



Fachhochschule Graubünden
University of Applied Sciences

Churer Schriften zur Informationswissenschaft

Herausgegeben von
Wolfgang Semar Bernard Bekavac, Ivo Macek, Armando Schär

Arbeitsbereich Bachelor of Science
in Information Science

Schrift 167

Deep Web und Bibliotheken: Stand der Dinge

Lorena Staiger

Chur 2023

Churer Schriften zur Informationswissenschaft

Herausgegeben von Wolfgang Semar,
Bernard Bekavac, Ivo Macek, Armando Schär

Schrift 167

Deep Web und Bibliotheken: Stand der Dinge

Lorena Staiger

Diese Publikation entstand im Rahmen einer Thesis zum Bachelor of Science in Information Science.

Referent: Simon Schultze

Korreferent: Prof. Dr. Rudolf Mumenthaler

Verlag: Fachhochschule Graubünden

ISSN: 1660-945X

Ort, Datum: Chur, Oktober 2023

Abstract

Das Deep Web ist aufgrund der Fülle und Qualität an Informationen integraler Bestandteil der akademischen Forschung. Die Recherche darin verlangt jedoch nach spezifischen Fähigkeiten, die Bibliothekar*innen und Informationsspezialisten zur optimalen Forschungsunterstützung einsetzen müssen. Um diese Rolle auszuführen, wird fundiertes Wissen zur gesamten Weblandschaft benötigt: Surface Web, Deep Web und Dark Web. Auch Schattenbibliotheken, die die Brücke zwischen allen drei Webschichten schlagen und eine immer grösser werdende Popularität verzeichnen, gehören dazu. Anhand drei quantitativer Befragungen und der Analyse von Zugriffsstatistiken auf Sci-Hub und Library Genesis zeigte sich ein vorhandenes grundlegendes Wissen des Deep Webs bei Bibliothekar*innen, jedoch ein Defizit bei den Schattenbibliotheken im Vergleich zur Nutzung des akademischen Mittelbaus. Ausserdem besteht ein zu wenig sichtbares oder vorhandenes Angebot von nachgefragter Rechercheunterstützung.

Vorwort

Diese Bachelorthesis wurde im Rahmen des Information Science-Studiums an der Fachhochschule Graubünden verfasst. Bereits zu Beginn des Studiums war die Themenrichtung für die Bachelorarbeit klar: Damals verfasste ich eine Studienarbeit zum Tor-Darknet und dessen Marktplätzen und war überzeugt, dass das Dark Web im weiteren Sinne mit Bibliotheken zusammenhängt. Der Fokus in dieser Arbeit lag auf dem Deep Web, dessen Verwobenheit mit dem Rest der Weblandschaft und was für eine Bedeutung Bibliotheken darin haben. Durch die Auswahl dieses Themas konnte ich als Abschluss des Studiums noch einmal gelernte Methoden anwenden und gleichzeitig persönliche Interessen im Bibliothekskontext verfolgen, die hoffentlich weiterführende Diskussionen auslösen werden.

Ich möchte mich bei meinem Betreuer, Herr Simon Schultze, und bei meinem Korreferenten, Herr Rudolf Mumenthaler, für die hervorragende Unterstützung und Begleitung im Prozess bedanken. Ebenso bei allen Teilnehmenden der Befragungen, die mir die Datengrundlagen für die Arbeit lieferten, und bei der ID Security der Uni Bern (insbesondere bei Herr Nenad Zegarac) für die Überwachung der Zugriffsstatistiken. Ein ganz herzliches Dankeschön geht auch an meinen Vorgesetzten, Herr Stefan Grosjean, der mir in der Anfangsphase wertvolle Inputs und Ideen zur Themeneingrenzung geliefert hat, und an meine Chefin Frau Dr. Michelle Schaffer, die mir unermüdlich bei der etwas mühsamen Verbreitung der Umfragen an der Uni Bern (und Swisslib) zur Seite stand.

Inhaltsverzeichnis

1	Einleitung	1
2	Theoretische Grundlagen.....	3
2.1	Internet.....	3
2.1.1	Hintergrund und Funktionsweise	3
2.1.2	Taxonomie des Internets	5
2.1.3	Deep Web im Bibliothekskontext	16
2.1.4	Dark-Web-Technologien	19
2.2	Akademische Forschung und Schattenbibliotheken	23
2.2.1	Sci-Hub und Library Genesis	24
2.2.2	Auswirkungen auf traditionelles Publikationsmodell und Bibliotheken.....	26
2.3	Forschungsunterstützung durch Bibliotheken	28
3	Methodik.....	29
4	Ergebnisse	31
4.1	Umfragen	31
4.1.1	Auswertung A und B	32
4.1.2	Auswertung C	42
4.2	Analyse Sci-Hub und Library Genesis	47
5	Diskussion.....	53
6	Fazit	57
7	Literaturverzeichnis	59

Abbildungsverzeichnis

Abbildung 1. Die Evolution des ARPANET von 1969-1982 (Leyden, 2019).....	4
Abbildung 2. Internetverkehr der drei Teilbereiche des Webs (eigene Darstellung nach Chertoff (2017), Eisberg: Lizenz Adobe Stock).....	6
Abbildung 3. Verbindungsaufbau einer Sitzung im Tor-Browser (Screenshot Tor-Browser).....	12
Abbildung 4. Circuit der Onion-Adresse von Facebook (Screenshot Tor).....	15
Abbildung 5. Häufigkeit Nutzung Deep Web, Gruppe A (eigene Darstellung).....	33
Abbildung 6. Häufigkeit Nutzung Deep Web, Gruppe B (eigene Darstellung).....	33
Abbildung 7. Zugriff auf das Dark Web, Gruppe A (eigene Darstellung).....	34
Abbildung 8. Zugriff auf das Dark Web, Gruppe B (eigene Darstellung).....	34
Abbildung 9. Häufigkeit der Nutzung von Sci-Hub/LibGen, Gruppe A (eigene Darstellung).....	36
Abbildung 10. Häufigkeit der Nutzung von Sci-Hub/LibGen, Gruppe B (eigene Darstellung).....	36
Abbildung 11. Gründe der Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Fachhochschule, Gruppe A (eigene Darstellung).....	37
Abbildung 12. Gründe der Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Universität, Gruppe B (eigene Darstellung).....	38
Abbildung 13. Wunsch nach Unterstützung bei der Recherche im Deep Web durch Bibliotheken, Gruppe A (eigene Darstellung).....	39
Abbildung 14. Wunsch nach Unterstützung bei der Recherche im Deep Web durch Bibliotheken, Gruppe B (eigene Darstellung).....	41
Abbildung 15. Häufigkeit der Nutzung von Google für Benutzendenanfragen (eigene Darstellung).....	42
Abbildung 16. Häufigkeit der Nutzung des Deep Webs im Berufsalltag (eigene Darstellung).....	43
Abbildung 17. N=32, Altersverteilung Nein-Stimmen zur Kenntnis von Sci-Hub/LibGen (eigene Darstellung).....	44
Abbildung 18. Altersverteilung der gesamten Umfrage C (eigene Darstellung).....	45
Abbildung 19. Vermutete Gründe für die Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Bibliotheken, Untergruppe C01 (eigene Darstellung).....	46
Abbildung 20. Vermutete Gründe für die Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Bibliotheken, Untergruppe C02 (eigene Darstellung).....	47

Tabellenverzeichnis

Tabelle 1. Wissenswertes für Bibliothekar*innen vor einer Suche im Deep Web (nach Prasad, 2017, S. 420–422)	17
Tabelle 2. Fragegruppen der Fragebögen A und B (eigene Darstellung).....	31
Tabelle 3. Fragegruppen des Fragebogen C (eigene Darstellung)	31
Tabelle 4. Länderverteilung der Anzahl Downloads, Juli 2023 (eigene Darstellung nach Sci-Hub, 2023c)	48
Tabelle 5. Zugriffe auf Sci-Hub und Library Genesis aus dem Netzwerk der Universität Bern, 12.06.-12.07.23 (eigene Darstellung nach N. Zegarac, persönl. Mitteilung, 13.07.2023a).....	49
Tabelle 6. nach N. Zegarac, persönl. Mitteilung, 13.07.2023b)	50
Tabelle 7. Zugriffe auf Mirror-Seiten von LibGen aus dem Netz der Universität Bern, alle Ports (eigene Darstellung nach N. Zegarac, persönl. Mitteilung, 13.07.2023b)	51

Abkürzungsverzeichnis

DOAJ	Directory of Open Access Journals
E2EE	End-to-end encryption
EFF	Electronic Frontier Foundation
HTTPS	Hypertext Transfer Protocol Secure
I2P	The Invisible Internet Project
ICCC	International Conference on Computer Communications
ID	Informatikdienste
ISP	Internet Service Provider
LibGen	Library Genesis
NRL	U.S. Naval Research Laboratory
OA	Open Access
P2P	Peer-to-peer
Tor	The Onion Router
WWW	World Wide Web

1 Einleitung

Bibliotheken sind seit jeher Teil des öffentlichen Lebens. Sie bieten Zugang zu Information, Wissen und Bildung in den unterschiedlichsten Schichten der Gesellschaft. Die Form der Dokumentation, Speicherung und Aufbewahrung dieses Wissens hat sich jedoch mit dem Wandel der Zeit geändert, und im heutigen digitalen Zeitalter schwinden analoge Aufbewahrungsmethoden. Nachschlagewerke in Form von Lexika sind nicht mehr die erste Wahl, sondern die Formulierung einer Suchanfrage in Google. Ben Gomes, Senior Vice President bei Google, beschreibt die Google-Suche als Revolution des freien Informationszugangs im Vergleich zur Zeit vor dem öffentlich zugänglichen Internet:

Today, we measure a fraction of a second the time it takes you to get information. [...] It can enable people around the world to have equal access to information. It's not just the people in some places who have access to the best libraries, everybody should have access to the highest quality information. (Google, 2020)

Die Google-Suche und damit das Surface Web reicht jedoch nur bis zur obersten Schicht des Webs und gibt dementsprechend nur einen Bruchteil der verfügbaren Informationen zurück. Das Deep Web, das nicht durch konventionelle Suchmaschinen erreicht wird und eine höhere Qualität an Informationsressourcen beinhaltet, liegt trotz seiner enormen Grösse aussen vor. Forschungsergebnisse werden primär digital über wissenschaftliche Verlage im Deep Web publiziert, und deren Verfügbarkeit – und allenfalls Unterstützung dafür – für Forschende und Studierende ist wiederum abhängig von Lizenzverhandlungen durch Bibliotheken und dem grundlegenden Wissensstand der Bibliothekar*innen. Während Bibliotheken jedoch mit Verlagen Vertragsverhandlungen zu Lizenzen führen, umgehen Studierende und Forschende die Einschränkungen des Deep Webs – und damit die Paywalls der Verlagsseiten – mit Schattenbibliotheks-Webseiten wie Sci-Hub und Library Genesis. Ebenso werden Cyber Security und Privatsphäre im Netz in Bibliothekskreisen immer wieder kritisch diskutiert, während die Kenntnis über den Tor-Browser, der gleichzeitig auch den Einstieg in das Tor-Netzwerk (Dark Web, Teil des Deep Webs) bietet, und dessen Technologie unbekannt ist. Zu diesem Umstand gibt es eine Handvoll Studien, die gar die Nutzung des Tor-Browsers innerhalb von Bibliotheken zum Schutz der Benutzenden empfiehlt. Darüber hinaus existieren jedoch wenige Thematisierungen im informationswissenschaftlichen Kontext, obwohl Bibliotheken sich selbst als Expert*innen der Information und Recherche in den Fokus stellen. Vertiefte Kenntnisse über das

Internet und das Web sollten also an der Tagesordnung sein, um als Bibliothek aktuell zu bleiben und den Wissenstransfer weiterzuführen.

Die vorliegende Arbeit beleuchtet die unterschiedlichen Aspekte des Internets und des darauf laufenden Webs mit seinen verschiedenen Schichten. Nebst grundlegender Erarbeitung der gesamten Internetlandschaft wird der Ursprung des «The Onion Router Project» und dessen Bestrebungen der Online-Privatsphäre aus neutraler Sicht aufgezeigt. Sci-Hub und Library Genesis werden in Zusammenhang mit der Verlagslandschaft gebracht und deren Auswirkungen darauf und auf Bibliotheken ergründet. Die Arbeit strebt an, die Rolle von Bibliothekar*innen in der Meisterung der Informationsbeschaffung im Web aufzuzeigen und herauszufinden, inwiefern Kenntnisse über das Deep Web vorhanden sind und diese Rolle bereits der Realität entspricht. Ebenso soll die Nachfrage dieser Fähigkeit ergründet und somit der Stand der Bibliotheken in diesem Kontext anhand folgender Forschungsfrage diskutiert werden:

«Wie gross ist die Diskrepanz im Kenntnisstand zwischen Bibliothekar*innen und Studierenden/Forschenden bezüglich des Deep Webs, und inwiefern beeinflusst dies die Aktualität und Relevanz von Bibliotheken?»

Die Methodik besteht nebst der Literaturrecherche zur aktuellen Forschung aus drei Umfragen, die jeweils einer anderen Zielgruppe vorgelegt werden: Schweizweite Bibliothekar*innen, Studierende an der FHGR und dem akademischen Mittelbau an der Universität Bern. Ausserdem werden Zugriffsstatistiken auf die Schattenbibliotheken Sci-Hub und Library Genesis aus dem Netzwerk der Universität Bern erhoben und diskutiert. Die Erkenntnisse der Literaturrecherche werden abschliessend mit der Datenauswertung in Verbindung gebracht.

2 Theoretische Grundlagen

Das Konzept der «Cognitive Invisibility» besagt zusammengefasst, dass eine Person nur so viel bei ihrer Recherche in Erfahrung bringt, wie sie aus ihrem Wissen über die Recherchemöglichkeiten schöpfen kann (Ford & Mansourian, 2006). Um das Thema dieser Arbeit behandeln zu können und das Wissen dafür aufzubauen, ist demnach eine Auslegung der Grundlagen des Internets und des World Wide Webs unumgänglich.

2.1 Internet

Der Begriff des Internets im heutigen Sinne wird als Synonym zum World Wide Web verwendet, auch wenn diese nicht dasselbe verkörpern (vgl. Kap. 2.1.2). Es ist Bestandteil des täglichen Gebrauchs für über 5.3 Milliarden Personen – respektive 66% – der Bevölkerung weltweit und befindet sich in einer kontinuierlich positiven Wachstumsrate seit Anbeginn der grossflächigen Nutzung (International Telecommunication Union, 2022). Dieser Umstand unterstreicht die Notwendigkeit von geschulten Fachleuten zur Navigation des Internets in mindestens zwei seiner Teilbereiche (vgl. Kap. 2.1.2). Um diese in ihrer Gesamtheit zu verstehen, sind grundlegende Kenntnisse des Internets und dessen Funktionsweise Voraussetzung.

2.1.1 Hintergrund und Funktionsweise

Die Entstehung des Internets ist auf das ARPANET zurückzuführen. Entwickelt vom Verteidigungsministerium der Vereinigten Staaten (Abteilung «Advanced Research Project Agency», ARPA) war es das erste Netz, das eine Kommunikation zwischen mehreren Computern ermöglichte. ARPAs eigentlicher Entwicklungsgrund im Jahre 1958 war militärischer Natur und soll den USA einen Kommunikationsvorsprung liefern, insbesondere nach dem Start von Sputnik (Abbate, 1994, S. 3).

ARPANET galt als Sensation aufgrund zweier entscheidender Neuerungen in der Technologie: Es kombinierte mehrere Computer verschiedener Hersteller in einem Netzwerk und erreichte gleichzeitig einen fortlaufenden Datenaustausch über längere Distanzen. Das Hauptziel des Netzwerks wurde darin gesehen, den Informationstransfer und die Informationsverarbeitung zwischen den ARPA-Stationen voranzutreiben, die Rechenkosten zu reduzieren sowie eine Zusammenarbeit von Forschenden an unterschiedlichen Standorten zu inspirieren. Die ersten Stationen fanden ihren Einsatz an vier unterschiedlichen Universitäten im Jahre 1969 (s. Abb. 1), und nach erfolgreicher Präsentation des Experiments an der ICCC folgten weitere Projekte. Von hoher Gewichtung ist jenes

Projekt, das mehrere Netzwerke miteinander verbinden, also «internetworken», sollte. Hier entstanden die Protokolle TCP/IP, die die Interaktion zwischen den Computern regelten – diese Netzwerkkommunikation mittels Protokollen begründete den Begriff des Internets (Abbate, 1994, S. 1–5).

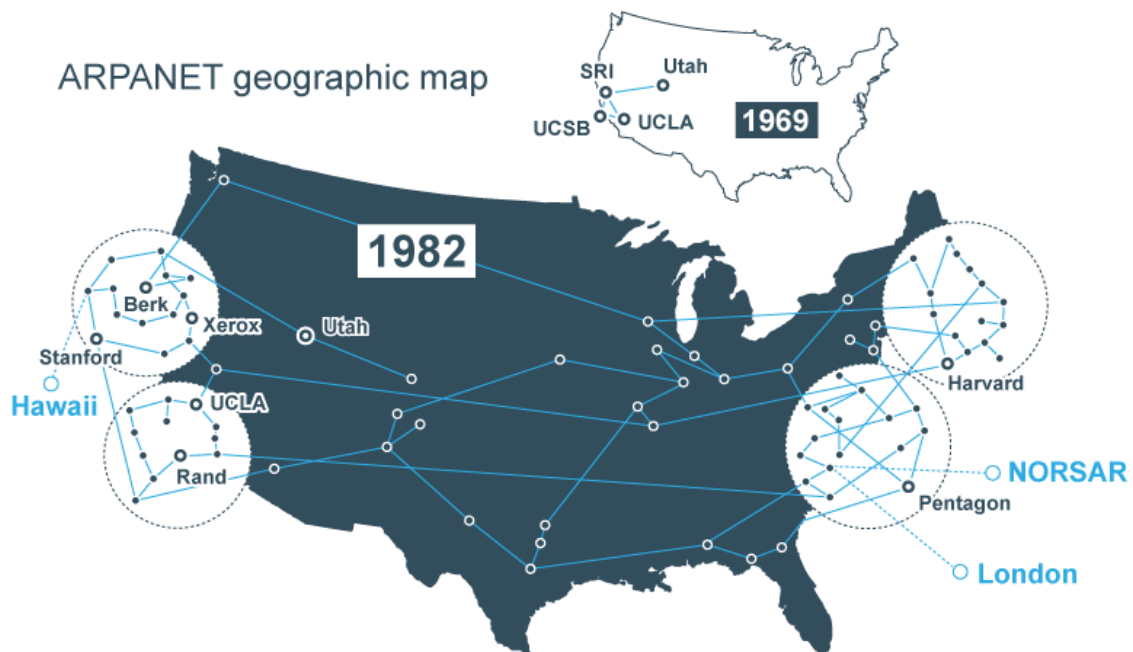


Abbildung 1. Die Evolution des ARPANET von 1969-1982 (Leyden, 2019)

Das gesellschaftliche Umdenken von Informationsbeschaffung und Kommunikation nach dem breitflächigen Einsatz des ARPANET beschreibt Abbate folgendermassen:

Before, people had typically shared computer resources within a single campus; with the ARPANET and Internet they could access resources across the country or around the world. Network services such as electronic mail, news, and bulletin boards allowed people to create virtual communities [...] that spread from academic and business uses to the personal realm. By the late 1980s networking had become a household word [...]. (1994, S. 6)

Ein zentraler Faktor für das Wachstum war der offene Zugang (Stichwort «open access» in heutigen akademischen Kreisen) zu den Protokollspezifikationen und der Möglichkeit, diese für eigene Projekte oder Softwareentwicklungen zu implementieren. Wie Leiner et al. (2009, S. 28) zusammenfassen, war die primäre Rolle des Netzwerks nicht der generelle Informationsaustausch, sondern der Austausch über den Aufbau und Betrieb des Netzes selbst. In den 1990er Jahren wurde ARPANET offiziell beendet und an die

National Science Foundation abgegeben (NSFnet), dem sich diverse Länder anschlossen und ein weltweites Wachstum des Internets auslöste (Cohen-Almagor, 2011, S. 53). Seither sah das Internet grundlegende technologische Veränderungen und Erweiterungen mit exponentiell zunehmender Geschwindigkeit. Das grundlegende Konzept des ARPANET, die Verbindung mehrerer Netzwerke als Ganzes sowie der Transfer von Paketen, ist jedoch noch immer vorhanden. Hatta (2020, S. 285–286) bricht die heutige Funktionsweise des Internets folgendermassen herunter: Der Zugang wird durch die Korrespondenz anhand einer hohen Anzahl an Nachrichten zwischen einem Client – dem Browser und/oder Gerät des Benutzers – und einem Server bewerkstelligt. Geschieht ein Aufruf einer Webseite, wird ein GET-Request (Anforderungsnachricht für den Seiteninhalt) mittels Protokollspezifikationen vom Client an den Server geschickt. Der Webserver sendet anschliessend die angeforderten Daten an den Client. Zu diesem Zeitpunkt wird die Nachricht minutiös in Daten einer bestimmten Grösse (Pakete) unterteilt. An jedes Paket werden Kontrollinformationen in einem «Header» angehängt, die unter anderem am Anfang des Headers die IP-Adresse des Absenders und des Ziels beinhalten. Die paketempfangende Seite stellt eine Verbindung her, rekonstruiert die Nachricht und ergreift entsprechende Massnahmen. Solche Pakete werden im Internet in einem Paket-Relay über mehrere Server an den Zielserver gesendet – dieser Paketfluss wird als «Verkehr» bezeichnet. Das Lesen des Headers und die Entscheidung, wohin und wie das Paket zu senden ist, nennt sich «Routing». Ein «Router» ist dementsprechend das Gerät oder die Software, die jene Entscheidungen trifft und die Pakete übermittelt.

