmungen entsprechen. Konkret bedeutet dies, zu verhindern, dass Bibliotheken perso-

2 Siehe Medienmitteilung des EuGH: [Der Gerichtshof erklärt den Beschluss 2016/1250 über die Angemessenheit des vom EU-US-Datenschutzschild gebotenen Schutzes für ungültig](https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:62016/JC01250&from=doctrines) (europa.eu)

3 Roßnagel, A. (Hrsg.). Das neue Datenschutzrecht, Europäische Datenschutz Grundverordnung und deutsche Datenschutzgesetze. Baden-Baden 2018

4 S. näher Roßnagel (2019, S. 2 f.). EuGH, C-101/01, ECLI:EU:C:2003:596, Rn. 86 – Lindquist; EuGH, C 131/12, ECLI:EU:C: 2014:317, Rn. 68 – Google Spain; EuGH, C-362/14, ECLI:EU:C:2015:650, Rn. 93 – Schrems; BVerfGE 84, 192 (195); 117, 202 (229); Roßnagel/Pfaffmann/Garstka, Modernisierung des Datenschutzrechts, Gutachten i.A. des BfM, 2001, S. 48 f.

5 Siems, Renke: Das Lesen der Anderen. Die Auswirkungen von User Tracking auf Bibliotheken, 3. Bibliothekssysteme, in: o-bib 9 (1), 2022, S. 1-25. Online: <<https://doi.org/10.5282/o-bib/5797>>.

6 <https://www.o-bib.de/bib/article/view/5879>

7 <https://stoptrackingscience.eu>

8 [https://www.bibliotheksverband.de/sites/default/files/2020-11/2019\\_11\\_27\\_dbv\\_Stellungnahme\\_Empfehlungen%20zu%20Methoden%20zur%20Kontrolle%20des%20Zugriffs%20auf%20wissenschaftliche%20Informationsressourcen.pdf](https://www.bibliotheksverband.de/sites/default/files/2020-11/2019_11_27_dbv_Stellungnahme_Empfehlungen%20zu%20Methoden%20zur%20Kontrolle%20des%20Zugriffs%20auf%20wissenschaftliche%20Informationsressourcen.pdf)

nenbezogene Daten in die USA übermitteln und dass US-Behörden von amerikanischen Unternehmen oder deren Tochterunternehmen in Europa die Herausgabe personenbezogener Daten verlangen können. Entweder arbeiten Bibliotheken mit solchen Unternehmen nicht zusammen oder es wird durch technisch-organisatorische zusätzliche Schutzvorkehrungen wie zum Beispiel durch Anonymisierung oder Verschlüsselung der Daten verhindert, dass amerikanische Behörden personenbezogene Daten verarbeiten können.

3. Die Bibliotheken haben keine Pflicht, ihre Nutzenden gegen möglicherweise vom Anbieter verlangte Zustimmung zur Datenauswertung zu schützen. Die Nutzenden müssen aktiv den Datenschutzregeln zustimmen können. Es würde jedoch den Aufgaben einer Bibliothek entsprechen, wenn sie ihren Nutzenden die Möglichkeit böte, sich selbst zu schützen.

Wer im Internet recherchiert, hinterlässt Spuren und diese Daten werden von Internet-Konzernen und Produktanbietern analysiert. Cookies sammeln, wenn sie denn einmal von den im Internet Surfenden akzeptiert wurden, auf deren Geräten wie Tablets, Computern und Smartphones Daten zu ihrem Surfverhalten und ihren Interessen. Internet-Nutzende werden diesen Daten entsprechenden Werbezielgruppen und Angeboten zugeordnet und sollen so zum Kauf angeregt werden. Wer also im Internet auf angeblich frei zugängliche Inhalte zugreift, muss sich bewusst sein, dass er diesen Zugriff mit den eigenen Daten bezahlt.

Anders verhält es sich, wenn lizenzierte wissenschaftliche Informationsangebote genutzt werden, wenn z.B. wissenschaftliche Artikel eines Fachgebiets über den Zugang einer wissenschaftlichen Bibliothek recherchiert und gelesen werden. Für derartige Zugriffe wird ja mittels einer Lizenzgebühr der Bibliothek für den Gebrauch des Informationsangebots bezahlt. Gerade für diese Zugriffe möchte man nicht mittels Daten offenlegen, an was man interessiert ist oder woran geforscht wird. Dies hält Verlage und Informationsanbieter aber nicht davon ab, nach genau gleicher Weise zu verfahren und die Daten für Marketingzwecke abzuschöpfen, nach dem von Amazon bekannten Motto: „Kunden, die diesen Artikel angesehen haben, haben auch angesehen...“. Dem gegenüber steht der berechnete Anspruch der Nutzenden, dass die gewünschte wissenschaftliche Publikation im jeweiligen Onlineangebot gut recherchier- und auffindbar ist.

