

Science Tracking auf der Spur – die Rolle von Bibliotheken im Kontext von Reading Analytics bei wissenschaftlichen Verlagen

Tracking science – the role of libraries in the context of reading analytics at academic publishers

Abstract

This paper contains the written thoughts and selected discussion points from the presentation held by Professor Brembs, Professor Förstner, and Petra Labriga at the annual meeting of the AGMB in September 2021. It addresses the accusations made there against academic publishers regarding allegedly unlawful data tracking in science and focuses on the resulting perspective of the strategic license management at ZB MED. As a starting point this paper proposes a concerted effort by German and European libraries to review the allegations, a joint effort to obtain legal advice, possibly under the leadership of the German Research Foundation (DFG), and a resulting inclusion of relevant clauses in license agreements to prevent the practice of unlawful data tracking and a possible trade of data.

Keywords: science tracking, reading analytics, academic publishing

Zusammenfassung

Dieser Fachbeitrag enthält verschriftlichte Gedanken und ausgewählte Diskussionspunkte aus dem Vortrag „Science Tracking – das ‚neue‘ Geschäftsmodell der Wissenschaftsverlage“ von Professor Brembs, Professor Förstner und Petra Labriga auf der AGMB-Tagung im September 2021. Er befasst sich mit den dort erhobenen Vorwürfen gegenüber den wissenschaftlichen Fachverlagen hinsichtlich des unrechtmäßigen Datentrackings in der Wissenschaft und fokussiert im Besonderen auf die daraus resultierende Perspektive des strategischen Lizenzmanagements bei ZB MED. Als erste Lösungsvorschläge werden eine konzentrierte Aktion der Überprüfung der Vorwürfe durch deutsche und europäische Bibliotheken, eine gemeinsam angestrebte Rechtsberatung, eventuell unter Führung der Deutschen Forschungsgemeinschaft (DFG), und eine daraus resultierende Aufnahme relevanter Klauseln in Lizenzverträge angeregt, die die Praxis des unrechtmäßigen Datentrackings und des eventuellen Handelns mit Daten unterbinden sollen.

Schlüsselwörter: Datentracking, wissenschaftliche Fachverlage

Petra Labriga¹

1 Strategisches
Lizenzmanagement, ZB MED
– Informationszentrum
Lebenswissenschaften, Köln,
Deutschland

Überwachung der Nutzenden oder Cyber-Sicherheit?

Wer sich im Internet bewegt, hinterlässt Spuren und wird von Internet-Konzernen und Produkthanbietern beobachtet. Cookies sammeln, so sie einmal akzeptiert werden, auf Computern Daten zu Surfverhalten und Interessen. Internet-Nutzende werden entsprechenden Werbezielgruppen und Angeboten zugeordnet und sollen zum Kauf angeregt werden. Wer also im Internet vermeintlich kostenlosen Content liest, sollte sich darüber bewusst sein, dass er diesen mit den eigenen Daten bezahlt.

Wie verhält es sich aber, wenn wissenschaftliche Inhalte im Abonnement genutzt werden, wenn z.B. wissenschaftliche Artikel eines Fachgebiets über einen institutionellen Zugang einer Bibliothek gelesen werden? Schließlich wird für den Lesezugriff in diesem Fall bezahlt und niemand möchte seine Daten dahingehend preisgeben, womit man sich gegenwärtig wissenschaftlich beschäftigt. Gleichzeitig wünschen sich Leser:innen von wissenschaftlichen Inhalten, dass diese online gut auffindbar und von überall aus erreichbar sind. Das Ziel von Bibliotheken ist es, Wissenschaftler:innen Zugriff auf subskribierte Inhalte auf für diese bequeme Weise auf dem Campus, aber auch zu Hause oder anderswo zu gewähren. Verlage betrachten zunehmend Zugriff auf Verlagsinhalte über die IP-Adresse der Bibliotheken für veraltet und drängen auf den Einsatz von Technologie zur persönlichen Authentifizierung der einzelnen Lesenden. Sie begründen dies zum einen mit Sicherheitsbedenken: Seit dem – illegalen – Herunterladen großer Datenmengen von wissenschaftlichen Artikeln durch *Sci-Hub* [1], [2] wollen Fachverlage einen erneuten Content-Diebstahl, ausgelöst durch mögliche Sicherheitslücken auf Universitätsservern oder durch Mithilfe von *Sci-Hub* wohl gesonnenen Bibliotheksnutzer:innen in Zukunft verhindern. Die Scholarly Networks Security Initiative (SNSI) [3] möchte Bibliotheken, Universitäten und ihre Nutzenden davon überzeugen, dass nicht nur die Verlage und ihre Inhalte gefährdet, sondern auch die Institutionen potenziell kriminellen Handlungen ausgesetzt sein könnten.