Das Internet ist demnach zusammengefasst ein Netzwerk von Netzwerken, das Computer weltweit miteinander verbindet und eine Kommunikation zwischen den Computern anhand von Internetprotokollen, Satelliten, Telefonleitungen und Kupferleitungen oder Glasfaserkabeln ermöglicht. Für das Internet gibt es keine zentrale Verwaltung, weder bei der technischen Umsetzung noch bei den Zugangs- und Nutzungsrichtlinien; jedes einzelne Netz legt seine eigenen Richtlinien fest (Kavallieros et al., 2021, S. 4).

2.1.2 Taxonomie des Internets

In seiner Definition ist das Internet die Infrastruktur verschiedener Netzwerke – Kabel, Router, Server, Clients und alle Regulationen, Sprachen und Protokolle –, während das WWW ein auf dem Internet aufbauender Dienst aus miteinander verknüpften Webadressen ist (MDN Web Docs, 2023). Der Einfachheit halber und zur besseren Unterscheidung zu den drei Teilbereichen des Webs wird in den nachfolgenden Ausführungen mit dieser

Definition des Internets gearbeitet, obschon umgangssprachlich der Begriff des Internets als Synonym zum World Wide Web verstanden wird.

Wie in Abbildung 2 aufgeführt lässt sich das Web in drei Oberkategorien unterteilen: Das Surface Web, Deep Web und das Dark Web. Während das Surface Web eine klare Definition aufweist und vergleichsweise verständlich erschlossen ist, existieren für das Deep Web mittlerweile mehrere Begriffsbestimmungen und unterschiedliche Deep Webs – sofern im Detail betrachtet. Das Dark Web geht dabei noch eine Stufe tiefer und lässt sich nicht als Ganzes definieren, sondern bedarf einer Betrachtung der unterschiedlichen Arten von Dark Webs und deren geschätzten Verbreitung. Aufgrund der Popularität des Tor-Netzwerks und dessen Erwähnung in der Literatur in Zusammenhang mit der Privatsphäre in Bibliotheken wird dieses Dark Web genauer erklärt.

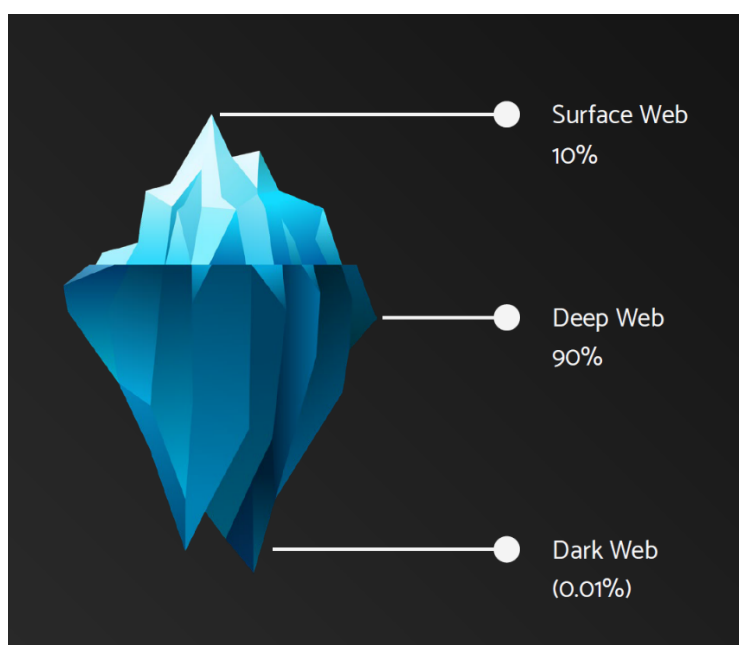


Abbildung 2. Internetverkehr der drei Teilbereiche des Webs (eigene Darstellung nach Chertoff (2017), Eisberg: Lizenz Adobe Stock)

Diverse Quellen aus den 2000er-Jahren berichten über unterschiedliche Grössendimensionen der drei Webs, von 84-99% des Internets zum Deep Web gehörend, wobei die genaue Grösse undefiniert gelassen wird (Bergman, 2000; Devine & Egger-Sider, 2014, S. 15; Lyman & Varian, 2003, S. 11). Wie Finklea ausführt, wurde das Deep Web in frühen Annahmen auf 4000-5000-mal grösser als das Surface Web geschätzt; die sich verändernde Dynamik des Informationszugangs und der Informationsdarstellung bedeutet jedoch, dass das Deep Web mit einer Geschwindigkeit wächst, die sich nicht quantifizieren lässt (2017, S. 2–3). Chertoff führt anstatt der Grösse den Internetverkehr zu 90% auf das Deep Web zurück, mit einem Anteil von 0.01% des Dark Webs (2017, S. 27). Die

Abbildung 2 stützt sich demnach aufgrund des Fehlens von aktuellen Zahlen auf die Schätzung des jeweiligen Volumens des Internetverkehrs.

2.1.2.1 Surface Web

Die zugänglichste und in der Gesellschaft am weitesten verbreitete Schicht des Webs wird gemäss dem gängig illustrierten Modell des Eisbergs als die oberste und sichtbare Fläche bezeichnet. Das Surface Web – auch Clearnet, Visible Web oder Open Web genannt (Prasad, 2017, S. 418) – wird demnach von der westlichen Durchschnittsperson im Alltag verwendet; zu Unterhaltungs-, Informations- oder Arbeitszwecken. Bestehend aus einer Sammlung von Webseiten wird es durch Standardsoftware wie einem Webbrowser zugänglich und durch Suchmaschinen wie Google indexiert und durchsuchbar gemacht (Hawkins, 2016, S. 4).

Alshammari definiert eine Webseite als ein in HTML oder einer anderen Markup-Sprache geschriebenes Dokument, das unterschiedliche Arten von Webressourcen beinhalten kann; dies können Stilinformationen oder Scripts zur Veränderung des Aussehens und der Dynamik einer Webseite sein und Medien wie Bilder, Ton und Videos (2017, S. 10). Das Laden einer Webseite erfordert das Abrufen der HTML-Datei und allen eingebetteten Medien beim Webserver (vgl. Kap. 2.1.1).

Die erste offizielle Webseite wurde am 30.04.1993 vom CERN veröffentlicht und auf Tim Berners-Lees NeXT-Computer gehostet. Sie beinhaltete grundlegende Informationen zum Projekt des World Wide Web, wie der Zugriff zu Dokumenten von anderen Personen funktionierte und wie ein eigener Server aufgesetzt werden konnte (Giampietro, 2013). Bis im Jahr 2000 wuchs die Zahl der Webseiten auf 17 Millionen und stieg exponentiell – im 2021 betrug die Anzahl der sich im Surface Web befindenden Webseiten 1.88 Milliarden (Armstrong, 2021).

Gekennzeichnet wird das Surface Web durch dessen Vernetzung, Indexierung und Durchsuchbarkeit mithilfe etablierter Suchmaschinen. Um den Unterschied zur darunterliegenden Schicht des Deep Webs aufzuzeigen, sind Kenntnisse der Funktionsweise einer Suchmaschine vonnöten. Aufgrund eines Suchmaschinen-Marktanteils von 83.94% (Lohmeier, 2023) werden die nachfolgenden Komponenten anhand von der Google-Suche erläutert.

2.1.2.1.1 Die Google-Suche

Google stellt einen umfassenden Leitfaden zur Funktionsweise der Google-Suche zur Verfügung (2023). Zusammengefasst und vereinfacht dargestellt erscheinen die Ergebnisse nach der Eingabe einer Suchanfrage anhand diverser Faktoren:

- **Crawling:** Um Seiten im Web aufzufinden und zu indexieren, setzt Google den Crawler «Googlebot» ein. Dieser durchsucht das Web konstant anhand von Verlinkungen bestehender Seiten und lädt dessen Inhalt herunter. An dieser Stelle geschehen bereits Abgrenzungen zum Deep Web: Das Crawling kann einerseits durch den Site Owner durch eine Robots.txt-Datei gezielt verhindert werden, oder der Seiteninhalt ist aufgrund der Notwendigkeit einer Anmeldung nicht sichtbar (vgl. Kap. 2.1.2.2.).
- **Indexierung:** In der Phase der Indexierung werden anschliessend die gecrawlten Seiten anhand deren Inhalt analysiert, auf Duplikate überprüft und dem Index (einer Art Datenbank) hinzugefügt. Ähnliche Seiten werden in einem Cluster gesammelt, um daraus die repräsentativste Seite als Suchergebnis anzuzeigen (Kanonisierung). Während der Indexierung werden weitere Metadaten gesammelt, beispielsweise die geografische Herkunft oder die Nutzerfreundlichkeit. Jedoch ist die Indexierung einer Seite nicht gewährleistet.
- **Suchergebnis:** Wird eine Suchanfrage («search query») getätigt, bestimmt Googles Algorithmus aus den indexierten Webseiten die Ergebnisse mit der wahrscheinlich höchsten Relevanz. Die Reihenfolge («ranking») wird anhand diverser Faktoren wie Sprache, Standort, Gerät des Nutzers und Qualität der Webseite bestimmt. Zum besseren Verständnis der Semantik einer Suchanfrage wird ausserdem die KI «BERT» (Bidirectional Encoder Representations from Transformers) eingesetzt.

Grundsätzlich kann das Ranking einer Webseite aus Sicht eines Site Owners mithilfe von SEO (Search Engine Optimization) verbessert werden. Dazu gehört beispielsweise die Platzierung von inhaltlich relevanten Stichworten im Titel, im Textkörper, im Alt-Text oder in Linktexten.

2.1.2.1.2 Privatsphäre im Surface Web

Mit der Nutzung des Internets und des Webs im Alltag ist die Privatsphäre während dem «Surfen» ein wiederkehrendes Thema. Inwiefern Google oder diverse Anbieter von Webseiten Daten über ihre Nutzenden sammeln und weiterverkaufen, ist nur bedingt der breiten Öffentlichkeit bekannt. Dennoch wird ein reger Handel mit persönlichen Daten betrieben: «Personal data has become a new source of economic value. Once processed and classified they provide relevant information for companies about people's interests and activities [...]» (Esteve, 2017, S. 36). Nebst der Einhaltung der länderspezifischen Gesetze zur Verarbeitung von personenbezogenen Daten gibt es keinen Hinderungsgrund, ein Unternehmen jener Art zu betreiben. Esteve führt jedoch weiter aus, dass sowohl Google als auch Facebook für ihre Nachverfolgung von Nutzendenverhalten für

Marketingzwecke – zum Beispiel personalisierte Anzeigen – auf Kritik bezüglich der Wahrung von Privatsphäre stossen (2017, S. 37).

Die Nutzung des Surface Webs ist mit der Sammlung von Informationen durch Webseiten und Suchmaschinen anhand der Informationen der Nutzenden verbunden. Jene Informationen werden zur Erstellung von Profilen verwendet, die einerseits Marketingzwecken, andererseits der Verbesserung der User Experience dienen (Ruiz-Martínez, 2012, S. 1475). Grundsätzlich können Daten auf drei verschiedenen Ebenen gesammelt werden: TCP/IP, HTTP und in der Anwendungsschicht. Daraus können die IP-Adresse (inklusive Domain-Name, Standort und ISP), verwendeter Webbrowser, zuvor besuchte URL und E-Mail-Adresse gezogen werden. Ausserdem können in der Anwendungsschicht eingebettete Objekte (wie JavaScript, Plug-Ins, Werbung, etc.) zu Zwecken des «Fingerprinting» – das Identifizieren eines Nutzenden anhand von Browserdaten – missbraucht werden (Ruiz-Martínez, 2012, S. 1475–1477). Als eine mögliche anwenderfreundliche Lösung für Bedenken hinsichtlich der Privatsphäre schlagen unter anderem Alshammari und Ruiz-Martinez den Browser Tor vor (vgl. Kap. 2.1.2.3 und 2.1.4) (2017; 2012, S. 1478).

2.1.2.2 Deep Web

Im Auftrag von BrightPlanet war Michael K. Bergman der erste, der den Begriff «Deep Web» in seiner Studie aus dem Jahr 2000 verwendete und definierte. Bergman beschreibt die Entdeckung des Deep Webs als signifikant aufgrund der Qualität der darin gefundenen Ressourcen:

There are literally hundreds of billions of highly valuable documents hidden in searchable databases that cannot be retrieved by conventional search engines. [...] If the most coveted commodity of the Information Age is indeed information, then the value of deep Web content is immeasurable. (S. iii)

Das Deep Web kennzeichnet sich laut Bergman ausserdem dadurch aus, dass dessen Inhalt nur auf direkte Anfragen und dynamisch dargestellt wird, während das Surface Web dank des Crawlings und der Indexierung mithilfe unabhängiger Suchmaschinen durchsuchbar ist (vgl. Kap. 2.1.2.1.1) (2000, S. 24). Hatta fasst die von Bergmans Studie durchsuchten Deep-Web-Inhalte in folgende Kategorien zusammen (2020, S. 279): Datenbanken innerhalb einer Organisation, Seiten mit einer Paywall, Seiten mit sich bei jedem Aufruf änderndem dynamischen Inhalt, Seiten mit zwingend zu nutzendem integrierten Suchsystem, E-Mail und Chats.

Seither hat die generelle Definition des Deep Webs eine neue Sicht erhalten. In der Literatur wird jedoch noch immer mit beiden Definitionen gearbeitet, um das Deep Web zu veranschaulichen. Devine und Egger-Sider nennen die klassische technologische Definition und die kognitive Definition (2014, S. 4–15):

Die technologische Definition deckt sich in der Kernaussage mit der frühen Studie von Bergman (2000). Jedoch vermerken Devine und Egger-Sider zusätzlich, dass mit der Entwicklung einer neuen Suchmaschine gleichzeitig ein neues Deep Web erschaffen wird: «Each search engine creates its own Invisible Web consisting of all the information resources that it does not include in its indexing.» (S. 5). Die Grösse des Indexes einer Suchmaschine steht ausserdem in einer negativen Korrelation zu deren Informationstiefe; je mehr Inhalt eine einzelne Seite aufweist, desto weniger genau ist diese im Index repräsentiert (Wouters, Reddy & Aguillo, 2006, S. 113). Die ausgelassenen Informationen gehören demnach ebenfalls zum Deep Web. Ebenso rein technische Aspekte wie veraltete oder neue Formate, die die Suchmaschine nicht mehr unterstützt oder noch nicht adaptiert hat, oder die Zeit zwischen den Crawls einer Webseite und deren allenfalls nicht bemerkten Änderungen. Seiten ohne Verlinkung von anderen Webseiten sind ausserdem unauffindbar für den Crawler und nur durch Kennung der URL aufrufbar – somit gehören diese zum Deep Web. Für letzteren Fall kann jedoch eine manuelle Anfrage zur Aufnahme der Seite in den Index bei der entsprechenden Suchmaschine getätigt werden.

Die kognitive Definition richtet den Fokus weg von der Technik hin zur Wahrnehmung des Menschen im Prozess der Informationsbeschaffung. Sie wird unter dem von Ford und Mansourian definierten Begriff der «Cognitive Invisibility» verwendet, der das Deep Web bei den Kenntnissen und Recherche-Fähigkeiten einer Person ansetzt (2006, S. 590). Grundsätzlich ist die Cognitive Invisibility also all jenes, das eine recherchierende Person nicht auffinden kann. Gründe dafür können schlichtes Unwissen, aber auch falsch eingegebene Suchbegriffe sein. Das kognitive Deep Web ist an einen einzelnen Menschen gebunden und unterscheidet sich von Person zu Person. Devine und Egger-Sider führen aus, wie Recherchierende von der Anzahl Suchergebnissen in konventionellen Suchmaschinen bereits überfordert oder zufrieden mit den erhaltenen Suchergebnissen sein können – somit ist das Bedürfnis von Informationen aus anderen Quellen nicht vorhanden (2014, S. 13). Auch der Faktor der Gewohnheit ist miteinzuberechnen; selbst wenn das Wissen über Informationsressourcen von besserer Qualität vorhanden ist, kann zur schnellen Informationsbeschaffung wieder zur gewohnten Suchmaschine gegriffen werden. Die Suche im Deep Web dagegen ist aufwendiger und erfordert den Einsatz verschiedener Suchstrategien und -tools mit entsprechenden Rechercheskills. Generell wird der geringen Nutzung des Deep Webs jedoch das Unwissen darüber

zugeschrieben und die Lösung anhand von Weiterbildungen vorgeschlagen, um diese Lücke zu wertvolleren Informationsressourcen zu schliessen (Devine & Egger-Sider, 2014, S. 23).

2.1.2.3 Dark Web

Aufgrund der Vielzahl an illegalen Aktivitäten und der anonymen Natur des Dark Webs steht dieser Schicht der Web-Landschaft die grösste mediale Aufmerksamkeit zu: «The darknet(s) as its name depicts recalls images of shadowy alleys, malicious, hard-faced individuals and socially damaging activities, covering a range from political protestors – rebels – to drug dealers, to terrorist and gun dealers, to paedophiles and everything in between.» (Kavallieros et al., 2021, S. 7). Jedoch werden auch diverse neutral bis positiv gewertete Aktivitäten damit verrichtet: Die Vermeidung von Identitätsdiebstahl oder Marketing-Tracking, die Umgehung von Zensur und Betreibung von Forschung in Ländern mit kritischen Ansichten zu gewissen Thematiken.

Das Dark Web ist ein Teil des Deep Webs (vgl. Kap. 2.1.2.) und macht gemäss Chertoff 0.01% des Internetverkehrs aus (2017, S. 27). Es ist allerdings unklar, wie gross der Anteil des Dark Webs am Deep Web ist und wie viel des Inhalts legale oder illegale Aktivitäten sind (Finklea, 2017, S. 3). Im Gegensatz zum Deep Web, das mithilfe eines regulären Webbrowsers je nach Zugangsberechtigungen aufrufbar ist, ist das Dark Web gänzlich unzugänglich ohne dafür vorgesehene Software. Webseiten im Dark Web tendieren zudem zur Unzuverlässigkeit: Sie kommen und gehen regelmässig und hinterlassen Verzeichnisse mit nicht funktionierenden Links, sobald die Seite verschwindet oder ihre URL ändert (Coffey, 2020, S. 2).

Innerhalb des Dark Webs residieren mehrere Darknets, die mit unterschiedlichen Technologien und Softwares unabhängig voneinander die dem Dark Web charakteristisch zugehörigen anonymen Netzwerke bilden (Kavallieros et al., 2021, S. 5–7). Am weitesten verbreitet ist das Tor-Netzwerk, zugänglich innerhalb und mithilfe des gleichnamigen Browsers Tor. Es stützt sich auf ein Netzwerk von freiwilligen Computern, die den Internetverkehr der Nutzenden durch eine Reihe von Computern anderer Nutzenden leiten, sodass der Verkehr nicht zum ursprünglichen Nutzenden zurückverfolgt werden kann (Weimann, 2016, S. 198). Die Software ist im Surface Web frei zugänglich und kann via Google und der Webseite des Tor Projects heruntergeladen werden. Die Nutzung von Tor an sich gestaltet sich als regulärer Browser und kann auch ohne Absicht des Zugriffs auf das Darknet in Verwendung sein. Im Unterschied zu Browsern wie Chrome oder Edge werden jedoch Suchergebnisse in Suchmaschinen ungefiltert geliefert (d.h. die Erstellung von Profilen durch die Suchmaschine ist gehindert), da aufgrund der Funktionsweise

von Tor bei jeder verbundenen Sitzung (s. Abb. 3) eine neue Identität zugewiesen wird (The Tor Project, Inc., 2023a).

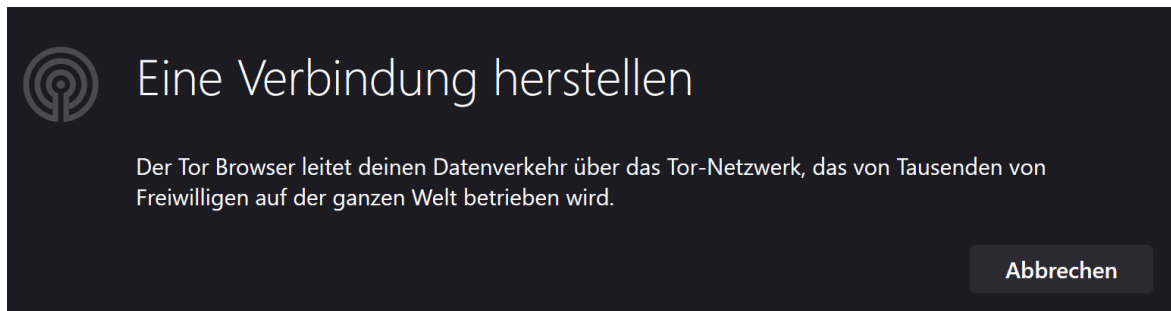


Abbildung 3. Verbindungsaufbau einer Sitzung im Tor-Browser (Screenshot Tor-Browser)

Zum besseren Verständnis des Zusammenhangs von Tor und der Privatsphäre-Diskussion des Surface Webs wird nachfolgend die Entstehung und Funktionsweise von Tor erläutert.