Der Anspruch der Wissenschaftlerinnen und Wissenschaftler zum Zugriff auf die Inhalte digitaler wissenschaftlicher Publikationen hat sich über die Jahre dank der Fortschritte in der technologischen Entwicklung gewandelt. Von der verteilten Authentifizierung und Auto-

risierungen für Webanwendungen und Webservices (Dan Shibolet-Verfahren) und ähnlichen Technologien darf heute erwartet werden, dass man sowohl vom Büro wie auch von zuhause aus und nicht nur in der Bibliothek, Zugriff auf die lizenzierten Inhalte hat.

Ein besonderes Augenmerk sollte auf die Studierenden während ihres Studiums gerichtet werden. Wie eine Untersuchung der Johannes-Gutenberg-Universität Mainz im Forschungsverbund mit der Goethe-Universität in Frankfurt in einer dreijährigen Studie gezeigt hat, zeigen Berufsanfängerinnen und Berufsanfänger in Jura, Medizin und Lehramt Defizite bei der kritischen Nutzung von Online-Informationen in ihrem Fachgebiet.<sup>9</sup> Fachdatendatenbanken und Inhalte des Internets nehmen eine immer zentralere Rolle in der Ausbildung ein, die durch die Pandemie noch einen zusätzlichen Schub erfahren hat.

Die Ansprüche der Informationsanbieter und Verlage gehen heute aber weiter. Zugriffe über die IP-Adresse oder über einen IP-Range erscheinen ihnen zunehmend veraltet. Sie hätten gerne den Einsatz von Technologien, die die persönliche Authentifizierung der Nutzerinnen und Nutzer gestattet. Der kritische Umgang mit diesen Informationsquellen muss im Studium gefördert werden und die Studierenden müssen durch ein gezieltes Angebot während des Studiums für das Datentracking sensibilisiert werden.

Mit dem Science Tracking würde die Metrisierung des Wissenschaftsbetriebs und seines Publikationswesens mit seiner Kommerzialisierung auf eine neue Stufe der Kommerzialisierung gehoben.

Für die heutige Wissenschaftsgesellschaft sind eine Anwendung solcher Technologien und Funktionssysteme für Marketingzwecke nicht grundsätzlich problematisch, denn Wissen wird nicht weniger, wenn es geteilt und ökonomisch verwertet wird. Als problematisch muss aber die Verschiebung der Gleichgewichte zwischen Wissenschaftler und Bibliothek einerseits und den Verlagen und Informationsanbietern andererseits angesehen werden, wenn die Kontrolle über das Wissen dahingehend verändert wird, dass die Wissenschaft und ihre Publikationsmöglichkeiten zu einem ausschließlichen Teilsystem der privatwirtschaftlichen Aktivitäten der Verlage und Informationsanbieter wird. Dann besteht das Risiko, dass die Steuerung des verfügbaren Wissens der Wissenschaft von einem kleinen Kreis weniger Informationsanbieter und Verlage übernommen wird. Das verfassungsmäßige Grundrecht der Freiheit der Wissenschaften wäre damit in Frage gestellt.

Science Tracking wäre somit ein weiteres Steuerungselement in der Hand der privatwirtschaftlichen Wissenschaftsverlage, welches das Gleichgewicht zwischen Wis-

9 <https://bildungsklick.de/schule/detail/internet-kompetenz-von-berufseinsteigern-defizitaer>

senschaft und Informationsanbietern sowie Verlagen noch einmal zuungunsten der ersteren verschieben würde. Problematisch ist dabei, dass die Verlage die Trends und Entwicklungen in den verschiedenen Wissenschaftsgebieten auswerten und diese möglicherweise auch beeinflussen können. Das mit öffentlichen Mitteln finanzierte Wissen auf ihren Servern ist dann für die Wissenschaft nur noch hinter einer Bezahlschranke zugänglich.