Die Verlagsbranche hat darauf mit dem Projekt RA 21 reagiert, welches später in der SeamlessAccess.org [4] mündete und aus dem mittlerweile der Dienst GetFTR (Get Full Text Research) [5] entstanden ist. Es geht hier im Schwerpunkt um Single Sign-On-Angebote der Verlage für Bibliotheken und ihre Nutzer:innen, die den von ihnen benötigten Content durch persönliche Authentifizierung von überall her abrufen können.

Reading Analytics – normales Geschäftsgebaren oder illegales Vorgehen?

Während die Verlage sich auf die Sicherheit berufen, lassen sich auch andere Vorteile für die Verlagsseite vermuten. Diese erkennt, wer liest, zu welcher Institution

die Lesenden gehören, für welches Fachgebiet und aktuelle Themen sie sich interessieren. Es lassen sich fachliche Zusammenhänge zwischen Lese- und Suchverhalten schließen und es besteht die Möglichkeit, weitere Literatur zum Thema vorzuschlagen. Sind diese Artikel z.B. nicht in der Subskription vorhanden, lässt sich über weitere Artikelverkäufe zusätzlich Geld verdienen. Die Analyse von Suchanfragen hilft dem Verlag darüber hinaus dabei, neue Themen und Autoren zu identifizieren und gibt Hinweise darauf, auf welchen Gebieten sich eventuell Neugründungen von Zeitschriften lohnen könnten.

Diese Form von Reading Analytics ist etabliert und akzeptiert, da sie der Angebotsverbesserung dient. In letzter Zeit ist die Profilbildung bei Verlagen jedoch unter Verdacht und in starke Kritik geraten. Marktveränderungen und der Einsatz technischer Möglichkeiten scheinen der Überwachung von Nutzer:innen Vorschub zu leisten. Zum einen verändern sich große Verlage wie Elsevier und SpringerNature von Fachliteratur publizierenden Verlagen hin zu Daten-Analyse-Firmen. Nach dem Motto „Workflow is the new content.“ (so Lorcan Dempsey in einem Tweet vom 04.09.2019 [6]) haben Verlage durch zahlreiche Zukäufe anderer Firmen und eigene Weiterentwicklung nun Produktportfolios, die den vollständigen Forschungszyklus abbilden und Forschende, so das deklarierte Ziel, auf diesem gesamten Zyklus begleiten wollen. Die Macht der großen Player macht die Wissenschaft zunehmend misstrauisch. Wenn dann Untersuchungen einzelner Informationsspezialisten ergeben, dass auf Verlagsseiten verschiedene Tracking-Methoden eingesetzt werden ([7], dort erwähnte Tracking-Methoden: Harvesting Bid Streaming, Port Scanning, Microtargeting und Trojaner), um Lesende potentiell auszuspionieren, erreicht das Unbehagen hinsichtlich der Marktmacht der großen Verlage und ihrer Konzerne ein neues Niveau. Verlage, die sich gerne als wichtige Partner der Wissenschaft präsentieren, agieren als Monopolisten und diktieren den Bibliotheken seit Jahrzehnten zum Teil enorme Preissteigerungen. Auch der Paradigmenwechsel hin zur Open-Access-Publikation, für die die Autor:innen bzw. ihre Institutionen bezahlen, hat das Kostenproblem für die Wissenschaft und Bibliotheken nicht gelöst. Abhängig vom Prestige des Verlags und des Journals werden auch weiterhin exorbitant hohe Preise, nun für die Veröffentlichung statt für das Leserecht, bezahlt. Die generelle Unzufriedenheit mit dieser Situation verstärkt die gerechtfertigte Empörung, die ein mögliches Fehlverhalten auf Verlagsseite hinsichtlich Datentracking auslöst.