2.1.2.3.1 Geschichte von Tor

Das Darknet des Tor-Netzwerks hat seinen Ursprung in den Anfängen von ARPANET (vgl. Kap. 2.1.1). Mit dem Wachstum der Computer-Interaktionen ergaben sich vermehrt isolierte und geheim gehaltene Netzwerke parallel zu ARPANET (Lacson & Jones, 2016, S. 42 zitiert nach McCormick, 2013, S. 22). Das U.S. Naval Research Laboratory erkannte in den 90er-Jahren die Gefahr der Überwachung des Internets und fragte sich, inwiefern eine Verbindung zwischen Computern möglich ist, ohne Absender und Empfänger identifizieren zu können (The Tor Project, Inc., 2023b). Das NRL machte schließlich von den isolierten Netzwerken Gebrauch und entwickelte erste Ansätze des «onion routing», die Technologie des Tor-Browsers im Zwiebelschicht-Prinzip: «The goal of onion routing was to have a way to use the internet with as much privacy as possible, and the idea was to route traffic through multiple servers and encrypt it each step of the way.» (The Tor Project, Inc., 2023b). Tor basiert noch heute auf diesem Grundsatz des Routings (vgl. Kap. 2.1.2.3.2) und erhielt dadurch auch die Abkürzung des Namens von «The Onion Router».

Das NRL nutzte Tor zum Schutz von amerikanischen Agenten und Dissidenten in Übersee und erhielt im Jahre 2004 Finanzierungen durch die EFF. Damit einher ging jedoch auch die öffentliche Verfügbarkeit der Software – Tor-Domains zu Drogenhandel, Kinderpornografie und Terrorismus kamen auf (Lacson & Jones, 2016, S. 42). Kurz darauf wurde die Non-Profit-Organisation «The Tor Project» gegründet; Bestrebungen zur Möglichkeit der Umgehung von Regierungs-Firewalls und der damit verbundenen Zensur in einigen Ländern führten zur Popularität von Tor bei Aktivisten und Privatsphäre-

liebenden Personen. Da es aus technischer Sicht jedoch umständlich war, das Tor-Netzwerk zu benutzen, startete im Jahr 2008 die Entwicklung des Tor-Browsers. So spielte Tor beispielsweise im Arabischen Frühling eine zentrale Rolle, um der Bevölkerung Zugang zu zensierten Informationen und Medien zu ermöglichen (The Tor Project, Inc., 2023b). Auch haben Mitglieder des Tor Projects den Browser auf Computern in Schutzeinrichtungen für häusliche Gewalt in den USA installiert, um ihre Bewohner*innen und den Standort der Einrichtungen zu schützen (Macrina & Phetteplace, 2015, S. 18). Für diese und weitere Engagements gewann Tor im Jahre 2010 die Auszeichnung für Projekte mit sozialem Nutzen der Free Software Foundation (Coffey, 2020, S. 4).

Spätestens nach dem Whistleblowing durch Edward Snowden bei der NSA wuchs das Bedürfnis nach Tools zum Schutz vor Massenüberwachung in der breiten Bevölkerung (The Tor Project, Inc., 2023b).

2.1.2.3.2 Funktionsweise von Tor

Gemäss dem Tor-Browser-Handbuch hat die Nutzung von Tor mehrere Hauptmerkmale (2023a):

- Der Internet Service Provider kann die Internetaktivität nicht zurückverfolgen. Dies schliesst die Namen und URLs der besuchten Webseiten mit ein.
- Ohne ausdrückliche und gewollte Enthüllung der Identifikation sehen Webseitenbetreiber und Beobachter eine Verbindung des Tor-Netzwerks, nicht die eigentliche IP-Adresse. Das Ziehen von Rückschlüssen auf eine Person (oder einen Client) ist somit nicht möglich.
- Das Fingerprinting (vgl. Kap. 2.1.2.1.2) oder die Identifizierung anhand der Browser-Konfiguration wird mit Tor gehindert.
- Tor speichert keine Suchverläufe.
- Cookies sind nur für die aktuelle Sitzung bzw. die aktuelle Identität gültig.

Möglich wird dies durch das Senden des Internetverkehrs über mindestens drei zufällige Server im Tor-Netzwerk, genannt «Relays» oder «Nodes» (Knoten). Dieser Vorgang wird als «Tor Circuit» bezeichnet. Im ersten Relay gelangen die Daten des Nutzers in den Circuit, indem sich Tor mit einem zufälligen und öffentlichen Eingangsknoten verbindet. Im zweiten Relay werden die Daten verschlüsselt und durch eine Reihe von Knotenpunkten gesendet. Um die Anonymität zu gewährleisten, kennt jeder Knoten nur die Identität des vorhergehenden und nachfolgenden Knotens. Der letzte Server, der «Exit Relay», entschlüsselt die Daten und schickt den Verkehr vom Circuit in das öffentliche Internet zum Zielsystem (Ngo, Marcum & Belshaw, 2023, S. 163). Im 2022 griff Tor noch auf rund 7000 Relays zurück und verzeichnete eine Nutzendenzahl von 2.5 Millionen (Kumar,

Akash, Arun & Dixit, S. 2214), mit Stand Juli 2023 sind es bereits knapp 5 Millionen Nutzende mit durchschnittlich gleichbleibender Anzahl an Relays (The Tor Project, Inc., 2023e).

Ngo et al. betonen jedoch, dass Tor keine komplette Anonymität gewährleisten kann (2023, S. 163): Der ISP kann noch immer sehen, dass der Nutzende Tor im Einsatz hat. Ebenso kann Tor die Eingangs- und Ausgangsknoten im Surface Web nicht verschleiern – die Serverbetreiber dieser Knoten können die IP-Adresse des Nutzenden identifizieren. Zusätzlich ist Tor aufgrund des Onion Routings vergleichsweise langsam.

Das Darknet innerhalb des Tor-Netzwerks basiert jedoch nicht auf der Verbindung zu Servern im Surface Web, sondern besteht aus den «Onion Services» (früher «Hidden Services» genannt). Steinebach, Schäfer, Karakuz, Brandl und Yannikos beschreiben die Onion Services als Online-Applikationen, die im Vergleich zu regulären Surface-Web-Internetservices anonym betrieben und aufgerufen werden können und ohne Tor nicht zugänglich sind (2019, S. 1). Wie bereits in Kapitel 2.1.2.3.1 aufgegriffen, werden die Tor-Domains bzw. die Onion Services zu unterschiedlichen Zwecken benutzt: Organisationen und Presseorgane wie die New York Times oder The Guardian machen zur anonymen Kommunikation mit Whistleblowern davon Gebrauch, andere umgehen damit die Zensur oder betreiben einen Informationsaustausch in Ländern mit hoher staatlicher Überwachung, und wieder andere nutzen sie zum Handel von illegalen Waren und Inhalten (Steinebach et al., 2019, S. 1).

Als Ergänzung zur Erklärung von Steinebach et al. (2019, S. 1) bieten gemäss dem Tor-Browser-Handbuch die Onion Services die folgenden Vorteile gegenüber regulären Internetdiensten (The Tor Project, Inc., 2023c):

- Der Standort und die IP-Adresse eines Onion Services ist verschleiert, sodass sich die Zensur oder die Identifizierung der Betreiber für Widersacher als schwierig gestaltet.
- Der Verkehr zwischen Tor-Nutzenden und Onion Services ist Ende-zu-Ende-verschlüsselt (E2EE). Eine Verbindung via HTTPS ist damit hinfällig.
- Die Adresse eines Onion Services wird automatisch generiert – Betreiber müssen keine Domännennamen erwerben. Die URL besteht dabei aus 56 Buchstaben und Zahlen und endet auf «.onion». Sie dient Tor unter anderem dazu, eine korrekte und unverfälschte Verbindung sicherzustellen.

Onion Services werden durch Tor differenziert gehandhabt und verlassen den Tor-Circuit nicht durch Ausgangsknoten, der Verkehr bleibt also innerhalb des Circuit (s. Abb. 4). Dadurch können Sicherheitsprobleme wie die Überwachung des Datenverkehrs durch

Ausgangsknoten (die Schwäche von Tor im Surface Web) verhindert werden (Çalışkan, Minárik & Osula, 2015, S. 11).

Diverse Social-Media-Plattformen und Presseorgane betreiben gespiegelte Onion Services ihrer Seiten für den Zugang aus zensurierten Ländern, darunter Facebook, BBC News und The New York Times (BBC, 2019; Facebook, 2023; Sandvik, 2017). Die Onion-Seite ist dabei ein exaktes Replika der Seite im Surface Web, sowohl optisch als auch in der Bedienung. So lautet Facebooks Onion-Adresse beispielsweise:

facebookkwkhpilnemxj7asaniu7vnjjbiltxjqhye3mhbshg7kx5tfyd.onion

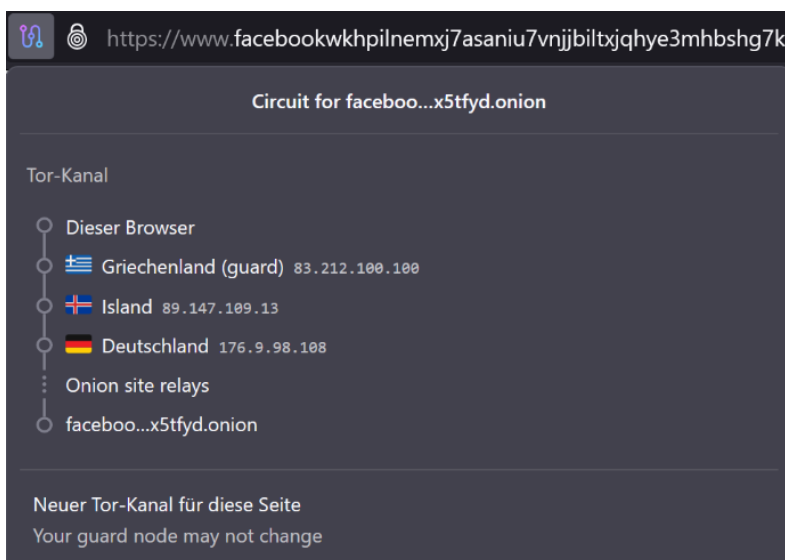


Abbildung 4. Circuit der Onion-Adresse von Facebook (Screenshot Tor)

Die Tor-Metrics zeigen mit Stand 22. Juli 2023 eine sichtbare Anzahl Onion Services von 758'625 an (The Tor Project, Inc., 2023d). Gegenüber den 27'424 Onion Services am 25.12.2014 (erster Tag der statistischen Verfügbarkeit der Services Version 2) entspricht dies einem Wachstumsfaktor von gerundet 28.

Zusammengefasst kann Tor somit sowohl als regulärer Browser mit erhöhter Privatsphäre und Konfigurationsmöglichkeiten im Surface oder Deep Web operiert werden, oder aber auf die isolierten Onion Services im Darknet des Tor-Netzwerks zugreifen.

2.1.3 Deep Web im Bibliothekskontext

Intellectual freedom, privacy, and unrestricted access to information remain the core values of information professionals. Libraries function as community education centers serving diverse populations with differing information needs. Information seeking using digital platforms requires specialized skills and tools to extract applicable results. (Coffey, 2020, S. 1)

Um Informationen effektiv und gewinnbringend aus dem Deep Web ziehen zu können, bedarf es spezialisierter Fähigkeiten, Tools und Techniken. Informationsspezialisten und Bibliothekar*innen spielen dabei als Experten der Informationsbeschaffung eine zentrale Rolle und können sich in der Hochschul- und Forschungsgemeinschaft entsprechend mit Rechercheunterstützung positionieren. Dafür benötigen Bibliothekar*innen das Verständnis des Konzepts und der Charakteristiken des Deep Webs, um bessere Suchergebnisse über die Google-Suche hinaus zu liefern (Prasad, 2017, S. 420).

Prasad definiert drei Punkte, die Bibliothekar*innen vor der Formulierung eines Suchstrings beachten sollten (2017, S. 420–422):

Kategorien der verfügbaren Informationen im Deep Web	
Sondersammlungen	Juristische, akademische und Forschungsdatenbanken, Datenbanken für Wissenschaft, Technik, Medizin, Bibliothekskataloge, etc.
Proprietäre Datenbanken	Kommerziell und der Regierung angehörig
Intranets	Organisatorische und Unternehmensseiten
Soziale Medien	Kommunikation, Profile und Vernetzungen
Merkmale von Deep Web Ressourcen	
Inhalt	Der Inhalt der Datenbank wird dynamisch generiert und kann nicht über eine Suchmaschine gefunden werden. Z.B.: Bibliothekskatalog, DOAJ
Gebühren	Ein Grossteil der kommerziellen Datenbanken verlangt eine gesamthafte Lizenz oder Gebühr pro Artikel.
Login	Einige Seiten verlangen ein reguläres Login oder eine IP-Authentifizierung, um Zugang zu erhalten. Der Inhalt dahinter kann kostenfrei oder gebührenbasiert sein.

Echtzeit	Informationen über aktuelle Ereignisse können von den üblichen Suchmaschinen nicht durchsucht werden.
Nicht verlinkte Webseiten	Führen keine Links auf eine Webseite, kann sie von Crawlern nicht gefunden und indexiert werden.
Neuer Inhalt	Die Zeit zwischen der Aufschaltung einer Webseite oder dem Hinzufügen neuen Inhalts.
Robots.txt-Dateien	Webseiten-Betreiber können Crawler bewusst am Indexieren ihrer Seite hindern.
Inhaltsreiche Webseiten	Einige Suchmaschinen nehmen inhaltsreiche Seiten nur teilweise auf. Je mehr Inhalt, desto weniger wird indexiert.
Formate	Veraltete Formate können nicht mehr unterstützt werden, neue Formate können noch nicht aufgenommen sein.
Strategien zur Deep-Web-Suche	
Für ein breites Thema: Standardsuchmaschine verwenden	
Für ein spezifisches Thema: Durchsuchbare Fachdatenbank verwenden	
Gewisse Datenbanken bieten Bibliothekar*innen mittels Registration kostenfreien Zugang an.	
Erstellen einer Liste für fachspezifische Webseiten von Institutionen und Organisationen, z.B. Regierungswebseiten, Forschungs- und Entwicklungsinstitute, Universitäten, Berufsverbände, Stiftungen und Verzeichnisse.	
Gezielt nach Webseiten mit verwandten Themen suchen, um spezifische Informationen zu erhalten.	
Vorschläge von Fachpersonen und Empfehlungen für Fachinformationen einholen, wenn das Thema zu spezifisch ist.	

Tabelle 1. Wissenswertes für Bibliothekar*innen vor einer Suche im Deep Web (nach Prasad, 2017, S. 420–422)

Wie Coffey aufführt besteht ein Bedarf an Informationsspezialisten, die als digitale Guides bei der Navigation durch die Informationsressourcen dienen (2020, S. 2). Da die meisten Bibliotheken und Universitäten ihren Nutzenden eine Vielzahl an Datenbanken zur Verfügung stellen, ist es für Informationsfachleute von hoher Wichtigkeit, das Deep Web und seine Inhalte zu verstehen und das Wissen über dessen Funktionen für die akkurate Findung von Information zu besitzen. Inwiefern Bibliothekar*innen diese Rolle bereits abdecken und wie gross die Nachfrage danach tatsächlich ist, geht aus der Literatur nicht eindeutig hervor. In Kapitel 3 und 4 werden die Untersuchung und Ergebnisse ebendieser Frage betrachtet.

Mit den von Devine und Egger-Sider (2014, S. 13) diskutierten Punkten zur Problematik, weshalb das Wissen und die aktive Nutzung des Deep Webs gering ist (Gewohnheit der Nutzenden an die Google-Suche und Lieferung von schnellen Antworten) (vgl. Kap. 2.1.2.2), stellt sich gleichzeitig die Frage: Weshalb sind diese Ressourcen von Bedeutung und wieso braucht es geschulte Personen, die sie systematisch aufdecken und das Wissen dazu weitergeben können? Bergman (2000) analysierte als erster und letzter das Deep Web in seiner quantitativen Studie und verglich damals vorhandene Webseiten auf dem Surface Web mit Seiten im Deep Web. Spätere Versuche in dieser Grössenordnung und Flächenabdeckung wurden aufgrund des exponentiellen Wachstums des Webs nicht unternommen. Eines der Studienziele war dabei die Qualitätsmessung der Seiten in fünf unterschiedlichen Fachgebieten: Landwirtschaft, Medizin, Betriebswirtschaft/Finanzen, Naturwissenschaften und Recht (S. 12). Durchschnittlich resultierte aus der Analyse eine dreimal höhere Qualität pro Dokument, mit einer tausendfach höheren Gesamtqualität des Deep Webs gegenüber dem Surface Web (S. 13). Ausserdem verzeichneten mediane Webseiten im Deep Web einen durchschnittlich 50% höheren Verkehr und wurden knapp doppelt so häufig verlinkt. Bergman schreibt dazu: «This suggests that well-known deep Web sites are highly popular, but that the typical deep Web site is not well known to the Internet search public.» (2000, S. 14). Das Deep Web verzeichnete in Bezug auf die untersuchten Fachgebiete ausserdem eine stabile inhaltliche Abdeckung, was eine Relevanz für jeden Markt und Informationsbedarf aufzeigt. Bergman betont – um auf die zuvor aufgestellte Frage zurückzugreifen – dass bei der Suche nach einer zwingend erforderlichen Information in Form einer «Nadel im Heuhaufen» das Deep Web eine Quelle von hoher Wichtigkeit darstellt (2000, S. 21). Dass überhaupt die Qualität und nicht die Quantität entscheidend ist, beruht darauf, dass eine effektive Suche sowohl die gewünschten Informationen ermitteln als auch sie in der Reihenfolge ihrer potenziellen Relevanz oder Qualität präsentieren sollte (S. 26).

Insbesondere bei der Recherchearbeit im beruflichen oder akademischen Sinne, die hohe Anforderungen an die Qualität der Ressourcen stellt, sind demnach Suchen ausserhalb von konventionellen Suchmaschinen oftmals unumgänglich. Da Bibliotheken die zentralen Anlaufstellen für lizenzierte Inhalte im Deep Web sind, benötigt es geschultes Personal, das direkt bei Auskunftstheken ansprechbar ist und Auskunft geben kann (vgl. Kap. 2.3).

2.1.4 Dark-Web-Technologien

Wie bereits in Kapitel 2.1.2.3 aufgeführt ist das Dark Web der Sammelbegriff für die darin enthaltenen Darknets, die mit entsprechender Software abweichend von regulären Browsern aufgerufen werden. Nebst dem populärsten Tool Tor gibt es eine Reihe anderer Softwares und Arten von Darknets – darunter zwei, die von diversen Informationsspezialisten als Mittel zur Erhöhung der Privatsphäre in Bibliotheken empfohlen werden. Diese Diskussion wird nachfolgend ebenfalls aufgegriffen, um die Thematisierung des Webs in der bibliothekarischen Literatur möglichst vollständig abzubilden und in Relation zu den Privatsphäre-Kontroversen im Surface Web zu setzen.

Um das Dark Web weiter einzuschränken, können zwischen zwei Arten unterschieden werden: File-Sharing-Netzwerke (Wood, 2010) und Overlay-Anonymitäts-Netzwerke wie Tor. Bei Overlay-Netzwerken gibt es erneut verschiedene Arten der Vernetzung, von breit verfügbaren und vergleichsweise einfach bedienbaren Softwares bis zu privaten Netzen, die nur über persönliche Bekanntschaften zu Mitgliedern Zugriff gewähren («Friend to Friend») (Hatta, 2020, S. 280–285). Alternativen zum Tor-Darknet sind beispielsweise Hyphanet (Freenet), I2P und Brave Browser. Freenet erhielt im Juni 2023 den neuen Projektnamen Hyphanet (Dougherty, 2023), folglich wird nachfolgend mit dieser Bezeichnung gearbeitet.

I2P (The Invisible Internet Project) ist ein verschlüsseltes Peer-to-Peer-Overlay-Netzwerk, bei dem Start- und Zielpunkt sowie der Inhalt des Verkehrs verschleiert sind (I2P, 2023c). Es wurde gegründet, um Tors Schwäche eines zentralisiert basierten Systems zu überwinden und ein höheres Level an Anonymität und Sicherheit zu gewährleisten. Im Grunde ist die Technologie ähnlich zu Tor: Anonymer Zugang zu Online-Inhalten, Peer-to-Peer-Routing und mehrschichtige Ende-zu-Ende-Verschlüsselung (Negi, 2017, S. 2542). Peer-to-Peer (P2P) ist ein Netzwerk dann, wenn Teilnehmende ein Teil ihrer Hardware-Ressourcen zur Verfügung stellen – sie sind also Gebende sowie Nehmende. Ein Hybrid-P2P erlaubt zentralisierte Entitäten, während ein reines P2P ausschliesslich von den Peers abhängig ist und bei Entfernung eines beliebigen Peers keine Netzwerk-Verluste erleidet (Schollmeier, 2002, S. 101–102). Negi (2017, S. 2542–2543) erklärt den Grundsatz von I2P folgendermassen: Der I2P-Kommunikationsmechanismus beruht stark auf paketbasiertem Routing. Jeder Nutzende im Relay-Netzwerk hat zwei zugehörige Tunnel; einen eingehenden und einen ausgehenden (ein Tunnel ist eine Folge von Peers, die Nachrichten in eine Richtung weiterleiten). Ein Nutzer, der eine Nachricht senden will, leitet diese durch einen ausgehenden Tunnel, während eingehende Nachrichten über den eingehenden Tunnel empfangen werden. Die Länge dieser Tunnel ist wählbar

und bestimmt den Grad der Anonymität und Latenzzeit. Jegliche Daten werden in vier Schichten verschlüsselt – inklusive Endknoten. Diese Art des Routing wird als «Garlic Routing» bezeichnet (Hoang, Kintis, Antonakakis & Polychronakis, 2018, S. 380). I2P muss ausserdem mittels dem «RP installer» manuell installiert werden (im Gegensatz zum Tor-Browser). Sobald der Computer nach der Installation am Netzwerk teilnimmt, fungiert er zeitgleich als Router. So entsteht ein dezentralisiertes und dynamisches Netzwerk. Analog zu Tor bietet auch I2P eigene Hidden Services an – die «Eepsites».