Es ist keine Schwarzmalerei, dass Science Tracking das Ungleichgewicht des Wissens nicht nur zugunsten der Privatisierung und eines privatwirtschaftlichen Monopols verschiebt, wie der Beitrag von Renke Siems in o-bib zeigt. Wissenschaftsverlage haben daher ein großes Interesse daran, zu wissen, wer an welchen Institutionen und in welchen Forschungsgebieten tätig und an deren wissenschaftlichen Erkenntnissen interessiert ist. Diese Kenntnisse sind für die Verlage und Informationsanbieter, aber auch für die Steuerung der staatlichen Forschungspolitik an den Fachhochschulen und den Universitäten künftig von großer Bedeutung. Darüber hinaus erschließt sich für die Verlage und Informationsanbieter mit der Analyse der Forschungsdaten ein neues Geschäftsfeld über ihr angestammtes Feld als Content Provider hinaus. Generell verfolgen alle Digital-Industrien die Tendenz, in sich geschlossene Umwelten zu entwickeln, die die Nutzerinnen und Nutzer nicht mehr verlassen können. Apple, Metaverse und weitere Firmen haben dies bereits vorgemacht. Die sogenannten DEAL-Verträge brachten keine Transparenz in Sachen Science Tracking. Genauere Analysen des Science Tracking wären nötig, um vollen Aufschluss zum Datenabfluss durch dieses Verfahren zu bekommen. Ohne diese Kenntnis kann dieser Praxis auch nicht Einhalt geboten werden.

Bereits steht den wissenschaftlichen Bibliotheken und ihren Universitäten sowie Fachhochschulen mit KI-Anwendungen die nächste Herausforderung ins Haus. Die Herkunft der Daten, auf die sich ChatGPT und andere Chatbots stützen, ist nicht einmal bekannt. Die wenigsten Universitäten und Fachhochschulen haben bisher eine Regelung gefunden, wie mit diesem Verfahren umgegangen werden soll. Verbote des Gebrauchs von ChatGPT wie an amerikanischen und australischen Universitäten machen die Runde. An der Universität Straßburg mussten Online-Prüfungen wegen des nicht erlaubten Gebrauchs von ChatGPT im Präsenzverfahren wiederholt werden. Für wissenschaftliche Bibliotheken stellt sich die Frage, wie sie mit Publikationen umgehen sollen, die möglicherweise unter Zuhilfenahme von ChatGPT erstellt wurden. Es stellen sich dabei nicht nur Fragen der wissenschaftlichen Transparenz, sondern auch urheberrechtliche Fra-

gen. Erste Kriterien zur Erkennung solcher Publikationen wurden bereits formuliert, wie beispielsweise das Fehlen von Zitatenachweisen, das Vorhandensein von wenig prägnanten Formulierungen, die mangelnde Vertiefung bei gewissen Themengebieten, die nicht im unmittelbaren Interesse der Forschung stehen. Corinne Mühlemann, Professorin für Kunstgeschichte an der Universität Bern, hat ChatGPT aktiv in ihren Unterricht eingebaut, indem sie in diesem Frühlingsemester ein Seminar zur Rolle von Textilhistorikerinnen in der Forschungsgeschichte der textilen Künste veranstaltete. Die Studierenden nutzten ChatGPT, um Fakten zu verschiedenen Textilhistorikerinnen zu recherchieren und danach ihre Seminararbeit zu schreiben, wie die Zürcher Tageszeitung „Tagesanzeiger“ schreibt.<sup>10</sup> Im Seminar ging es auch [sc. neben dem Forschungsthema] darum, das Bewusstsein für die Stärken und Schwächen des Einsatzes von ChatGPT zu schärfen. Die Stilistik der durch den Algorithmus erzeugten Texte erwies sich teilweise als verblüffend, aber auch bei der Zusammenstellung der wissenschaftlichen Fakten zeigten sich Schwächen.

Abschließend bleibt festzuhalten, dass das Bekanntwerden von ChatGPT im November vergangenen Jahres und das hohe Tempo der Weiterentwicklung dieser Technologie die Bibliotheken vor völlig neuartige Herausforderungen stellen. Mit Künstlicher Intelligenz und den technologisch darauf beruhenden Chatbots werden sich Bibliotheken neu positionieren müssen. Die Herkunft der Daten, mit denen die Chatbots trainiert werden, ist nicht bekannt. Die von den Chatbots zusammengestellten Recherche-Resultate beruhen nicht zwingend auf Daten aus Bibliotheksbeständen oder -angeboten. Was dies im Einzelnen für die Bibliotheken bedeutet, kann erst nach eingehender genauer Beobachtung der weiteren Entwicklung dieser neuen Technologien vertieft beurteilt werden. ■



**Stephan Holländer**

Lehrbeauftragter, Basel  
stephan@stephan-hollaender.ch

<sup>10</sup> Tagesanzeiger, 4.07.2023, GPT – diese Berner Professorin wagt ein Experiment. Link: <https://www.tagesanzeiger.ch/unis-zittern-vor-chat-gpt-diese-berner-professorin-wagt-ein-experiment-142577951107>