Datentracking in der Wissenschaft

Auf die Problematik des Datentrackings in der Wissenschaft wurden zuerst Informationsspezialist:innen in den USA aufmerksam. Hier soll das Sammeln, die Weitergabe bzw. der Verkauf der Daten an die US-amerikanische Behörde ICE (Immigration and Customs Enforcement) durch eine Firma des RELX Konzern, zu dem auch Elsevier gehört, zur Folge gehabt haben, dass Menschen aufge-

spürt und deportiert worden seien [8]. Das Sammeln von Daten zur Profilbildung wirft rechtliche und ethische Fragen auf und eine so gestaltete potentielle Überwachung von Wissenschaft richtet sich eindeutig gegen die Wissenschaftsfreiheit.

Informationsspezialist:innen und Wissenschaftler:innen nehmen diese Vorwürfe ernst und in Deutschland werden die kritischen Stimmen von der Deutschen Forschungsgemeinschaft (DFG) [9] unterstützt, die sich meines Wissens bisher als einzige nationale Forschungsgesellschaft klar gegen das Science Tracking positioniert hat. Die Verlage Elsevier [10] und Digital Science¹ haben auf das DFG Papier „Datentracking in der Wissenschaft“ reagiert und die Vorwürfe des Datentrackings im Sinne von kommerziellem Science Tracking weit von sich gewiesen. Es gebe, so diese zwei Verlage, weder eine Überwachung der Wissenschaft noch einen Datenverkauf.

Wie sollen Bibliotheken auf diese Anschuldigungen reagieren?

Bibliotheken verhandeln Zugriffs- und Besitzrechte von Content und bewegen sich seit der Zeitschriftenkrise der 80er und 90er Jahre auf schwierigem Parkett, da sie die Literaturversorgung in einem Monopolmarkt bei gleichzeitig schrumpfenden Bibliotheksbudget organisieren müssen. Gleichzeitig begleiten sie Wissenschaftler:innen und Verlage beratend durch die Transformation hin zu Open-Access-Publikation und stellen ihre Erwerbungs- auf Informationsbudgets um. Welche Möglichkeiten bieten sich Bibliotheken in dem Versuch, Nutzer:innen vor potentiell datensammelnden Verlagen zu schützen?

Eigene Untersuchungen anstellen

Bibliotheken können durch eigene Untersuchungen zunächst einmal überprüfen, ob sich die Vorwürfe des Daten Trackings und des Handels mit Daten durch Fachverlage und ihre Mutterkonzerne in Europa beweisen lassen. Stichproben von Kolleg:innen haben ergeben, dass sich im Jahr 2021 wesentlich weniger Tracker auf einer Seite im Einsatz befanden als bei einer früheren Überprüfung in 2020. Eine Stichprobe in Fachjournal *Nature* unter Einsatz von *Ghostery* ergab im August 2020 den Einsatz von 73 Trackern im Vergleich zu 2 Trackern bei einer Stichprobe im September 2021 (siehe Vortrag zu Science Tracking auf der AGMB-Tagung 2021 [11], ab Minute 5:15). Das entkräftet die Vorwürfe nicht völlig, mag aber ein Hinweis darauf sein, dass Tracking-Aktivitäten eventuell bereits zurückgefahren oder stark eingeschränkt worden sind. Hier bedarf es einer regelmäßigen Überprüfung durch versierte Kolleg:innen, eventuell durch eine über Deutschland verteilte Arbeitsgruppe, die sich diese Aufgabe hinsichtlich der Vielzahl der Verlage und Journale aufteilt.