Ein weiteres, jedoch konzeptmässig differenziert zu betrachtendes System ist Hyphanet. Als Browser-Plattform, die das Veröffentlichen, Teilen und Suchen von Inhalten anonym ermöglicht, war das Ziel der Gründung von Hyphanet ein zensurfrees Kommunikationsnetzwerk. Der Verkehr zwischen den Knotenpunkten ist verschlüsselt und wird über mehrere Knoten geleitet, um die Identifikation der Partizipanten zu erschweren und den Inhalt der Kommunikation zu verbergen – dieser Aspekt ist mit Tor und I2P gleichzusetzen. Auch ähnlich wie I2P ist jeder Nutzende verpflichtet, Bandbreite und Speicherplatz der eigenen Hardware für den Dateispeicher freizugeben. Die HDD oder SSD eines Nutzers wird dazu benutzt, Dateien aus dem Netzwerk abzuspeichern und so ein ganzheitlich dezentralisiertes Netzwerk zu schaffen. Jene Dateien sind jedoch verschlüsselt, somit hat die Person keine Einsicht auf den Inhalt der auf ihrem Gerät gespeicherten Dateien. Inhalte werden jedoch nicht unendlich akkumuliert – nicht gefragte Dateien werden automatisch aussortiert und Platz für neue Dateien wird geschaffen. Hyphanet lässt den Nutzenden entscheiden, mit wem eine Verbindung hergestellt werden soll, und bietet einen geschlossenen Netzwerk-Modus an (Verbindung ausschliesslich zu Freunden). Eigene Seiten sind ebenfalls verfügbar, genannt «Freesites» (Negi, 2017, S. 2543).

Die Vorteile von Tor liegen vor allem in dessen Grösse, finanzieller Unterstützung und unter anderem der Anzahl an Softwareentwicklern. Auf dieser Basis kann The Tor Project Probleme effizient lösen und neue Features effektiv einbinden. Tor hat den Fokus auf anonymem Surfen im Surface Web und auf einfachem Zugang zu deren Technologie, dementsprechend spricht es weite Teile der Bevölkerung an und erfordert keine überdurchschnittlichen Kenntnisse der Internet-Architektur. I2P ist konzipiert und konzentriert auf dessen Hidden Services bzw. Eepsites (Aufbau des eigenen Netzwerks), die somit schneller sind als Onion Services. Die Technologie von I2P ist zudem grundlegend auf Anonymität ausgelegt und weniger auf Funktionalität oder Performance (I2P, 2023b). Hyphanet lässt sich nur begrenzt mit I2P und Tor vergleichen, zumal es auf einem gänzlich anderen System basiert und der Fokus auf der Dateispeicherung liegt. Trotzdem gewährt es ein hohes Level an Anonymität und bietet sichere Kommunikationsmöglichkeiten (I2P, 2023a). Sowohl I2P als auch Hyphanet benötigen zur Installation und Teilnahme

am Netzwerk vertieftes Wissen über die Mechanismen der Internet-Architektur und können den Normalnutzenden unter Umständen überfordern (Negi, 2017, S. 2543). Bezüglich Forschung ist Tor ebenfalls Hauptvertreter in der Literatur, während Hyphanet und I2P als Vergleichswerte herangezogen werden. Hoang et al. (2018) untersuchten I2P isoliert und weisen ebenfalls auf die mangelnde Forschung der Tor-Alternativen hin. Als Gründe nennen sie:

First, I2P's purely distributed network architecture, which lacks any centralized directory service, makes it harder to measure. Second, the intermittent availability of exit nodes causes I2P to operate as a self-contained network (which only serves hidden services) most of the time, making it less attractive to users who want to casually browse websites on the public Internet. (Hoang et al., 2018, S. 380)

Für Internet-Normalgebrauchende geht aus dieser Analyse eine klare Favorisierung zur Nutzung von Tor als Privatsphäre-Tool hervor. Auf Basis der unterschiedlichen Projektfokusse von I2Ps und Hyphanets Netzwerken gegenüber Tor ist auch keine Änderung dieses Umstands zu erwarten. Wird Tor also für den Einsatz in Institutionen wie Bibliotheken betrachtet, finden sich vereinzelte Studien und Meinungsäußerungen von Informationsspezialisten dazu. Inwiefern ein solcher Einsatz zielführend sein könnte, wird in Kapitel 5 ausführlicher diskutiert.

Lund und Beckstrom (2021, S. 60) argumentieren, dass nach Ereignissen wie das Datenleck von Facebook in Verbindung mit Cambridge Analytica der Schutz der Daten von Bibliotheksbenutzenden wichtiger denn je sei. Viele Bibliotheken können sich nur limitierten finanziellen und arbeitstechnischen Ressourcen bedienen, was folglich zu verminderten Möglichkeiten im Bereich der Hardware- und Software-Beschaffungen und -Updates führt. Insbesondere OPLs oder kleinere Bibliotheken ohne technischen Support sind gefährdet und auch diejenigen, die potenziell von der kostenfreien Technologie von Tor profitieren. Die Studie vergleicht Leitbilder und Ziele diverser öffentlichen Bibliotheken mit jenen von Tor und dokumentiert klare Überschneidungen, die eine Anregung zur Differenzierung von Tors Ruf der illegalen Inhalte auf den Onion Services und der Technologie an sich anstreben. Universitätsbibliotheken sind nicht explizit erwähnt, zumal diese in Universitäten integriert sind und genügende Sicherheitsmassnahmen der universitären Firewalls verzeichnen können. Auch hier werden beide Seiten von Tor beleuchtet und betont, dass es ursprünglich nicht für illegale Aktivitäten geschaffen wurde: «Illegal content exists everywhere on the Internet, including in areas offered on public library computers [...]. Tor was not designed for crime, just as the Internet in general was

not, it was designed to be an ultra-secure and private network.» (Lund & Beckstrom, 2021, S. 63). Macrina und Phetteplace sagen dazu: «Online anonymity should not be dismissed as something desired only by those with something criminal or nefarious to hide, but as a vital tool for human rights, speech, privacy, security and intellectual freedom. Cue libraries.» (2015, S. 18).

Coffey bringt die Frage der Relevanz mit ein (2020, S. 4): Bibliotheken haben traditionell die Aufgabe, unterschiedliche Nutzende bei der Recherche und Lehre zu unterstützen. Um jedoch in der digitalen Revolution relevant zu bleiben, müssen sie neue Rollen annehmen. Nebst der Informationskompetenz könnte eine davon das Aneignen und Vermitteln von Datenschutzkompetenzen sein: «Producing digital-savvy librarians creates the framework for shaping digital-savvy patrons.» (Coffey, 2020, S. 5).

In den USA wurden bereits Bestrebungen zur Integration von Tor in Bibliotheken unternommen, die von der Regierung (Departement of Homeland Security) ausgehend auf Widerstand stiessen (Brooks, 2015). Unterstützung fand die Bibliothek durch das Library Freedom Project, das sich für Informationsfreiheit und Informationsdemokratie einsetzt (Library Freedom Project, 2023) und eine Kollaboration mit Tor mit Ziel der breitgefächerten Bildung in Privatsphäre-Themen eingegangen ist (The Tor Project, 2015). Darauf folgten weitere Bibliotheken, die diesem Beispiel folgten (Lund & Beckstrom, 2021, S. 65). Jedoch machen auch Lund und Beckstrom darauf aufmerksam, dass eine Integration von Tor einer ausführlichen Prüfung und Planung bedarf, ausgehend von der Art der Bibliothek und den Nutzenden; gegebenenfalls müsste die Option des Tor-Browsers auf gewissen Geräten gehindert werden, beispielsweise in öffentlichen Bibliotheken mit Zugang für Kinder (2021, S. 68). Ein solches Vorhaben zieht ausserdem eine Reihe von Schulungen mit sich, sowohl für Mitarbeitende als auch für Benutzende.

Gemäss dem gemeinsamen Projekt von Tor und Library Freedom sind Bibliotheken die optimalen Entitäten für das Hosting von Endknoten, um das Tor-Netzwerk weiter zu stabilisieren und Möglichkeiten zur Umgehung von Zensur zu verbessern (The Tor Project, 2015). Jedoch gibt es keine kritische Auseinandersetzung mit diesem Vorhaben in der Literatur, die den negativen Aspekt der Stärkung des Tor-Netzwerks aufgreift (vgl. Kap. 5).

Eine Alternative zu Tor ist der Brave Browser, der im Jahre 2016 von zwei ehemaligen Führungskräften bei Mozilla gegründet wurde. Die Funktionen beinhalten integrierte Ad-Blocker, Webtracker-Verhinderung und HTTPS (zwingendes HTTPS für alle Seiten, sofern vorhanden) (Lund, 2021, S. 15). Hinter Brave steckt keine Dark-Web-Technologie, sondern eine Privatsphäre-optimierte Variante eines regulären Browsers – Daten werden

nicht verschlüsselt oder über mehrere Relays geroutet. Trotzdem bietet es gemäss Lund (2021) einen erhöhten Schutz der Benutzendendaten und eignet sich demnach zur Integration in einer Bibliothek.

Wie Coffey beschreibt, eignet sich das Dark Web im weitesten Sinne als Applikation für die Wahrung der Privatsphäre von Bibliotheksbenutzenden (2020, S. 1). Dark-Web-Technologien sind jedoch nicht geeignet für die akademische Forschung wie das Deep Web selbst, auch wenn es per Definition ein Teil davon ist; der kurzlebige Charakter der meisten Inhalte des Dark Webs disqualifiziert sie als zuverlässige Informationsquelle für die wissenschaftliche Forschung (Coffey, 2020, S. 3).

2.2 Akademische Forschung und Schattenbibliotheken

Das Deep Web mit Ressourcen in Fachdatenbanken, Verlagsrepositorien und Bibliothekskatalogen ist eine bedeutungsvolle Quelle für die Recherche und Forschung im akademischen Sinne. Jedoch ist der Zugang zu jener wissenschaftlichen Literatur erschwert, einerseits durch Paywalls, andererseits durch die Cognitive Invisibility (vgl. Kap. 2.1.2.2; 2.3) (Himmelstein et al., 2018, S. 1). Die im Jahr 2018 durchgeführte Studie von Piwowar et al. zeigt eine Open-Access-Verfügbarkeit von Artikeln von 27.9%, bei neueren Artikeln (jünger als 2 Jahre) sind es 47% (S. 10). Die Fachgebiete Biomedizin und Mathematik verzeichnen OA-Verfügbarkeiten von 50%, während es in der Chemie und im Ingenieurwesen weniger als 20% sind (S. 13). Trotz wachsender OA-Bewegung bleibt der Zugang zu wissenschaftlicher Literatur limitiert und Lizenzen auf Institutionen wie Universitäten oder medizinische Einrichtungen beschränkt (Himmelstein et al., 2018). Kleinere Institutionen oder Entwicklungsstaaten sind mit zusätzlich erschwertem Zugang konfrontiert (Bendezú-Quispe, Nieto-Gutiérrez, Pacheco-Mendoza & Taype-Rondan, 2016).

Himmelstein et al. (2018, S. 1) führt die Problematik hierbei folgendermassen auf: Zusätzlich dazu, dass nur ein kleiner Teil der Bevölkerung Zugang zu wissenschaftlicher Literatur hat, sind Veröffentlichungen in Journals das wichtigste Ergebnis der Wissenschaft. Das Produkt der Forschung und das Wissen wird durch das Vermarkten des Urheberrechts geschützt. An diesem Punkt scheitern die Lizenznehmenden wie Bibliotheken und Universitäten, die aufgrund der Einzigartigkeit eines jeden Artikels keine Substitute ohne Wissensverlust ihren Nutzenden gegenüber liefern können. Während infolgedessen die Preise für Zeitschriftenabonnements stetig gestiegen sind, müssen Bibliotheken budgetbedingt andere Dienstleistungen kürzen.

Weiter angeregt wurde die Diskussion um diese Problematik durch das Aufkommen von Schattenbibliotheken, die Zugang zu Raubkopien von Artikeln und Zeitschriften bieten und trotz Open Access weiter Zuwachs an Nutzenden erhält. Die Existenz und rege Nutzung dieser Webseiten hinterfragen die Ethik und Wirksamkeit des klassischen Publikationsmodells weiter (vgl. Kap. 2.2.2) (Piwowar et al., 2018, 2). Neben dieser Art des «Black Open Access» wird auch das Peer-to-Peer File-Sharing betrieben, das in die Dark-Web-Technologien als Mittel des Transfers hineingreift (Greshake, 2017, S. 3). Letzteres wird in der vorliegenden Arbeit jedoch nicht genauer erläutert.

2.2.1 Sci-Hub und Library Genesis

Schattenbibliotheken lassen sich definieren als Dokumentensammlungen in einer Datenbank im Surface Web, die die Paywalls von Zeitschriften und Verlagen umgehen und Raubkopien frei zugänglich zur Verfügung stellen. Sie verletzen dabei das Urheberrecht und gelten daher als Piraterie (Piwowar et al., 2018, 2). Online-Sammlungen von raubkopierten E-Books und E-Journals kamen bereits in den frühen 2000er-Jahren auf; die meisten davon persönliche Scans, die via Social Media geteilt wurden. Vereinzelt bildeten sich grössere Kollektionen, die zu Archiven heranwuchsen: Darunter Hansi library, Aaaaarg und Library Genesis (LibGen) (Karaganis, 2018, S. 1). Später stiess Sci-Hub dazu.

LibGen wurde 2008 durch russische Wissenschaftler gegründet und unterscheidet sich von klassischen Schattenbibliotheken. Anstatt den Inhaltszuwachs zu kontrollieren, sammelt LibGen frei verfügbare wissenschaftliche Texte aus dem Internet und führt sie in einer einzigen, offenen Datenbank zusammen. Das Suchen und Abrufen von Texten ist zwar eine nutzbare Funktion, dennoch liegt der Schwerpunkt auf der Verbreitung der eigenen Bibliotheksinfrastruktur – einschließlich des Quellcodes, des Katalogs und der Terabyte-Sammlung – an jede Person, die ihre eigene Bibliothek aufbauen möchte. Kopien der gesamten Datenbank und Instruktionen für das «Mirroring» können heruntergeladen und damit begonnen werden, Texte unter den eigenen Bedingungen zu verbreiten. Diese Offenheit hat zur Entstehung eines lebendigen Ökosystems von Schattenbibliotheken rund um die LibGen-Kernsammlung geführt. Ausserdem kann LibGen seine begrenzten Ressourcen auf die Pflege einer qualitativ hochwertigen wissenschaftlichen Sammlung konzentrieren, während die Mirror-Sites um den besten Service für die Nutzenden konkurrieren, die Kosten für die Verbreitung tragen, die LibGen-Sammlung der Öffentlichkeit zur Verfügung stellen und gleichzeitig die Wahrscheinlichkeit ihres langfristigen Überlebens erhöhen (Bodó, 2018a, S. 27–28).

Im Jahr 2011 veröffentlichte Alexandra Elbakyan (Informatik-Doktorandin aus Kasachstan) Sci-Hub, welches das File-Sharing von wissenschaftlichen Artikeln durch die Möglichkeit der Suche und des direkten Downloads vereinfachte und beschleunigte (Karaganis, 2018, S. 1). Elbakyan entwickelte Sci-Hub aufgrund der Zugangsprobleme; sie benötigte tausende von Artikeln für ihre Forschungen, fand jedoch keinen legalen Weg nicht für jeden Artikel durchschnittlich 30 Dollar bezahlen zu müssen (Hoy, 2017, S. 74).

Die ursprüngliche Tagline der Sci-Hub-Webseite war «to remove all barriers in the way of science» (Hoy, 2017, S. 73). Mittlerweile ändert sich die Tagline mit jedem Refresh der Seite, so liest sich beispielsweise: «Knowledge as a human right», «knowledge must be free», «breaking academic paywalls since 2011» oder «common ownership of the means of production, free access to articles of consumption» (Sci-Hub, 2023b). In der Beschreibung positioniert sich Sci-Hub als das umstrittenste Projekt der modernen Wissenschaft mit dem Ziel, freien und uneingeschränkten Zugang zu allen wissenschaftlichen Erkenntnissen zu bieten (Sci-Hub, 2023a). Per Juli 2023 verzeichnete Sci-Hub 88'343'822 Dokumente in dessen Datenbank (vgl. Kap. 4.2) (Sci-Hub, 2023b).

Die Suche in Sci-Hub gestaltet sich nicht als Discovery-Tool, sondern verlangt nach eindeutigen Identifikatoren wie DOI, PubMed ID, URL oder vollständigen Titeln. Bis im Jahr 2014 leitete Sci-Hub Suchanfragen als erster Schritt an LibGen weiter und schickte den gegebenenfalls vorhandenen Artikel an den Nutzenden (Elbakyan, 2017; Hoy, 2017, S. 74). Seither leitet Sci-Hub nur noch im Falle einer Suchanfrage eines Buches an LibGen weiter, andernfalls bedient es sich der eigenen Datenbank. Befindet sich noch keine Kopie des Artikels im Repository, durchsucht Sci-Hub die Zeitschriftendatenbanken. Möglich wird dies durch freiwillig zur Verfügung gestellte oder gestohlene institutionelle Zugangsdaten (González-Solar & Fernández-Marcial, 2019, S. 2; Hoy, 2017, S. 74), wobei Elbakyan Letzteres dementiert: «I cannot disclose the source of the credentials. I assume that some credentials coming to Sci-Hub could have been obtained by phishing. Anyway, Sci-Hub is not doing any phishing by itself. The credentials are used only to download papers.» (2017).

Sci-Hub positioniert sich selbst als legal: «The project is legal, while restricting access to information and knowledge is not» (Sci-Hub, 2023a). Am 03.06.2015 reichte Elsevier jedoch Klage gegen Sci-Hub und LibGen auf Grundlage von Urheberrechtsverletzung ein und gewann am 21.07.2017 den Rechtsstreit: Elsevier wurden 15'000'000 Dollar Schadenersatz gewährt, während Sci-Hub ein Verbot zur Kopierung jeglicher Artikel von Elsevier erhielt (District Court, S.D. New York 1:15-cv-04282). 2015 wurde die Webseite «sci-hub.org» ausserdem aufgrund der Klage abgeschaltet, woraufhin sie 18 Tage später

ausserhalb der USA gehostet wieder online ging (Hoy, 2017, S. 75). Seither änderten der Domain-Name und der Standort des Hostings mehrfach, darunter Laos, Hong Kong, Mongolei, Taiwan und Russland, und es wird ein Onion Service der Seite geführt (González-Solar & Fernández-Marcial, 2019, S. 2). Gemäss Lawson (2017, S. 27) ist Sci-Hub aufgrund der Natur des Webs, der diversen Absicherungen seitens Elbakyan und der Hacker-Gemeinschaft nicht zu stoppen.

Nebst den urheberrechtlichen Diskussionen und den Auswirkungen auf die Verlage werden ethische Fragestellungen in Zusammenhang mit dem Zugangsproblem in den Fokus gestellt. So berichten Bendezú-Quispe et al. (2016) über das ethische Dilemma der Nutzung von Sci-Hub im medizinischen Sektor in Peru: Ärzt*innen benötigen die aktuellste medizinische Literatur, um komplexe Krankheiten zu behandeln. In vielen Fällen haben sie jedoch keinen Zugang dazu, da keine institutionelle Lizenz besteht, keine Finanzierungen für Abonnements vorhanden sind und Konferenzen (nebst Interessenskonflikten) diese Themenfelder nicht abdecken. Das ethische Dilemma besteht somit in der Notwendigkeit einer illegalen Plattform, um Patienten mit einer effektiven und effizienten Behandlung versorgen zu können. Diese Art von Dilemma steht beispielhaft für eine Reihe von Ländern mit vergleichbaren Problemen. Hoy kommentiert dazu: «If Sci-Hub can improve medical care or save lives, is there a moral obligation to make that information available? Do doctors have a moral obligation to use that information, even if it is obtained via copyright infringement?» (2017, S. 75).

Ein weiterer Grund für die Nutzung von Sci-Hub, insbesondere in den westlichen Ländern (González-Solar & Fernández-Marcial, 2019, S. 2) und frei von den ethischen Fragestellungen, ist die Bequemlichkeit und Schnelligkeit der Artikelsuche. Aus Sicht des Recherchierenden liefert ein Download von Sci-Hub oder ein Download über die Verlagswebseite mit Institutionszugang dasselbe Resultat, zumal der heruntergeladene Artikel eine exakte Kopie des Originals ist (Lawson, 2017, S. 26–27).

2.2.2 Auswirkungen auf traditionelles Publikationsmodell und Bibliotheken

Die Meinungen rund um Schattenbibliotheken und ihre Auswirkungen gehen stark auseinander: Destabilisierung der wissenschaftlichen Verlagsbranche, verminderte Einnahmen der Verlage, veränderte Nutzung wissenschaftlicher Bibliotheken. Trotzdem wird besonders Sci-Hub trotz dessen Illegalität (in Bezug auf Copyright) mehr gelobt als verurteilt. Dies birgt die Frage in sich, wie Bibliotheken damit umgehen sollen (Gardner, McLaughlin & Asher, 2017, S. 569). Auf der einen Seite steht die Dienstleistung gegenüber den Kund*innen, auf der anderen Seite die Verlage und deren Lizenzverträge.

In der Klage gegen Sci-Hub liess Elsevier einen jährlichen Verlust von 5000 Dollar verlauten. Darin miteinbezogen waren jedoch nur 100 dokumentierte Artikel von Elsevier, die in Sci-Hubs Datenbank vorhanden waren (District Court, S.D. New York 1:15-cv-04282). Laut der Studie von Greshake (2017, S. 7) sind allerdings rund 24% der gesamten Downloads auf Sci-Hub Artikel von Elsevier. Mit den damaligen Zahlen waren dies knapp 8'000'000 Downloads. Der effektive Verlust lässt sich somit um ein Vielfaches vermuten. Banks (2016) betont ausserdem, dass nicht nur Verlage wie Elsevier betroffen sind, die unter Umständen wenig Sympathien aufbringen – kleine Verlage könnten aus dem Geschäft gedrängt werden.

Lawson (2017, S. 28) argumentiert, dass die traditionelle Publikationstätigkeit der Verlage noch immer von Bedeutung und die Piraterie von Artikeln davon abhängig ist. Würden alle Abonnenten ihre Lizenzverträge zurückziehen (bzw. sich nur auf Sci-Hub verlassen) und die Verlage keine Einnahmen mehr erhalten, stünde die Veröffentlichung jeglicher akademischen Produkte still. Sci-Hub funktioniert ohne traditionelles Publikationsmodell nicht und wäre in diesem Szenario eine Datenbank mit fortlaufend alternden Ressourcen (Hoy, 2017, S. 75). Wenn die Verlagsarbeit fortgesetzt werden muss, das Abonnementmodell jedoch nicht zweckmässig ist, sind alternative Finanzierungsformen notwendig (Lawson, 2017, S. 28).

González-Solar und Fernández-Marcial sagen dazu, dass Sci-Hub ein Symptom und eine Reaktion auf ein Problem sei, das in jedem Fall gelöst werden muss (2019, S. 5). Es kennzeichnet einen Wendepunkt und eine Herausforderung für wissenschaftliche Bibliotheken betreffend der häufig als komplizierter angesehenen angebotenen Retrieval-Systeme. Gewisse Verlage weisen gar die Schuld den Bibliotheken zu für nicht benutzerfreundliche Systeme und mangelnde Schulungen darin (S. 6).

Bibliothekar*innen können ihre Benutzenden nicht davon abhalten, Schattenbibliotheken zu nutzen und Verbote auszusprechen ist wenig zielführend. Filter auf der Institutions-Firewall sind ebenso ineffektiv, da von jeder Schattenbibliothek mehrere Mirror-Sites existieren und Domainnamen ändern können (oder ein Zugriff via Tor möglich ist). Hoy empfiehlt jedoch, das differenzierte Wissen darüber weiterzugeben und zu informieren, inwiefern und weshalb andere Quellen die bessere Wahl sein können: «As always, the best course of action for librarians is to educate their patrons when they see infringement happening.» (2017, S. 77). González-Solar und Fernández-Marcial (2019, S. 9) sprechen keine direkte Empfehlung aus, warnen jedoch vor dem sich ändernden Informationsverhalten von Forschenden und der Notwendigkeit von Bibliotheken, sich daran anzupassen.

2.3 Forschungsunterstützung durch Bibliotheken

Die Services von Bibliotheken im digitalen Bereich haben sich in den letzten Jahren stark verbessert, und trotzdem können sie mit der Schnellebigkeit und Einfachheit von Schattenbibliotheken nicht mithalten. Bibliotheken bleiben aber Anlauf- und Beratungsstelle für die Beschaffung von Informationsressourcen, somit müssen sie auf Anfragen zu jeglichen Themen rund um das Web vorbereitet sein – sei es für Suchanfragen in Google, in Datenbanken im Deep Web, Privatsphäre-Themen oder Schattenbibliotheken (die wiederum mit dem Deep Web verbunden sind) (González-Solar & Fernández-Marcial, 2019, S. 2).

In den bisherigen Kapiteln wurde vermehrt die Wichtigkeit von Bibliothekar*innen und Informationsspezialisten im Prozess der Recherche und Informationsfindung aufgezeigt. Es braucht geschultes Personal, das die Anfragen der Benutzenden abfangen und durch individuell angemessene Unterstützung weiterverarbeiten kann. So können die Cognitive Invisibility, also das persönliche Deep Web, systematisch durchbrochen und verborgene wertvolle Informationen gefunden werden. Eine mögliche Strategie hierfür ist die Entwicklung von Research Support Services, die an diversen Universitäten bereits Verwendung finden (Fernández-Marcial, Costa & González-Solar, S. 2). Gleichzeitig kann das dem sich verändernden Informationsverhalten von Forschenden Rechnung tragen (vgl. Kap. 2.2.2). Eine weitere Möglichkeit sind Kollaborationen mit Institutionen, die bereits von Nutzenden in Gebrauch sind. So zum Beispiel das Projekt von Google Books, das die erhöhte Sichtbarkeit des Webs anstrebt und im Zuge dessen mit Universitätsbibliotheken zusammenarbeitet. Google ist sich der begrenzten Erreichbarkeit des gesamten Webs durch deren Suchmaschine bewusst und ist durch das Projekt im «deep field indexing» involviert (Hatta, 2020, S. 279). Auch Google Scholar ermöglicht teilweise die Suche im Deep Web, jedoch kann es nur das Suchergebnis liefern und nicht den Zugang dazu – sofern nicht ein Lizenzvertrag mit dem betreffenden Verlag existiert und die Institution die Verlinkung zu ihren eigenen Datenbanken erlaubt. Aus der Sicht einer breitflächig ausgeführten Recherche besteht die Gefahr bei Google Scholar, dass Forschende und Studierende sich ausschliesslich darauf verlassen und weiterführende Ressourcen aus Fachdatenbanken aussen vor lassen. Trotzdem stellt es einen benutzerfreundlichen Übergang zu Deep-Web-Inhalten dar (Devine & Egger-Sider, 2014, S. 38).

Inwiefern Studierende und Forschende die Unterstützung von Bibliotheken in Relation zum Deep Web gebrauchen oder wünschen, ist in den Ergebnissen zur Evaluation genauer beschrieben (vgl. Kap. 4.1).

3 Methodik

Mit der vorliegenden Arbeit wurde die Diskrepanz im Wissen und im Kenntnisstand zwischen Bibliothekar*innen und den ausgewählten Zielgruppen im Bereich des Deep Webs (inklusive darin enthaltenem Dark Web) ergründet. Die Forschungsfrage hierfür lautete:

«Wie gross ist die Diskrepanz im Kenntnisstand zwischen Bibliothekar*innen und Studierenden/Forschenden bezüglich des Deep Webs, und inwiefern beeinflusst dies die Aktualität und Relevanz von Bibliotheken?»

Die Methodik bestand aus zwei Herangehensweisen. Einerseits die Zusammentragung von Recherche und Literatur, um den Forschungsstand der Internet- und Weblandschaft der behandelten Themen aufzuarbeiten. Andererseits wurden drei quantitative Umfragen durchgeführt, die den Kenntnisstand des Dark- und Deep Webs, Sci-Hub und Library Genesis und die Erfahrungen und Bedürfnisse mit Unterstützungsarbeit der jeweils anderen Zielgruppen erfragten. Die Befragungen wurden bei den folgenden Gruppen verbreitet:

- Bibliothekar*innen, schweizweit über den Verteiler «Swisslib» und unabhängig der Art der Bibliothek (keine Einschränkung auf wissenschaftliche Bibliotheken)
- Studierende der FHGR (keine Einschränkung der Studiengänge)
- Akademischer Mittelbau an der Universität Bern (Graduate School for Health Sciences (GHS) und Graduate School for Cellular and Biomedical Sciences (GCB))

Die Auswahl erfolgte durch die Verfügbarkeit der Gruppen seitens der Verfasserin und um Ergebnisse im akademischen Bereich von zwei unterschiedlichen Standpunkten zu erhalten: Forschenden und Doktorierenden mit geschätzt hohem Bedarf an wissenschaftlicher Literatur in ihren Bereichen und Studierenden an einer Hochschule mit hohem Praxisbezug. Die Umfrage der Bibliothekar*innen wurde durch den E-Mail-Verteiler für Informations- und Dokumentationsspezialisten geteilt, um möglichst breite und schweizweite Resultate zu erzielen. Es erfolgte dabei bewusst keine Einschränkung auf akademische Bibliotheken, um den Stand des Deep Webs in Bibliotheken im Allgemeinen evaluieren zu können.

In einem weiteren Schritt wurden effektive Zugriffszahlen auf Sci-Hub und Library Genesis von der IP-Adresse der Universität Bern aus gemessen. Dafür erfolgte in Zusammenarbeit mit der Informatikabteilung eine Regeleinrichtung auf der Firewall für den Zeitraum eines Monats – vom 12.06.2023 bis 12.07.2023 –, die die Zugriffe auf die URLs überwachte. Die Ergebnisse wurden anschliessend mit den Antworten des akademischen Mittelbaus in Verbindung gebracht. Im Falle von Sci-Hub, welches das Abrufen von

weltweiten aktuellen Statistiken ermöglicht, wurden die Zahlen in Relation zueinander gebracht.

Die Fragestellungen ergänzend zur Forschungsfrage lauteten:

- Was ist der prozentuale Anteil der Nutzung von Sci-hub, Library Genesis, Deep Web und Dark Web der drei Gruppen?
- Was ist der aktuelle Forschungsstand bezüglich des Deep Webs, Dark Webs, Sci-Hub und Library Genesis und wie stehen sie in Verbindung zu den Auswirkungen auf Bibliotheken?
- Was führt zur Nutzung von Sci-Hub und Library Genesis anstelle der Dienstleistungen der Bibliothek?
- Wie viele Zugriffe auf Sci-Hub und Library Genesis wurden im letzten Monat vom Netz der Universität Bern aus getätigt?
- Wie können Bibliotheken das Deep Web und Dark Web zu ihrem Vorteil nutzen?
- Welche Aussagen lassen sich durch die Unterschiede des untersuchten Kenntnisstands auf die Aktualität und Relevanz von Bibliotheken machen?

Die Umfragen wurden mit dem Tool «Lime Survey» erstellt und enthielten je 16 respektive 17 Fragen. Die Fragebögen für die Studierenden und den akademischen Mittelbau waren identisch, wobei jener für die Graduate Schools aufgrund der hohen Anzahl an internationalen Doktorierenden in englischer Sprache verfasst wurde. Die Fragen für die Bibliothekar*innen folgten einem ähnlichen Aufbau und dienten je nach Fragegruppe als Gegenstück zu den Fragen der anderen beiden Zielgruppen (s. Anhang 11_Fragebögen).

4 Ergebnisse

4.1 Umfragen

Zur besseren Unterscheidung werden die Fragebögen für die Auswertung entsprechend benannt:

Fragebogen **A**: Studierende FHGR

Fragebogen **B**: Akademischer Mittelbau UNIBE

Fragebogen **C**: Bibliothekar*innen

Fragebogen A und B enthielten die folgenden Fragegruppen (exkl. demografischer Angaben):

AB1	Internetnutzung	Zwei Fragen zur Nutzung des Surface Webs in Zusammenhang mit der Forschung/dem Studium
AB2	Deep Web	Vier Fragen zur Kenntnis und Nutzung des Dark- und Deep Webs im Studienalltag
AB3	Sci-Hub und Library Genesis	Vier Fragen zur Kenntnis, Nutzung und Grund und Häufigkeit derselben
AB4	Bibliotheksdienstleistungen	Drei Fragen zur Inanspruchnahme von Bibliotheksdienstleistungen in Zusammenhang mit dem Deep Web

Tabelle 2. Fragegruppen der Fragebögen A und B (eigene Darstellung)

Fragebogen C enthielt ähnliche und korrespondierende Fragegruppen (exkl. demografischer Angaben und beruflicher Werdegang):

C1	Internetnutzung	Zwei Fragen zur Nutzung des Surface Webs in Zusammenhang mit der beruflichen Tätigkeit
C2	Deep Web	Fünf Fragen zur Kenntnis und Nutzung des Dark- und Deep Webs im Berufsalltag
C3	Sci-Hub und Library Genesis	Fünf Fragen zur Kenntnis und Nutzung von Sci-Hub und Library Genesis und der Konfrontation derselben im Berufsalltag

Tabelle 3. Fragegruppen des Fragebogen C (eigene Darstellung)

Insgesamt wurden die Fragebögen zu Anteilen von mindestens 71.8% vollständig ausgefüllt:

A alle: 63

A vollständig: **55**

B alle: 90

B vollständig: **80**

C alle: 142

C vollständig: **102**

Die unvollständigen Antworten wurden in der Auswertung nicht berücksichtigt. Die Bögen A und B werden jeweils direkt verglichen, während C separat und anschliessend in Kapitel 5 A und B gegenübergestellt wird.

4.1.1 Auswertung A und B

Die Teilnehmenden Studierenden der FHGR repräsentierten 4 verschiedene Studienrichtungen: Informationswissenschaften (83%), Digital Business Management (11%), Betriebsökonomie (2%) und MSc Data Visualization (4%).

Der Akademische Mittelbau an der UNIBE vertrat die Forschungsfelder: Biomedical Engineering (14%), Biomedical Sciences (8%), Biology (8%), Cell biology (6%), Virology (4%), Biochemistry (4%), Microbiology (3%), Immunology (8%), Computer Science (1%), Veterinary Medicine (3%), Veterinary Virology (1%), Neuroscience (8%), Endocrinology (1%), Bioinformatics (1%), Medicine (15%).

4.1.1.1 Fragegruppe AB1

In der Nutzung von Suchmaschinen in Zusammenhang mit der Studien- oder Forschungsrichtung zeigen sich bereits signifikante Unterschiede. Während Gruppe A einen täglichen Gebrauch von 41.82% angab, waren es bei Gruppe B 81.25%. Dies ist jedoch wenig überraschend, da Doktorierende und Forschende einer Graduate School üblicherweise fortlaufend an Projekten oder Studien arbeiten.

In den Antworten zu was Suchmaschinen am meisten genutzt werden, gab Gruppe A überwiegend das Nachschlagen von Begriffen und die thematische Suche an. Vereinzelt wurden Recherchen für die Arbeit genannt. Bei Gruppe B fiel dagegen die Artikelsuche und die Gewinnung eines Überblicks über ein Thema am häufigsten.

4.1.1.2 Fragegruppe AB2

Frage 1 bezog sich darauf, ob der Begriff des Deep Webs schon einmal gehört wurde. Während Gruppe A zu 100% Ja antwortete, sind es bei Gruppe B rund die Hälfte (52.5%). Die Häufigkeit der Nutzung des Deep Webs verteilte sich bei beiden Gruppen (s. Abb. 5; 6).

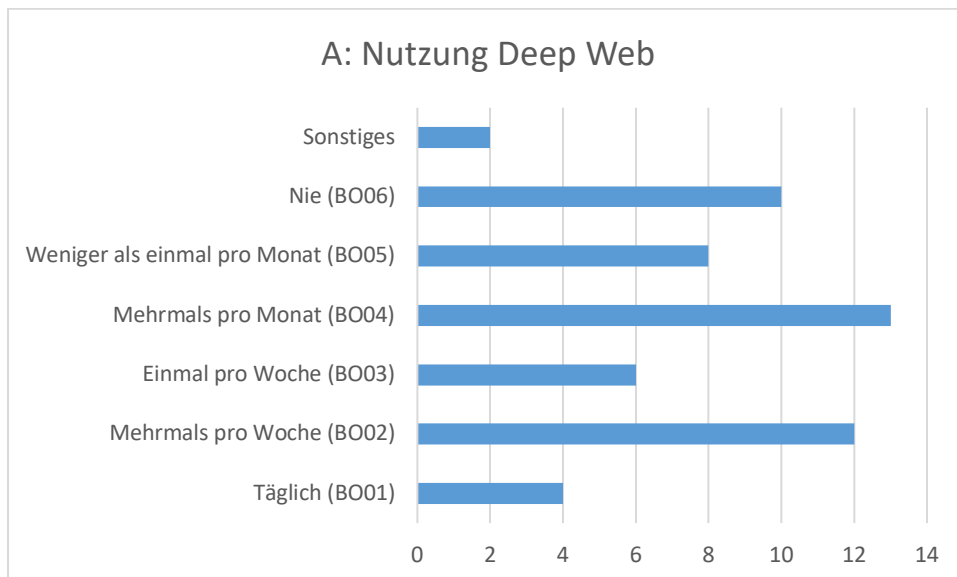


Abbildung 5. Häufigkeit Nutzung Deep Web, Gruppe A (eigene Darstellung)

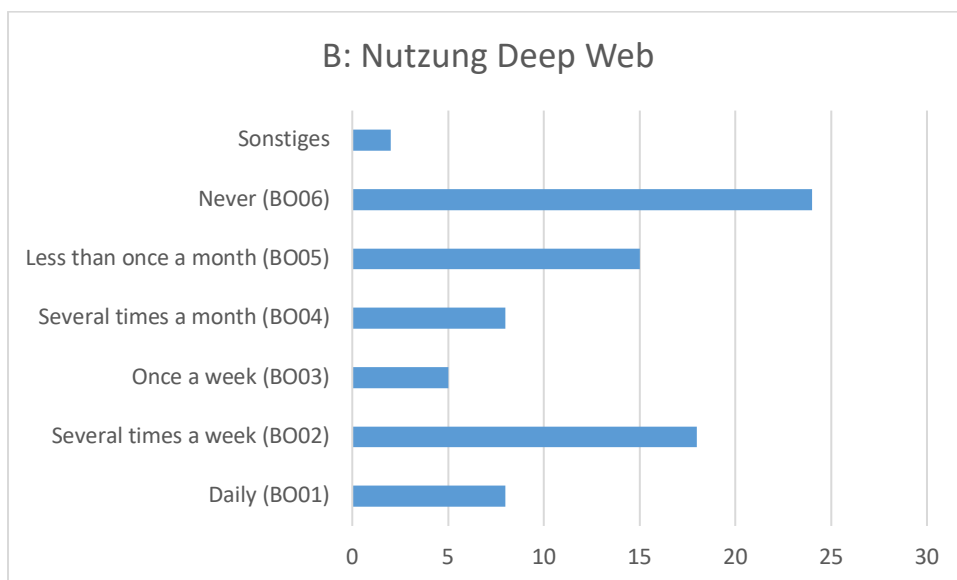


Abbildung 6. Häufigkeit Nutzung Deep Web, Gruppe B (eigene Darstellung)

Auffallend ist der hohe Wert von 24 bei Antwort BO06 im akademischen Mittelbau (30%). Hier stellt sich die Frage, ob das Deep Web tatsächlich nie in Gebrauch ist, oder ob die Begriffsdefinition in der Fragestellung zu wenig präzise formuliert war. Mit Vergleich zur Fragegruppe AB1 mit dem täglichen Gebrauch der Suchmaschinen und der

überwiegenden Artikelsuche darin liegt die Vermutung nahe, dass ein Klick in Google Scholar mit Weiterleitung zu einer Verlagsseite nicht als Deep Web interpretiert wird. Eine weitere Möglichkeit ist, dass jene 30% nur Google und Google Scholar benutzen und den direkten Weg zu einer Fachdatenbank nicht vornehmen – demnach könnte aus der Definitionsbeschreibung abgeleitet worden sein, dass das Deep Web nie in Gebrauch ist.

96.36% von Gruppe A haben den Begriff des Dark Webs bereits gehört, 87.5% sind es bei Gruppe B. Die Unterschiede zeigen sich erst beim Zugriff darauf (s. Abb. 7; 8).