Bewusstsein schaffen

Bibliotheken können Aufklärung und Bewusstsein schaffen. Obwohl Bibliothekar:innen zu den ersten *Whistleblowern* [12] gehörten, hat sich zunächst die Wissenschaft der Aufgabe angenommen, das Thema Datentracking in der Wissenschaft zu positionieren und so auf dessen Risiken und Gefahren aufmerksam zu machen. Hier bekommen die aktiven Wissenschaftler:innen nun zunehmend Unterstützung von Informationsfachleuten, also vor allem von denjenigen, die mit Bibliothekslicenzierung, Erwerbung und Publikationsberatung befasst sind. Es soll jedoch nicht unerwähnt bleiben, dass es nicht für alle Wissenschaftler:innen von Interesse zu sein scheint, ob oder welche Daten über sie gesammelt werden. Jene, die sich privat in sozialen Netzwerken wie z.B. Instagram oder Twitter bewegen, scheinen eine private Profilbildung bereits hinzunehmen und eine berufliche nicht zu fürchten. Auch hier kann Aufklärung dahingehend, was ein Missbrauch der Daten nach sich ziehen kann, zu mehr Bewusstsein führen.

Bestehendes Recht durchsetzen

Wo ein Verstoß gegen geltendes Recht (DSGVO) nachgewiesen werden kann, sollte dies zur Anklage gebracht werden. Bibliotheken beziehungsweise ihre übergeordneten Organisationen und Behörden sind in der Lage, eine solche juristische Auseinandersetzung, die einzelnen eher unmöglich erscheinen mag, zu führen. Hier können sich Bibliotheken für die freie Wissenschaft stark machen und engagieren.

Ergänzung der Lizenzverträge

Um eventuellen Rechtsstreitigkeiten vorzubeugen, möchte ich jedoch als nächsten Schritt empfehlen, Formulierungen in Lizenzverträgen aufzunehmen, die das Sammeln von Daten bezüglich einzelner Lesender grundsätzlich ausschließen. Die Verhandlungen darüber werden ein Lackmustest dahingehend sein, welche Prozedere momentan bei Verlagen aktiv sind. Bei Ablehnung von Klauseln kann die Bibliothek Begründungen erfragen und sich so ein Bild der gängigen Praxis machen. Auch bei der Formulierung solcher zu erstellenden Lizenzbausteine muss nicht jede Einrichtung das Rad neu erfinden. Es ist denkbar, dass die DFG durch Einholung von juristischer Beratung bei der Textformulierung unterstützen oder ein Bibliotheksverband die Kosten für Rechtsberatung übernehmen kann.

Fazit

Seit dem Start der Petition Stop Science Tracking [7] haben sich 920 deutsche und europäische Wissenschaftler:innen (Stand 12/2021) mit ihrer Unterschrift an dieser beteiligt. Das Thema Datentracking in der Wissenschaft wird momentan auf diversen Bibliotheksveranstaltung

thematisiert und es entsteht zunehmend ein Bewusstsein für die Risiken von Datensammlung durch Reader Analytics auf Verlagsplattformen. Die unterschiedlichen Expert:innen in Bibliotheken sollten aktiv werden, um Risiken abzuwägen, gesicherte Erkenntnisse zu erlangen und entsprechende Fragen, Antworten und Maßnahmen zu formulieren. In der Kommunikation mit den Verlagen sollten neue Lizenzen erarbeitet werden, die die illegale Sammlung von Lesenden-Daten und den etwaigen Handel mit diesen Daten ausschließen.

Anmerkung

¹ Digital Science äußert sich auf Anfrage zum Thema wie folgt: “Dimensions does not sell usage data, either on a per user or aggregated basis, or otherwise seek to monetize it. Where we do log usage information, we aim to be as transparent as possible and always treat the activity of our users as confidential. Like any web-based service, usage information is collected via our logs and cookies. This is necessary for any number of reasons, such as for security (the maintenance of proper logs is an essential function of security), to provide user support (we need to be able to see usage history to help users work out why they are getting e.g. a particular search) and to identify entitlements (e.g. so a university user can get seamless access to the full text of articles that their institution subscribes to) etc. So the collection of usage information in of itself should not be controversial. Put another way, there are lots of legitimate reasons for collecting usage information. What we do not do is try to monetize that information or use it in a way that gives us insight into the topics a user (or group of users) may be researching – perhaps unlike other businesses, this is not of interest to us.”

Interessenkonflikte

Die Autorin erklärt, dass sie keine Interessenkonflikte in Zusammenhang mit diesem Artikel hat.