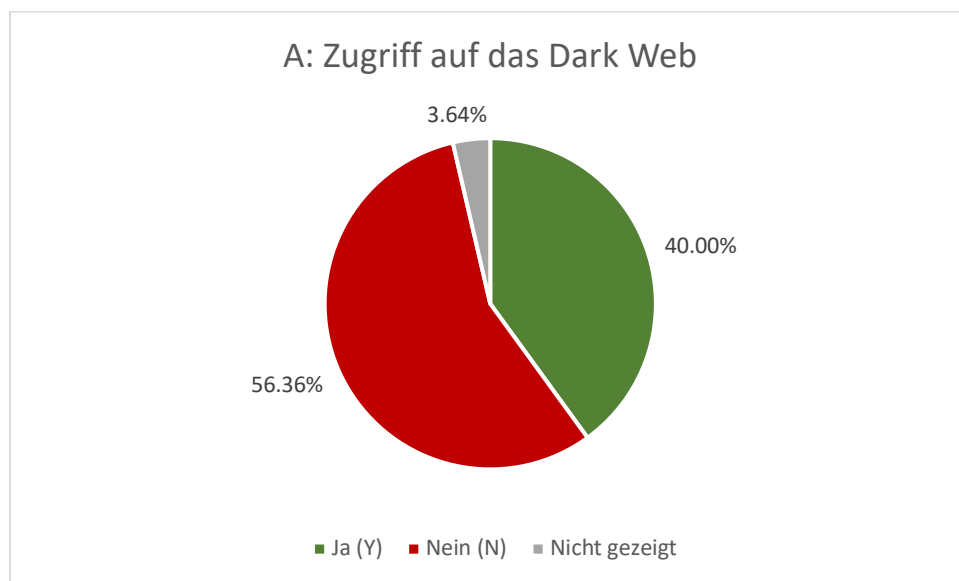


Abbildung 7. Zugriff auf das Dark Web, Gruppe A (eigene Darstellung)

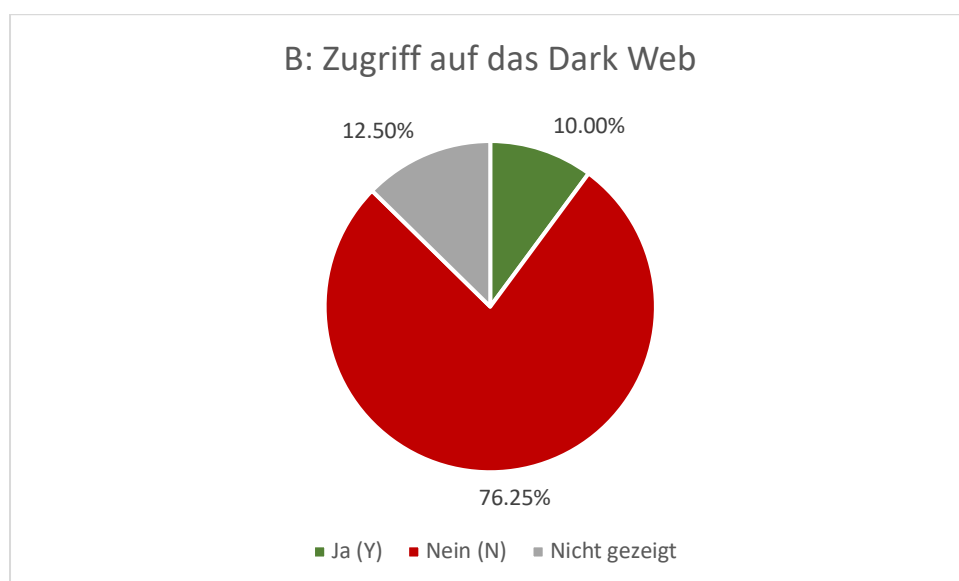


Abbildung 8. Zugriff auf das Dark Web, Gruppe B (eigene Darstellung)

«Nicht gezeigt» bezieht sich darauf, dass die Frage bei Unbekanntheit des Begriffs «Dark Web» nicht angezeigt wurde. Somit sind 40% (22) der 55 Befragten aus Gruppe A schon einmal mit dem Dark Web in Berührung gekommen, während es bei Gruppe B nur 10% (8) der 80 sind. Es gibt hierbei keinen Anhaltspunkt, weshalb das Dark Web bei Gruppe A verbreiteter ist – es wurde keine Kommentarspalte im Fragebogen zur Verfügung gestellt.

Gruppe A kann in der Repräsentativität der FHGR-Studierenden der Fragegruppe AB2 in Frage gestellt werden. Der hohe Anteil an Informationswissenschafts-Studierenden führt womöglich zu einer höheren Kenntnis des Deep Webs, zumal die Internetarchitektur in mehreren Modulen Teil des Lehrplans ist (Fachhochschule Graubünden, 2020).

4.1.1.3 Fragegruppe AB3

In der dritten Fragegruppe wurden die Kenntnis und Nutzung sowie der Grund für den Gebrauch von Sci-Hub und/oder LibGen gegenüber den Lizenzen der Bibliothek abgefragt.

Beide Gruppen haben eine hohe Rate an Befragten, die bereits von den zwei Schattenbibliotheken gehört haben: Bei A sind es 72.37% (40), bei B 88.75% (71). Dies kann wiederum auf die hohe Studien- und Forschungsarbeit der Graduate Schools zurückgeführt werden, was die Differenz in den Nutzungszahlen bestätigt: Während nur die Hälfte der 40 Personen aus Gruppe A mindestens eine der Schattenbibliotheken benutzt hat, sind es 60 von 71 beim akademischen Mittelbau (75% von N=80).

Die Mehrheit greift mehrmals oder weniger als einmal pro Monat auf Sci-Hub und/oder Library Genesis zu (s. Abb. 9; 10).

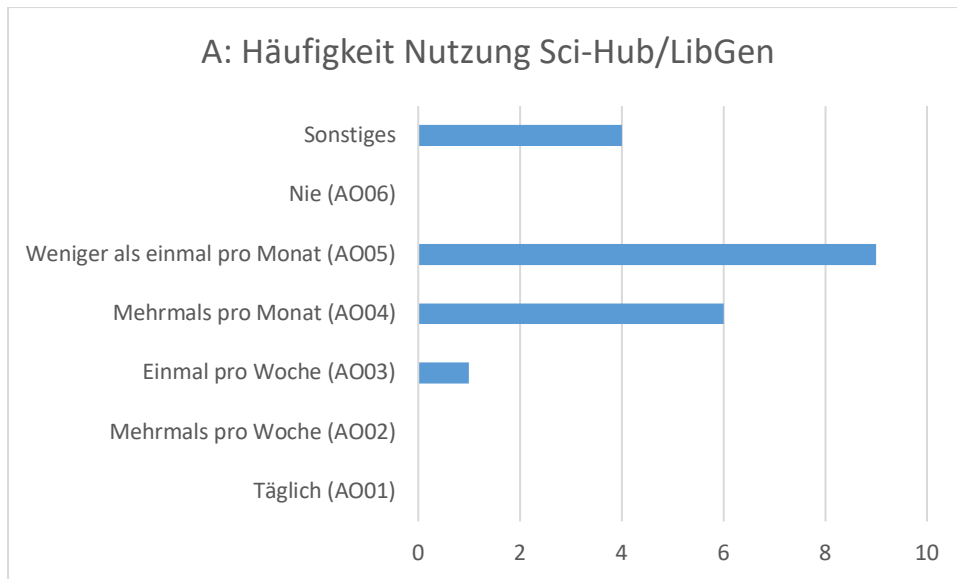


Abbildung 9. Häufigkeit der Nutzung von Sci-Hub/LibGen, Gruppe A (eigene Darstellung)

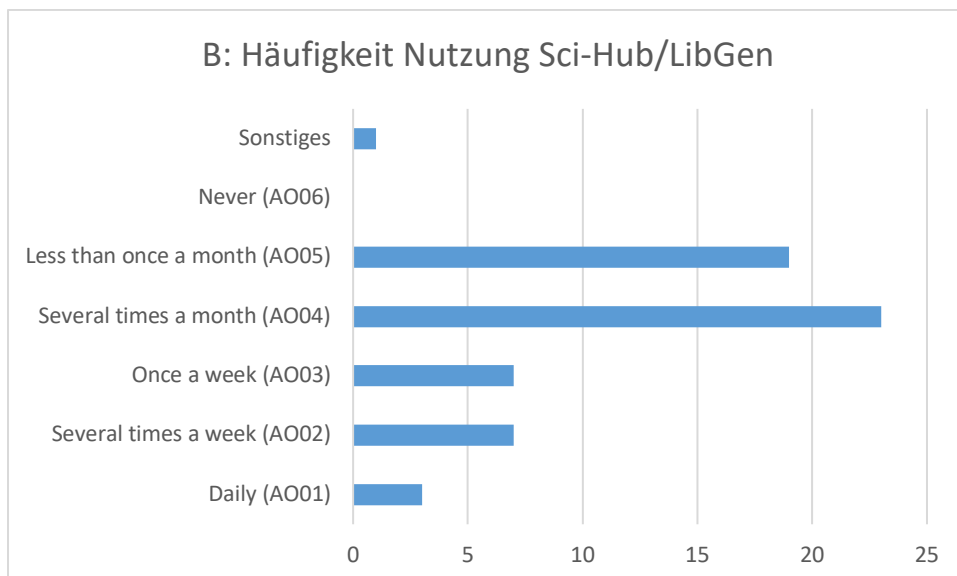


Abbildung 10. Häufigkeit der Nutzung von Sci-Hub/LibGen, Gruppe B (eigene Darstellung)

Kommentare der Gruppe A zu Abbildung 9:

- «Immer häufiger seit dem Modul IRES» (3x)

Kommentar der Gruppe B zu Abbildung 10:

- «When I was not a registered student I used it it frequently but now that I have access to almost all journals I don't use it as often as before.»

Der Gebrauch der Schattenbibliotheken primär bei mangelndem Zugang der Universität wurde auch bei den Begründungen mit 63.38% als häufigster Grund gewählt (s. Abb. 12). Bei Gruppe A dagegen waren es Zugriffsprobleme auf bereits lizenzierte Ressourcen (38.24%) und dass viele benötigte Artikel und E-Journals generell nicht lizenziert sind (34.29%) (s. Abb. 11).

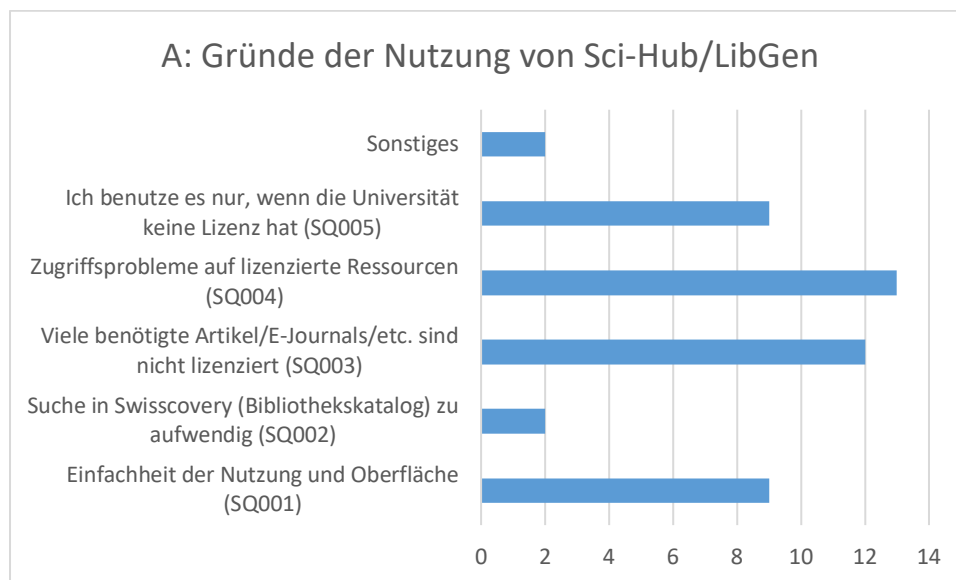


Abbildung 11. Gründe der Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Fachhochschule, Gruppe A (eigene Darstellung)

Abgegebene Kommentare bei «Sonstiges» zu Abbildung 11:

- «Auf Libgen finde ich auch Belletristik und Comics die ich als e-book lesen und so Geld sparen kann. Wenn mir etwas sehr gut gefällt, dann kaufe ich das Buch/Comic auch für meine Sammlung.»
- «Download von Fachliteratur, wenn nicht von Dozierenden zur Verfügung gestellt wird»

Der erste Kommentar deutet auf eine private Nutzung hin. Inwiefern dies ein Einzelfall ist oder mehrere Personen ausserhalb des akademischen Kontexts geantwortet haben, ist nicht bekannt.

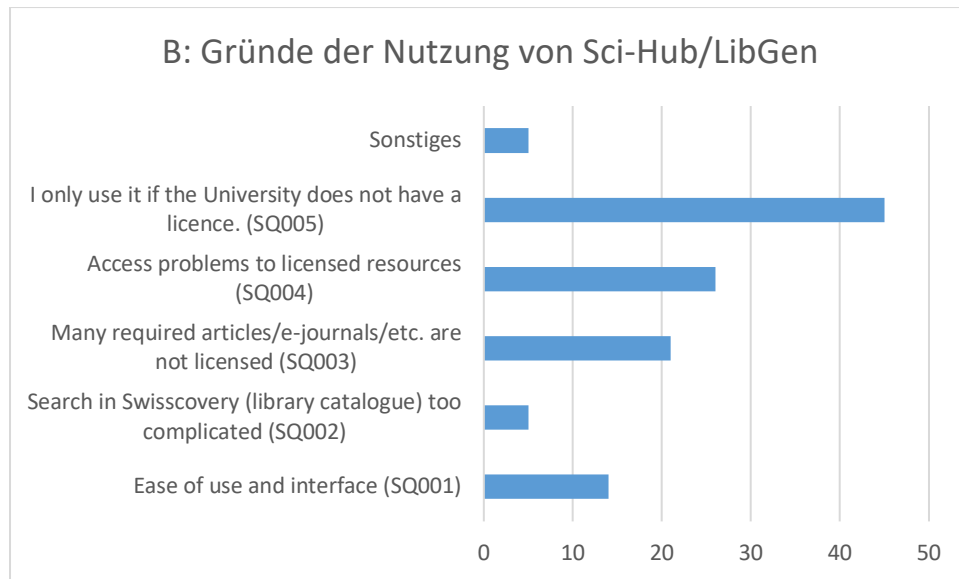


Abbildung 12. Gründe der Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Universität, Gruppe B (eigene Darstellung)

Auswahl abgegebene Kommentare bei «Sonstiges» zu Abbildung 12:

- «General support of the concept, Science should be free for everyone, not only for the ones who can afford it»
- «I use it when I want to take a look at the table of content of some books»
- «don't have to login»

Der erste Kommentar referenziert moralische Beweggründe, die in der Fragestellung nicht miteinbezogen waren. Das Anschauen eines Inhaltsverzeichnisses via Sci-Hub kann einerseits auf die Einfachheit des UI zurückzuführen sein, oder aber auf die Unwissenheit der Verfügbarkeit von gescannten Inhaltsverzeichnissen im Bibliothekskatalog. Generell gaben beide Gruppen nur zu wenigen Anteilen Swisscovery als zu kompliziert an. Bei Gruppe A ist die Wahrscheinlichkeit hoch, dass die Befragten Swisscovery gut kennen. Bei Gruppe B, mit Bezug zur Fragegruppe AB1 und der hohen täglichen Nutzung von Google und Google Scholar für die Artikelsuche, kann das Gegenteil der Fall sein; Swisscovery wird signifikant weniger genutzt als Suchmaschinen und dementsprechend nicht als ausschlaggebenden Grund für die Nutzung von Schattenbibliotheken angegeben. Es soll jedoch betont werden, dass Letzteres auf Spekulationen beruht. Für eine generalisierbare Aussage fehlen genauere Fragestellungen im Fragebogen.

4.1.1.4 Fragegruppe AB4

Die letzte Fragegruppe bezog sich auf die Dienstleistungen der Bibliotheken und inwiefern diese in Zusammenhang mit dem Deep Web bereits in Anspruch genommen wurden. Weder Gruppe A (16.36%) noch Gruppe B (10%) ersuchten sich zu hohen Teilen Hilfe bei Bibliothekar*innen. Jedoch wurden jene Anliegen zu 100% (A) vollständig und 87.5% (B) vollständig oder teilweise vollständig erfolgreich gelöst.

Als letzte Frage der Fragegruppe und gleichzeitig des gesamten Fragebogens wurde nach dem Wunsch nach mehr Deep-Web-Unterstützung durch Bibliotheken gefragt («Wünschen Sie sich mehr oder tiefergehende Unterstützung von Bibliotheken bei der Recherche im Deep Web?»).

Gruppe A antwortete mehrheitlich mit Nein, wobei trotz der hohen Vertretung von Informationswissenschafts-Studierenden 38.18% mit Ja abstimmten (s. Abb. 13).

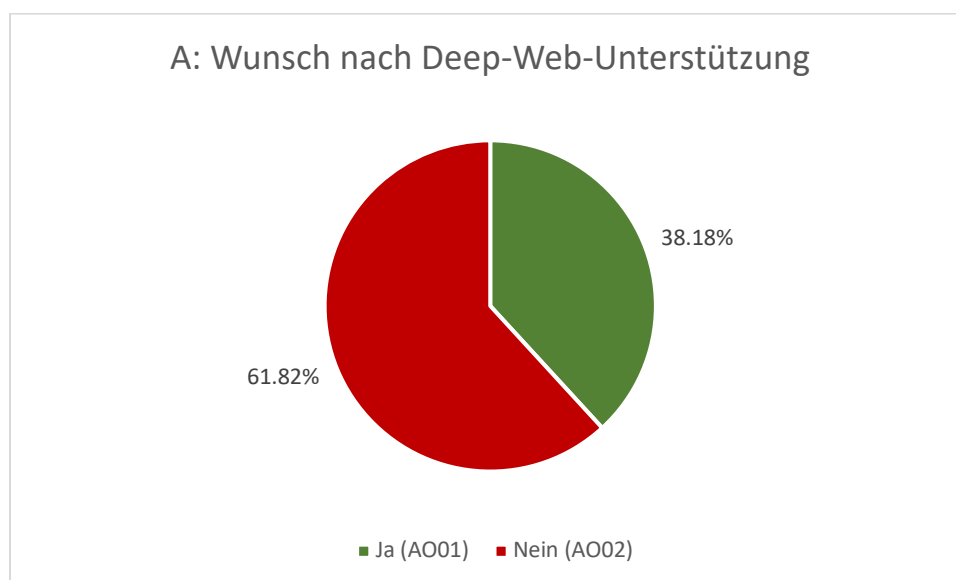


Abbildung 13. Wunsch nach Unterstützung bei der Recherche im Deep Web durch Bibliotheken, Gruppe A (eigene Darstellung)

Auswahl an Kommentaren von Gruppe A zu Abbildung 13:

- [Nein] «Zurzeit nicht, aber wahrscheinlich wäre es sinnvoll, wenn ich dann die Bachelorarbeit schreibe und auf Literaturdatenbanken zugreifen muss und etwas doch nicht auf Scihub/Libgen zu finden ist.»
- [Ja] «weil es eine viel ausführlichere Recherche ermöglichen würde und ich gerne bereits früher über diese Möglichkeiten Bescheid gewusst hätte»
- [Ja] «ein Punkt warum das Deep Web nicht genutzt wird, ist wahrscheinlich, weil Personen es nicht kennen und was man nicht kennt nutzt man weniger»

- [Nein] «Bisher war ich nicht gross auf das Deep Web angewiesen, daher reicht es mir momentan, zu wissen, wie man damit umgeht.»
- [Ja] «Jein, nicht direkt bei der Bereitstellung bestimmter Literatur für eine Arbeit sondern generell zur Horizonterweiterung, um mehr solcher Tools kennenzulernen.»
- [Ja] «Ich denke, dass ist abhängig vom Thema. Es ist gut zu wissen, dass man bei Schwierigkeiten jemanden als Unterstützung hat, wo andere Suchansätze und mehr Kenntnisse hat.»
- [Ja] «Ich denke es ist allgemein ein wichtiges Thema, bei dem ich persönlich das Gefühl habe, dass ich viel zu wenig darüber weiss. Ich denke nicht, dass es zwingend die Aufgaben von Bibliotheken ist, dieses Wissen zu vermitteln aber bestimmt wäre es eine Möglichkeit.»
- [Ja] «Ich denke es wäre sicher hilfreich wenn die Bibliotheken dafür ein Informationsangebot hätten. Ich denke es wäre auch wichtig darauf aufmerksam zu machen, damit die Leute überhaupt Bescheid wissen das es das Angebot gibt. Dies würde Studierenden vielleicht auch helfen, überhaupt daran zu denken das Deep Web in der Recherche zu bedenken.»
- [Ja] «Meine Erwartung ist, dass die Bibliotheken die eigenen Fachdatenbanken kennen und einen allgemeinen Marktüberblick haben, um mich an geeignete Stelle weiterzuverweisen. Ich erwarte auch Anleitungen und Zusammenfassungen zu den lizenzierten Ressourcen auf der Webseite der Bibliothek und eine Liste / Datenbank zum suchen von Fachdatenbanken.»

Ein Grossteil der Nein-Stimmen argumentierte mit dem bereits vorhandenen Wissen durch einen I+D-Hintergrund oder durch die von der Fachhochschul-Bibliothek zur Verfügung gestellten Ressourcen und Einführungen. Die Ja-Stimmen betonten mehrheitlich die aus ihrer Sicht wichtige Option dieses Angebots, mehr Informationen und die Verbreitung des Wissens über das Deep Web.

Knapp die Mehrheit (51.25%) antwortete bei Gruppe B mit Ja. Da dies die Gruppe mit der höheren Studien- und Forschungstätigkeit ist, sind diese Resultate signifikant und deuten auf eine zu berücksichtigende Nachfrage von Recherche-Unterstützung im Deep Web hin (s. Abb. 14).

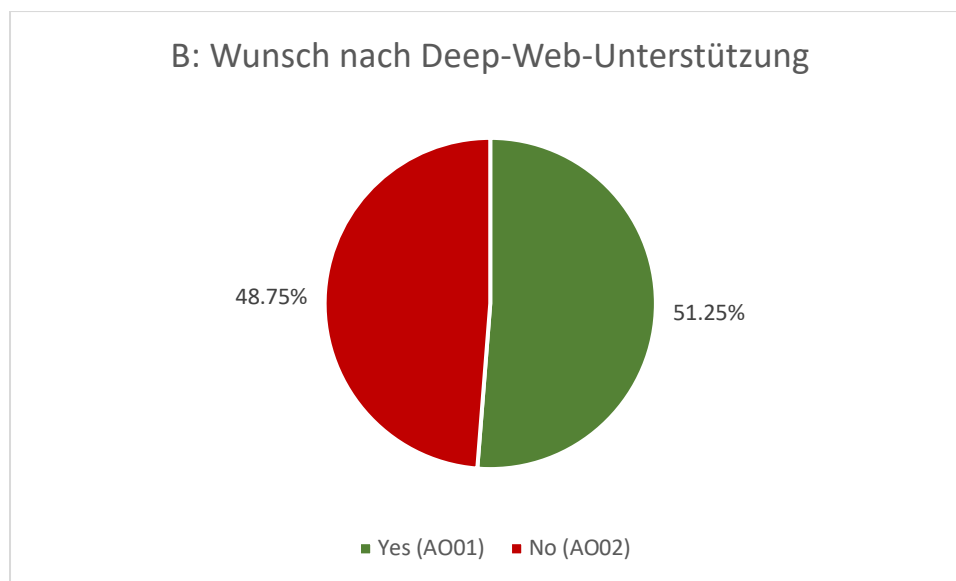


Abbildung 14. Wunsch nach Unterstützung bei der Recherche im Deep Web durch Bibliotheken, Gruppe B (eigene Darstellung)

Auswahl an Kommentaren von Gruppe B zu Abbildung 14:

- [Ja] «some advice how to proceed would be helpful»
- [Ja] «UniBern is quite scattered in departments and personally I have seen books in the library here but never a librarian.»
- [Nein] «I guess it could've been nice as part of my degree, but right now I'll make do with the research habits I have.»
- [Ja] «Maybe this is a terminology thing, but I'd love to hear more about deep-web specifically. I do use the Uni library website to search for articles, but don't know if this is what is referred to as the deep web.»
- [Ja] «I'm curious to know about the advantages of it, as I'm not familiar with it yet.»
- [Ja] «It will be good to have an orientation on how to use these resources when we register»

Weitere Kommentare deuten erneut darauf hin, dass das Deep Web an sich anhand der gegebenen Definition aus Fragegruppe AB2 nicht vollständig verstanden wurde. In Zusammenhang damit stellt sich die Frage, ob die Umfrage-Teilnehmenden ansonsten zu mehr Anteilen mit Ja geantwortet hätten.

4.1.2 Auswertung C

Der Fragebogen für die Bibliothekar*innen hatte zusätzlich zur thematischen Datenerhebung Fragen zum beruflichen Werdegang von Bibliotheken. Folgende Bibliothekstypen waren vertreten: Öffentliche Bibliothek (16.67%), Wissenschaftliche Bibliothek (61.76%), Spezialbibliothek (22.55%), Sonstiges (6.86%).

4.1.2.1 Fragegruppe C1

Bei der Frage zur Häufigkeit des Einsatzes von Google (oder einer anderen Suchmaschine) im Rahmen einer Benutzendenanfrage gab ein Grossteil als Antwort «Täglich» (35.29%) und «Mehrere Male pro Woche» an (24.51%) (s. Abb. 15).

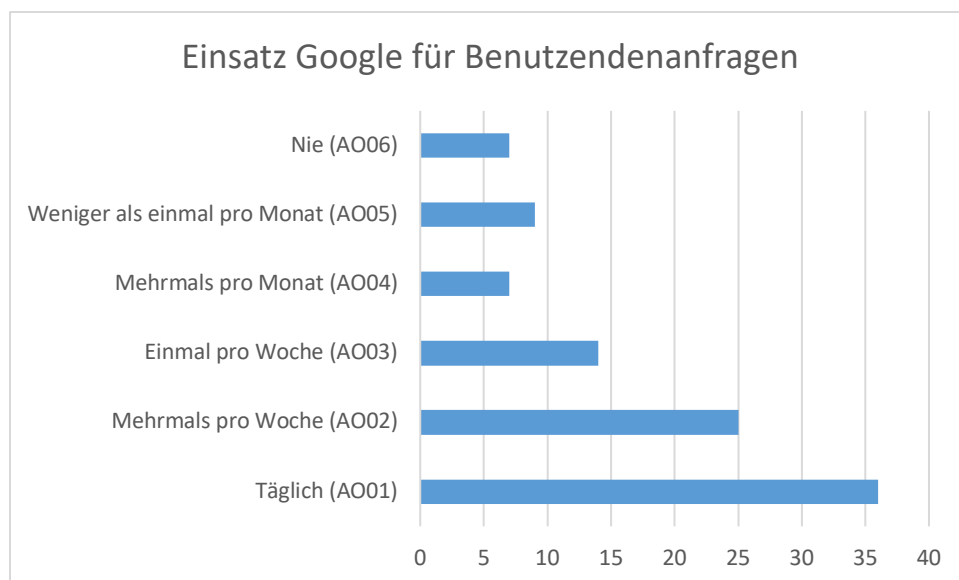


Abbildung 15. Häufigkeit der Nutzung von Google für Benutzendenanfragen (eigene Darstellung)

Dies deckt sich mit den Antworten der vorhergehenden Frage, die nach dem Suchmaschinen-Verhalten im allgemeinen Berufsalltag fragte. Die häufigsten Kommentare beinhalteten die generelle Informationssuche, kleinere bis grössere thematische Suchen, erste Sucheinstiege für Recherchen und das Finden von bibliografischen Nachweisen (s. Anhang 12_Statistiken).

4.1.2.2 Fragegruppe C2

Im Fragenset zum Deep Web gaben 95.10% der Bibliothekar*innen an, bereits einmal davon gehört zu haben. Zudem stimmten 70.59% der angegebenen Definition für das Deep Web zu, 24.51% teilweise. Davon gaben 11.76% korrekte Ergänzungen zur Definition an – da diese keine vollständige Aufzählung ist, deutet das auf ein grundlegendes theoretisches und praktisches Verständnis hin.

Etwas weniger als die Hälfte benutzen das Deep Web täglich im Zusammenhang mit einer Benutzendenanfrage oder in einem anderweitigen Kontext ihrer Bibliothek (s. Abb. 16), 22.55% mehrmals pro Woche. Da Bibliothekar*innen häufig wöchentlich oder alle paar Tage in Schichten an der Auskunftstheke oder im Kundenservice arbeiten, werden diese beiden Resultate betreffend der Bedeutung ihrer Häufigkeit als gleichwertig angesehen.

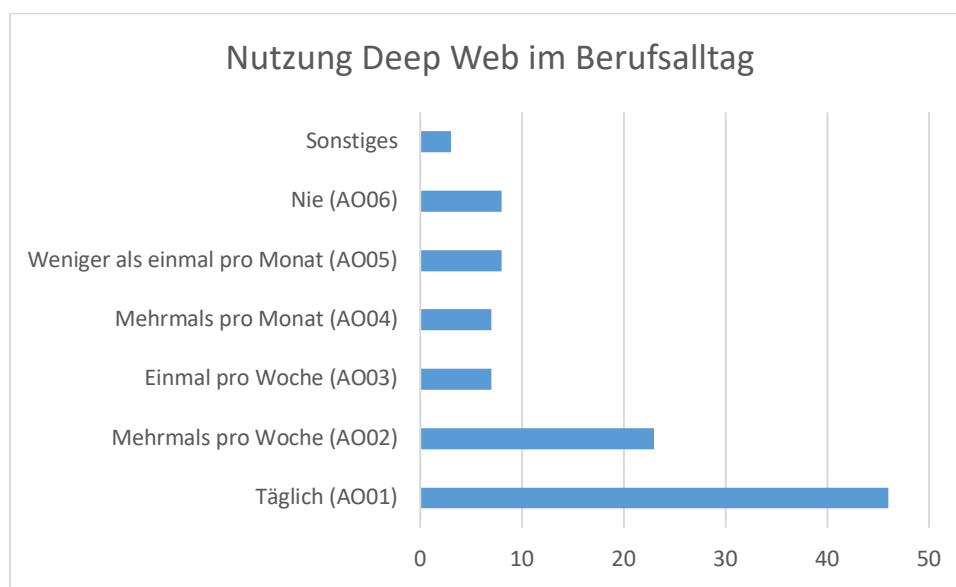


Abbildung 16. Häufigkeit der Nutzung des Deep Webs im Berufsalltag (eigene Darstellung)

Nicht übersehen werden sollten die Antworten zu den geringen Häufigkeiten, AO05 und AO06. Damit sind 15.86% nie oder selten mit dem Deep Web in Kontakt. Aus den Tätigkeitsbereichen der korrespondierenden Antworten kann keine direkte Begründung dafür gefunden werden. Die Interpretation liegt hier bei diversen Einflussfaktoren, wie beispielsweise kein Kundenkontakt, andere Wege der Informationsrecherche oder Unwissen über die Definitionsgrenze des Deep Webs.

Die Kenntnis über den Begriff des Dark Webs liegt mit 98.04% etwas höher als jene des Deep Webs. Zugriffe darauf fanden jedoch mit 10.78% nur wenige statt, wobei sich 4.9% der Antwort enthielten (aufgrund des potenziell sensiblen Charakters des Themas wurde diese Frage im Fragebogen als freiwillig markiert).

4.1.2.3 Fragegruppe C3

In einem letzten Schritt wurden die Bibliothekar*innen zu den Schattenbibliotheken befragt. Dabei ergab sich ein Unwissen darüber bei 31.37% (32). Aus dem Tätigkeitsbereich der Befragten, dem Bibliothekstyp ihrer Anstellung und ihrem

Ausbildungshintergrund lassen sich keine Aussagen zu Zusammenhängen ziehen, jedoch bei der Altersverteilung der N=32 Nein-Stimmen (s. Abb. 17):

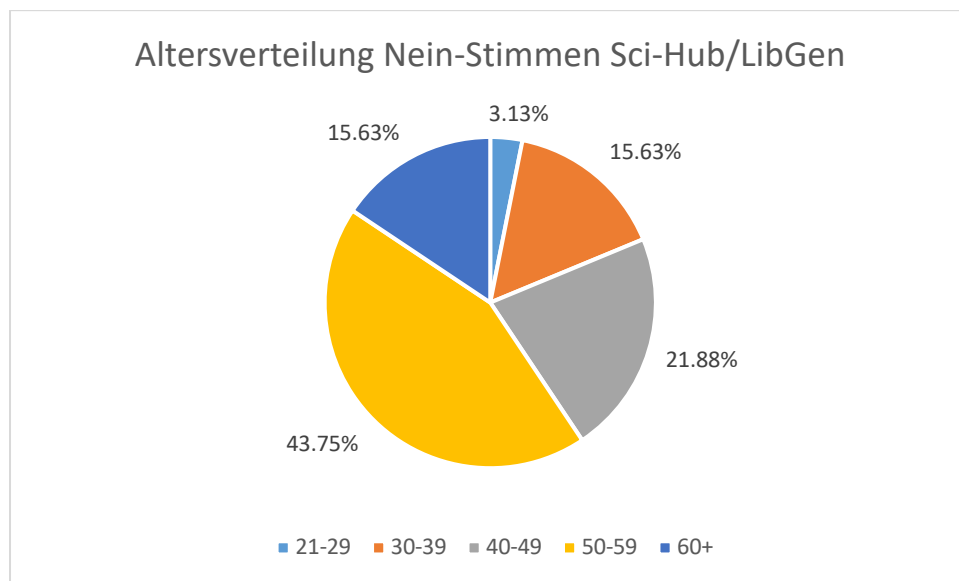


Abbildung 17. N=32, Altersverteilung Nein-Stimmen zur Kenntnis von Sci-Hub/LibGen (eigene Darstellung)

14 der 32 Personen (43.75%) sind zwischen 50 und 59 Jahre alt, mit 21.88% sind die 40-49-jährigen an zweiter Stelle. Dies kann mit der Altersverteilung der gesamten Umfrage C ins Verhältnis gesetzt werden (s. Abb. 18): 33.33% sind 50-59-jährig, was gleichzeitig der grösste Anteil der Nein-Stimmen ausmacht. Dicht gefolgt jedoch sind die 40-49-jährigen, die in der Grundgesamtheit 20.59% vertreten und damit einen kleineren Anteil als die 30-39-jährigen (30.39%) innehaben. Letztere machen nur 15.63% der Nein-Stimmen aus. Einen ebenso kleinen Anteil haben die 60+-jährigen, was allerdings primär auf die geringen Teilnehmendenzahlen in dieser Altersgruppe zurückzuführen ist (7 bzw. 6.86%); 5 dieser 7 Befragten stimmten mit Nein.

Daraus lassen sich erste Rückschlüsse auf das Alter oder die Generationen an Bibliothekar*innen und deren Kenntnisse über Sci-Hub und Library Genesis ziehen: Je älter die Bibliothekar*innen, desto weniger bekannt sind Schattenbibliotheken. Da diese in direktem Zusammenhang mit dem Deep Web stehen, das traditionelle Publikationsmodell der akademischen Forschung beeinflussen und beim Zielpublikum weit verbreitet sind, besteht eine hohe Gewichtung auf der Notwendigkeit der Kenntnis und Problemerkennung von Schattenbibliotheken. Hier kann somit ein Defizit gegenüber den Praktiken der Kundschaft identifiziert werden.

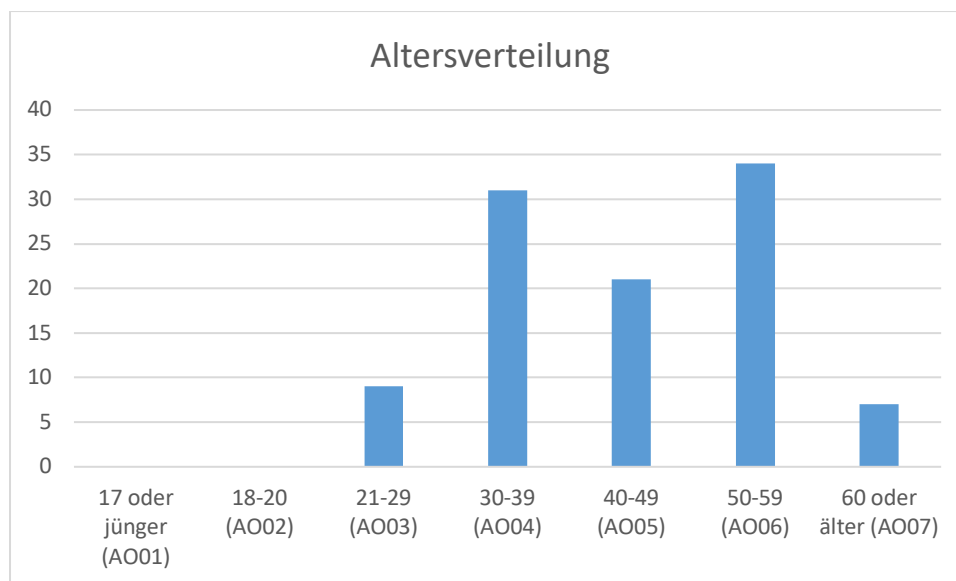


Abbildung 18. Altersverteilung der gesamten Umfrage C (eigene Darstellung)

25 der 70 Ja-stimmenden Personen wurde mindestens einmal eine Frage oder Aussage zu Sci-Hub und/oder LibGen durch einen Benutzenden gestellt, was 35.71% bei N=70 und 24.51% bei N=102 sind. 41 Befragte sahen sich im Kundenkontakt noch nie damit konfrontiert (58.57% bei N=70, 40.2% bei N=102) und 4 können sich nicht mehr erinnern. 30.39% aller Befragten haben Sci-Hub und/oder Library Genesis ausserdem selbst schon benutzt.

Auswahl an Kommentaren dazu (Frage: «Wurde Ihnen jemals eine Frage zu Sci-Hub/Library Genesis durch einen Benutzenden gestellt oder eine Aussage dazu gemacht?»):

- «Nicht direkt zu diesen Plattformen. Mehr allgemein, warum gewisse Daten nicht grundsätzlich als open access zur Verfügung stehen.»
- «Obwohl wir die Weisung haben, nicht mit Kunden über Sci-Hub zu reden, haben mich schon mehrere Benutzende darauf angesprochen. Ich erklärte ihnen kurz die Situation und Möglichkeiten von Sci-Hub, habe sie dann aber über unsere bibliotheksinternen Möglichkeiten zur Suche im Deep Web aufgeklärt und auch gleich gezeigt, wo sie entsprechende Artikel o.ä. finden. (Datenbankrecherche etc..)»
- «Die Studenten kennen Sci-Hub und interessieren sich dafür, ob es legal ist.»
- «Ich habe schon gehört, dass Sci Hub verwendet wird. Wir dürfen in der Bibliothek darauf nicht zugreifen (Weisung).»
- «Habe von Benutzenden gehört, dass sie Artikel auf Sci-Hub herunterladen, selbst wenn ihre Institution diese Artikel lizenziert hat. Grund: Benutzungsfreundlichkeit (kein langes Herumklicken, keine Registrierung oder Anmeldung notwendig).»

Dies lässt vermuten, dass Sci-Hub und LibGen in Bibliotheken den Benutzenden gegenüber von Weisungs-wegen her ein Tabu-Thema sind. Dazu müsste jedoch eine zusätzliche Erhebung erfolgen, um genauere Aussagen zu treffen.

Beiden Untergruppen – diejenigen, die Sci-Hub/LibGen kennen (C01) und jene die noch nie davon gehört haben (C02) – wurde als Abschluss die Frage gestellt, weshalb Benutzende die Schattenbibliotheken den Lizenzen der Bibliotheken vorziehen könnten. C02 bekamen hierfür eine kurze Definition der Begriffe und zu was genau die Schattenbibliotheken in Gebrauch sind. So konnte eruiert werden, was C02 ohne Erfahrung der Webseiten als Gründe vermutet und was C01 aus eigener Erfahrung oder aus dem Kundenkontakt mitbekommen hat. Die Resultate unterschieden sich bei mehreren Antwortmöglichkeiten (s. Abb. 19; 20):

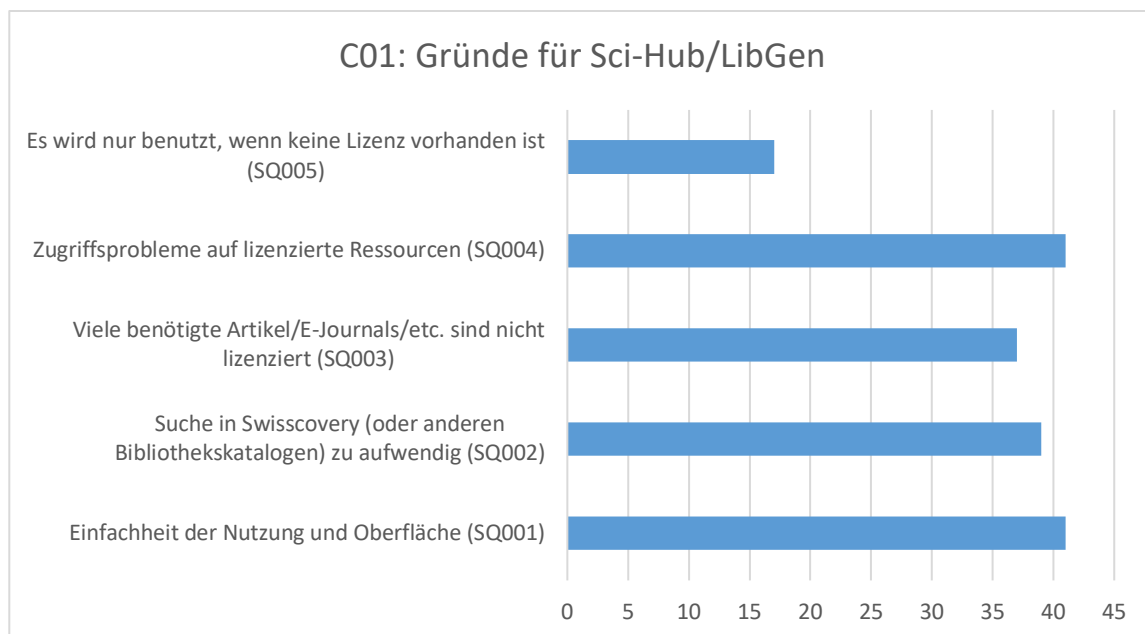


Abbildung 19. Vermutete Gründe für die Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Bibliotheken, Untergruppe C01 (eigene Darstellung)

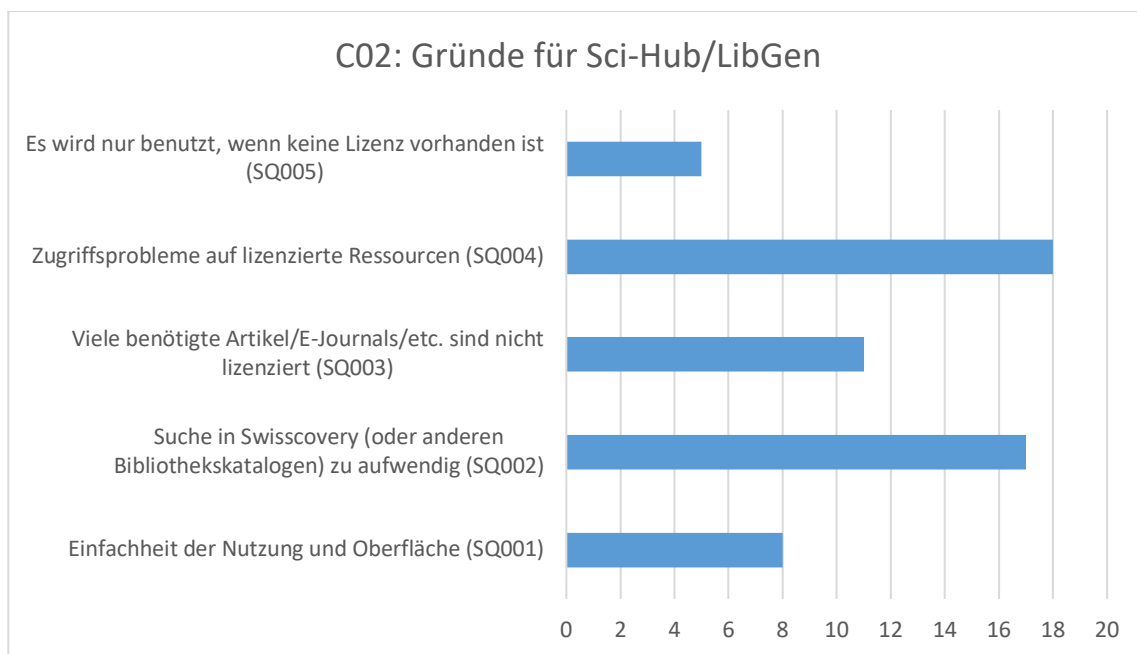


Abbildung 20. Vermutete Gründe für die Nutzung von Sci-Hub/LibGen gegenüber den Lizenzen der Bibliotheken, Untergruppe C02 (eigene Darstellung)

Während die Mehrheit von C01 alle Optionen abgesehen von 5Q005 als gleichwertige mögliche Gründe erachtete, wurden bei C02 die Antworten 5Q005 und SQ001 weniger häufig ausgewählt. Ebenso Antwort SQ003 im Vergleich zu C01 – dies kann jedoch auf das mangelnde Wissen von C02 über das Konzept und die Problematik hinter dem Phänomen der Schattenbibliotheken zurückzuführen sein. Ausserdem ist SQ001 durch C02 nicht direkt beantwortbar (UI nicht bekannt). Auffallend ist, dass beide Untergruppen Antwort 5Q005 weniger stark gewichteten, im starken Kontrast zu den Studierenden und zum Mittelbau.