ORCID

0000-0001-9892-652X

Literatur

1. Cummins E. How academic pirate Alexandra Elbakyan is fighting scientific misinformation. *Vice*. 2021 Jun 08 [cited 08.12.2021]. Available from <https://www.vice.com/en/article/v7ev3x/how-academic-pirate-alexandra-elbakyan-is-fighting-scientific-misinformation>
2. Graber-Stiehl I. Science's pirate queen. *The Verge*. 2018 Feb 08 [cited 08.12.2021]. Available from: <https://www.theverge.com/2018/2/8/16985666/alexandra-elbakyan-sci-hub-open-access-science-papers-lawsuit>

3. Scholarly Networks Security Initiative [Internet]. [cited 08.12.2021]. Available from <https://www.snsi.info/>
4. SeamlessAccess [Internet]. [cited 08.12.2021]. Available from: <https://seamlessaccess.org/>
5. GetFullTextResearch [Internet]. [cited 08.12.2021]. Available from: <https://www.getfulltextresearch.com>
6. Dempsey L. Workflow is the new content > If the map becomes the territory then we will be lost – Librarian of Things [Tweet]. Twitter; 2019 Mar 04. Available from: <https://twitter.com/lorcand/status/1102376546407538690>
7. Background information. Publishers track researchers for multiple reasons. In: Stop Tracking Science [Internet]. [cited 08.12.2021]. Available from: <https://stoptrackingscience.eu/background-information/#reasons>
8. Biddle S. LexisNexis to provide giant database of personal information to ICE. *The Intercept*. 2021 Apr 02 [cited 08.12.2021]. Available from: <https://theintercept.com/2021/04/02/ice-database-surveillance-lexisnexis/>
9. Deutsche Forschungsgemeinschaft. Datentracking in der Wissenschaft: Aggregation und Verwendung bzw. Verkauf von Nutzungsdaten durch Wissenschaftsverlagen. Bonn; 2021 Oct 28 [cited 08.12.2021]. Available from: https://www.dfg.de/download/pdf/foerderung/programme/lis/datentracking_papier_de.pdf
10. Davis A. Elsevier-Antwort auf DFG-Papier Datenverfolgung in der Forschung. *Elsevier*; 2021 Jun 18 [cited 08.12.2021]. Available from: <https://www.elsevier.com/de-de/connect/elsevier-antwort-auf-dfg-papier-datenverfolgung-in-der-forschung>
11. Brems B, Förstner K, Labriga P. Science Tracking – das neue Geschäftsmodell der Wissenschaftsverlage [Video]. In: Jahrestagung der Arbeitsgemeinschaft für medizinisches Bibliothekswesen (AGMB). YouTube; 2021. Available from: https://www.youtube.com/watch?v=Eei3_wVXWWY
12. Hanson C. User Tracking on Academic Publisher Platforms [Talk]. In: Coalition for Networked Information Spring 2019 Member Meeting; April 8-9, 2019; St. Louis, Missouri. Available from: <https://www.codyh.com/writing/tracking.html>

Erratum

Im Abschnitt Datentracking in der Wissenschaft sowie in der Anmerkung wurde Springer durch Digital Science ersetzt. Das Zitat wurde fälschlicherweise dem Verlag SpringerNature zugeordnet, die Information wurde jedoch von Digital Science zur Verfügung gestellt. SpringerNature und Digital Science gehören beide zur Holtzbrinck Publishing Group.

Korrespondenzadresse:

Petra Labriga
Strategisches Lizenzmanagement, ZB MED –
Informationszentrum Lebenswissenschaften, Gleueler
Straße 60, 50931 Köln, Deutschland
labriga@zbmed.de

Bitte zitieren als

Labriga P. Science Tracking auf der Spur – die Rolle von Bibliotheken im Kontext von Reading Analytics bei wissenschaftlichen Verlagen. GMS Med Bibl Inf. 2021;21(3):Doc18.
DOI: 10.3205/mbi000507, URN: urn:nbn:de:0183-mbi0005077

Artikel online frei zugänglich unter

<https://www.egms.de/en/journals/mbi/2021-21/mbi000507.shtml>

Veröffentlicht: 20.12.2021

Veröffentlicht mit Erratum: 18.01.2022

Copyright

©2021 Labriga. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.