4.2 Analyse Sci-Hub und Library Genesis

Im Gegensatz zu Sci-Hub führt Library Genesis keine aktuellen und einsehbaren Statistiken. Bodó analysierte LibGen jedoch von 2008 bis 2014 und dokumentierte die folgenden Zahlen (S. 53-76):

- Wachstum von 34'000 (2008) zu 1.2 Millionen Dokumenten (2014).
- Meistgefragte Fachgebiete nach DDC: Nicht klassifiziert (31%) Sozialwissenschaften und Anthropologie (15%), Technologie (14%), Naturwissenschaften und Mathematik (9%), Literatur, Rhetorik und Kritik (9%).
- Russland verzeichnete die meisten Downloads (31%), gefolgt von den USA (8%), Indonesien (6%), Indien (5%) und der Ukraine (5%).

- Die meisten Downloads pro 1'000 Einwohner hatte erneut Russland mit 6.0, anschliessend Litauen (5.5), Estland (4.2), Schweden (3.7), Lettland (3.4) und Ukraine (3.0).

Sci-Hub berichtet auf deren Webseite «sci-hub.se» über detaillierte Statistiken zur Art und Herkunft ihrer Nutzendenzahlen (Sci-Hub, 2023b). Insgesamt befinden sich 88'343'822 Dokumente in ihrer Datenbank, mit der grössten Verteilung in den Fachrichtungen Medizin (24'557'530), Chemie (16'460'921), Biologie (15'499'507), Geisteswissenschaften (12'592'316), Physik (8'658'518), Ingenieurwissenschaften (6'892'853), Mathematik (3'676'789), Ökologie (3'676'789), Informatik (2'768'241), Wirtschaft (2'572'842) und Geowissenschaften (2'571'177). Ca. 80% der Dokumente sind in Zeitschriften veröffentlichte Forschungsartikel, 6% Konferenzschriften, 5% Buchkapitel und 9% andere Dokumententypen. Die Deckung von wissenschaftlichen Verlagen beträgt dabei über 95%. 77% der Inhalte wurden zwischen 1980 und 2020 veröffentlicht, 36% zwischen 2010 und 2020.

Die Länderverteilung mit den höchsten Downloadzahlen des Monats Juli 2023 gestalten sich wie in nachfolgender Tabelle 4 aufgeführt:

Nr.	Land	Downloads	Nr.	Land	Downloads
1	China	44,157,924	11	Irland	1,067,240
2	USA	24,623,949	12	Neuseeland	910,454
3	Brasilien	7,861,862	13	Iran	844,172
4	Indien	3,006,688	14	Frankreich	840,556
5	Russland	2,310,458	15	Niederlande	830,045
6	Deutschland	1,474,928	16	Mexiko	825,879
7	Indonesien	1,394,804	17	Singapur	818,229
8	Kanada	1,219,385	18	Ukraine	745,201
9	Australien	1,131,649	19	Japan	698,761
10	UK	1,131,319	20	Malaysia	634,979

Tabelle 4. Länderverteilung der Anzahl Downloads, Juli 2023 (eigene Darstellung nach Sci-Hub, 2023c)

Die Schweiz befindet sich mit 136'277 Downloads im Monat Juli 2023 auf Platz 50 von 234. Zugriffe via VPN sind in dieser Statistik nicht separat aufgeführt; somit können eine gewisse Anzahl an Downloads effektiv aus anderen Ländern stammen als in der Tabelle angegeben. Ebenso sind Zahlen von Sci-Hub-Mirrors nicht inbegriffen, die in Ländern mit

gesperrten offiziellen Domains (sci-hub.se, sci-hub.st, sci-hub.ru) relativ hoch sein können (Sci-Hub, 2023c).

Für die Statistik der Universität Bern wurden in Kooperation mit den Informatikdiensten (ID Security) einen Monat lang, vom 12.06.2023 bis 12.07.2023, die Zugriffe auf die Domains sci-hub.se und libgen.li überwacht. Dafür wurde eine Regel auf der Firewall eingerichtet und so das gesamte Netzwerk der Universität, inklusive VPN, eingeschlossen. Für die offiziellen Domains wurden die Ports HTTP und HTTPS gelistet (s. Tabelle 5).

Destination FQDN	IP	# Hit Count
libgen.li	89.248.170.59	247
sci-hub.se	186.2.163.219	1845
scihub.se	199.59.243.223	2768

Tabelle 5. Zugriffe auf Sci-Hub und Library Genesis aus dem Netzwerk der Universität Bern, 12.06.-12.07.23 (eigene Darstellung nach N. Zegarac, persönl. Mitteilung, 13.07.2023a)

Zusätzlich erkannte die ID Security zahlreiche Zugriffe auf Mirror-Seiten der beiden Schattenbibliotheken, die teilweise um ein Vielfaches höhere Hit Counts erreichten (s. Tabelle 6; 7). Bei diesen wurden jedoch alle Ports miteingeschlossen, somit können sie nicht direkt mit den offiziellen Domains verglichen werden. Es handelt sich bei den Tabellen um eine Auswahl der Mirror-Seiten.

Destination FQDN	IP	# Hit Count
sci-hub.hkvisa.net	172.67.170.69	128
sci-hub.hkvisa.net	104.21.63.75	180
sci-hub.ee	104.21.10.133	9
sci-hub.ee	172.67.190.44	18
sci-hub.pub	172.67.223.15	1
sci-hub.tf	104.21.65.213	2
sci-hub.ren	104.21.70.220	2
sci-hub.ren	172.67.139.245	13
sci-hub.shop	104.21.9.230	212
sci-hub.shop	172.67.161.98	196
sci-hub.wf	104.21.93.119	22
sci-hub.wf	172.67.209.211	12
sci-hub.scihubtw.tw	104.21.94.48	16

sci-hub.live	141.164.37.59	1
sci-hub.it.nf	144.208.71.125	2
sci-hub.mksa.top	172.67.128.122	18
sci-hub.mksa.top	104.21.73.88	44
img.sci-hub.shop.cdn.cloudflare.net	172.67.161.98	196
sci-hub.3800808.com	172.67.176.155	5
sci-hub.3800808.com	104.21.83.133	3
sci-hub.tw	185.53.177.50	480
sci-hub.bz	185.53.177.52	44
zero.sci-hub.se	186.2.163.169	174
dacemirror.sci-hub.se	186.2.163.170	194
sci-hub.st	186.2.163.201	501
scihub.wikicn.top	172.67.152.105	14
scihub.wikicn.top	104.21.1.228	9
sci-hub.scihubtw.tw	104.21.94.48	16
sci-hub.scihubtw.tw	146.59.215.226	52
sci-hub.scihubtw.tw	162.241.8.17	1
scihubtw.tw	104.21.94.48	16
scihub.se	199.59.243.223	15'909
scihub.com	37.48.65.144	63
scihub.net.cn	59.110.54.99	1
whereisscihub.vercel.app	76.76.21.142	1558
whereisscihub.vercel.app	76.76.21.241	1310
sci-hub.ru	190.115.31.218	374
zero.sci-hub.st	190.115.31.42	20
dacemirror.sci-hub.st	190.115.31.43	16
ww25.sci-hub.do	199.59.243.223	15916
sci-hub.cc	69.16.230.228	6
sci-hub.41610.org	76.76.21.241	1322
sci-hub.41610.org	76.76.21.22	1823

Tabelle 6. nach N. Zegarac, persönl. Mitteilung, 13.07.2023b)

Destination FQDN	IP	# Hit Count
libgen.unblockit.app	103.224.182.238	3
libgen.ggfwws.net	104.21.13.236	19
libgen.ggfwws.net	172.67.133.110	11
libgen.unblockit.pet	172.67.154.204	1
libgen.rocks	104.21.68.171	46
libgen.rocks	172.67.197.62	39
libgen.onl	104.21.9.42	1
libgen.onl	45.33.2.79	298
libgen.me	104.21.95.37	2
libgen.fun	162.244.82.96	4
libgen.is	193.218.118.42	1463
libgen.fr	199.59.243.223	15914
libgen.org	217.19.248.132	258
libgen.it	45.79.19.196	478
libgen.it	198.58.118.167	924
libgen.lc	89.248.170.56	84
libgen.li	89.248.170.59	247
libgen.gs	89.248.170.61	6

Tabelle 7. Zugriffe auf Mirror-Seiten von LibGen aus dem Netz der Universität Bern, alle Ports (eigene Darstellung nach N. Zegarac, persönl. Mitteilung, 13.07.2023b)

Die 15'909 Zugriffe auf scihub.se können einerseits auf die falsche Eingabe der offiziellen Domain zurückgeführt werden, andererseits besteht die Wahrscheinlichkeit einer bewusst erstellten Mirror-Seite. Keine der stichprobenmässig ausprobierten Mirror-Seiten konnte jedoch erreicht werden, demnach könnte es sich bei den aufgelisteten Seiten aus Tabelle 6 und 7 um veraltete oder früher aktive Domains handeln, die noch immer aufgerufen werden. Ebenso erdenklich ist eine hohe Gewichtung der Cyber Security seitens jener Benutzenden und die dementsprechende Präferenz zur Nutzung von Mirrors. Die ID Security konnte keine Erklärung finden, weshalb zahlreiche Mirror-Seiten anstatt der offiziellen Domains – welche frei zugänglich und nicht durch die Firewall der UNIBE gesperrt sind – benutzt werden (N. Zegarac, persönl. Mitteilung, 13.07.2023b).

Da es sich bei diesen Statistiken um Zugriffe auf die URLs handelt, kann keine Aussage zu den eigentlichen Downloads gemacht werden. Dennoch ist die Wahrscheinlichkeit gross, dass ein Aufrufen einer Schattenbibliothek aus dem Netz einer Universität mit der Absicht auf das Suchen und Herunterladen einer Datei getätigt wird. Die Datenerhebung

fand bereits während der Semesterferien statt und erzielte trotzdem kombinierte Zugriffe beider Schattenbibliotheken von 2'092 – während der Prüfungsphase werden die Zahlen als entsprechend höher vermutet.

5 Diskussion

Die Zusammentragung des Deep Webs in der Literatur und die in dieser Arbeit behandelten Ergebnisse zeigen auf, dass eine effiziente und effektive Recherche das gesamte Web umfasst und auf den individuellen Informationsbedarf abgestimmt werden muss. Wie Coffey (2020, S. 2) beschreibt, braucht es Expert*innen, die diesen Bedarf identifizieren und extrahieren können. Wie die Umfrage bei den Bibliothekar*innen ergab, ist das grundlegende Verständnis für das Deep Web bereits vorhanden. Der Stand besteht demnach nicht direkt in der Kenntnis des Deep Webs selbst, sondern ob und inwiefern dieses Wissen auf die jeweilige Nachfrage hin passend vermittelt wird.

Jede der befragten Gruppen zeigte Gemeinsamkeiten und Unterschiede auf, die mit ihren jeweiligen Fachrichtungen in Verbindung gebracht werden können. So konzentrieren sich die Recherchen im Surface Web im Zusammenhang mit der jeweiligen Studien- oder Forschungsrichtung bei den Studierenden auf schnelle und allgemein thematische Suchen, während der Mittelbau vor allem die Artikelsuche darüber tätigt. Bibliothekar*innen nutzen das Surface Web sowohl für allgemeine Suchen im Berufsalltag als auch für Recherchen und erste Ansätze dafür.

Kenntnisse über den Begriff des Deep Webs waren bei den Studierenden und Bibliothekar*innen hoch, der Mittelbau hat zur Hälfte schon einmal davon gehört. Die Nutzung zeichnete sich beim Mittelbau als vergleichsweise tief ab, was jedoch mit dem mangelnden Verständnis des Begriffs in Verbindung gebracht werden kann (Hinweis darauf aus mehreren Kommentaren, s. Anhang 12_Statistiken): Bei der Nutzung von Suchmaschinen (inkl. Google Scholar) wurde ein täglicher Gebrauch von 81.25% angegeben. Jedoch geht aus der im Fragebogen angegebenen Definition nicht hervor, dass Google Scholar auf das Deep Web zugreift und somit ebenfalls zur Nutzung gezählt werden kann. Ebenso sind Zugriffe auf das Deep Web von Google selbst möglich, sofern der Institutionszugang zur Überbrückung der Paywall von der entsprechenden Verlagswebseite gegeben ist. Die effektive Häufigkeit der Nutzung des Deep Webs durch den Mittelbau kann somit näher der Prozentzahl der Häufigkeit der allgemeinen Suchmaschinen-Nutzung geschätzt werden. Aus den Kommentaren der Studierenden und dem Mittelbau zeichnet sich ausserdem ab, dass sie Unterstützungsmöglichkeiten durch Bibliotheken als sinnvoll erachten und an den vertieften Recherchemöglichkeiten interessiert wären (überwiegend Mittelbau). Gleichzeitig geht jedoch analog der Theorie von Devine und Egger-Sider (2014, S. 23) hervor, dass die bisherigen eigenen Recherchemethoden und -wege als genügend eingestuft werden (überwiegend Studierende). Auch hier stellt sich die Frage, ob der Mittelbau umso mehr an Unterstützung interessiert wäre, wenn die Inhalte und

Ressourcen des Deep Webs klarer vermittelt würden. Aus den Kommentaren lässt sich auch eine mangelnde Wahrnehmung der bereits vorhandenen bibliothekarischen Unterstützung ableiten.

Die signifikantesten Differenzen finden sich betreffend den Schattenbibliotheken. Während die Studierenden und Bibliothekar*innen Sci-Hub und Library Genesis zwischen 68% und 72% kennen, sind es beim Mittelbau 88.75%. Das kann auf die erhöhte Studien- und Forschungstätigkeit des Mittelbaus und das verstärkte akademische Umfeld zurückgeführt werden. Insbesondere sticht jedoch bei den Gründen zur Nutzung von Schattenbibliotheken gegenüber den Lizenzen der Universität/Fachhochschule/Bibliothek die Antwortmöglichkeit SQ002 («Suche in Swisscovery (Bibliothekskatalog) zu aufwendig») hervor: Sowohl Studierende als auch Mittelbau wählten dies mit Abstand am wenigsten als Grund an, während Bibliothekar*innen diese Option als gleichwertig (Untergruppe C01) oder gar am zweithäufigsten anwählten (Untergruppe C02) – direkt nach den Zugriffsproblemen auf bereits verfügbare Lizenzen (SQ004). Die Vermutung liegt nahe, dass Untergruppe C02 davon ausging, der primäre Weg zu lizenzierten Ressourcen führe bei der Kundschaft über den Bibliothekskatalog. Vielmehr wird allerdings direkt über Google, Google Scholar oder den Datenbanken selbst auf das Deep Web zugegriffen. Mit dieser Einschätzung liesse sich auch die am wenigsten angewählte Option SQ002 durch Gruppe A und B erklären. Dagegen ist auffallend, dass der Mittelbau die Option SQ005 («I only use it if the University does not have a license») am meisten anwählte. Dies kann mit den breitflächigen Lizenzierungen der Universität Bern zusammenhängen, insbesondere im Vergleich zur FHGR und den am höchsten gewichteten Optionen SQ003 und SQ004 durch die Studierenden. Die Einfachheit der Nutzung und Oberfläche wird vom Mittelbau zudem vergleichsweise gering gewertet, im Kontrast zu den Vermutungen der Bibliothekar*innen dazu. Auch hier kann jedoch das Argument von Google Scholar gebracht werden, das direkte Zugriffe zu Volltexten mit Institutionszugang ebenfalls benutzerfreundlich gestaltet.

Um die Cognitive Invisibility zu überwinden, sind sichtbare und ansprechende Unterstützungsangebote vonnöten. Aus den Ergebnissen der Umfrage kam hervor, dass Angebot und Nachfrage von beiden Seiten her da sind, sich jedoch nicht genügend zu kreuzen scheinen. Die teilnehmenden Bibliothekar*innen, die meisten davon mit akademischem Hintergrund in den Informationswissenschaften, wissen um die grundlegenden Aspekte des Deep Webs und wären dazu in der Lage, entsprechend erste Ansätze in der Unterstützung zu liefern. Die Frage ist jedoch, ob sie sich selbst als Lieferanten dieses Wissens sehen und ob Bibliotheken tatsächlich die Verantwortlichen für den Wissenstransfer des Deep Webs sind (wurde in der Umfrage durch jemanden aus der Gruppe der

Studierenden in Frage gestellt). Die Erarbeitung des Forschungsstandes und der theoretischen Grundlagen der Internetarchitektur deuten auf ein Ja. Ein Grossteil der Literatur zum Deep Web stammt aus Bereichen des Bibliothekswesens oder generell dem informationswissenschaftlichen Bereich. Es werden auch zusätzliche Aspekte des Deep Webs thematisiert, die in den Augen jener Autoren ebenfalls zur Aufgabe von Bibliotheken gehören sollten und in die sich wandelnde Rolle von Informationsinstitutionen im digitalen Zeitalter hineinspielen: Cyber Security und Privatsphäre im Internet.

Im Westen mögen Dark-Web-Technologien wenig Anklang finden, da Pressefreiheit einen vergleichsweise hohen Stellenwert hat und Privatsphäre und Schutz von personenbezogenen Daten im Gesetz verankert sind. Von Zensur betroffene Länder sind jedoch darauf und auf die Nationen, die die Technologien ermöglichen, angewiesen. Die Rolle von Bibliotheken kann in diesem Zusammenhang zukunftsgerichtet weltweit betrachtet und überdacht werden. Sie sind Träger und Vermittler von Information und sollten sich mit Privatsphäre und Fragen der Überwachung und Zensur der freien Informationen auseinandersetzen. Ein breites Wissen zu Technologien des Webs ist dabei Voraussetzung und rücken Bibliotheken in das Zentrum der Informationsfreiheit und Informationsdemokratie.

In der Diskussion zur Implementierung von Tor in Bibliotheken darf jedoch die negative Seite von Tor nicht ignoriert werden. Der Browser bietet Zugang zu Onion Services; geschieht ein Zugriff darauf in einer Bibliothek oder wird diese gar gezielt dafür genutzt (nicht eigener Computer, an öffentlichem Standort, heisst zusätzliche Anonymität), ist das der gegenteilige Effekt des eigentlichen Vorhabens. Eine mögliche Lösung wäre eine speziell zugeschnittene Version von Tor, eventuell in Kollaboration mit The Tor Project, oder eine Konfiguration über den Tor-Browser gelegt. Hierbei schwindet jedoch der Vorteil der Einfachheit von Tor. In Bezug auf das Library Freedom Project stellen sich ähnliche Fragen, mitunter moralische: Würde das Hosting von Tor-Endknoten in Bibliotheken nicht gleichzeitig zur Stabilität von Kriminalität beitragen? Oder braucht es diese Bestrebungen, um eben dem entgegenzuwirken und das Netzwerk ins Positive zu leiten? Haben demnach Bibliotheken der westlichen Welt, die nicht unter Zensur der eigenen Regierung leiden, auch den weltweiten Auftrag, zur Informationsfreiheit beizutragen und über ihre Landesgrenzen hinauszublicken?

Zu diesen moralischen Überlegungen anknüpfend zählt die Problematik der Schattenbibliotheken. Nicht jedes Land hat Zugang zu freier Information oder die finanziellen Mittel, um die immer teurer werdenden Lizenzen der wissenschaftlichen Verlage zu erwerben und somit Zugang zu aktueller Forschung zu erhalten. In jenen Ländern ist die Nutzung

von Sci-Hub und Library Genesis besonders hoch und omnipräsent in der Medizin und Chemie. Auch hier zeigt sich ein solidarischer Kreislauf: Jene Länder sind auf den Westen angewiesen, der durch die Zugangsdaten der Institutionen zum Betrieb von Sci-Hub beiträgt. Tor kann in dieser Konstellation ebenfalls erwähnt werden, da es wiederum den Zugang zu Wissenschaft durch den Onion Service ermöglicht (im Falle der landesweit gesperrten Domains). Dennoch sind Sci-Hub und Library Genesis nur Symptome eines grösseren Problems, zumal sie ohne die Publikationsarbeit der wissenschaftlichen Verlage ihre Existenzberechtigung im jetzigen Sinne verlieren.

23 der 80 Personen des Mittelbaus (28.75%) gaben an, mehrmals pro Monat Sci-Hub oder LibGen zu benutzen. Wird dies in Kombination mit den uniweiten Zugriffszahlen von 2'092 betrachtet, ergibt sich eine nicht zu ignorierende Bedeutsamkeit der Popularität von Schattenbibliotheken. Insbesondere deshalb, da die Universität Bern ein breites Spektrum an lizenzierten Inhalten anbietet und Sci-Hub und Library Genesis trotzdem hohe monatliche Besuche verzeichnen. Beide Webseiten gewannen Jahr für Jahr mehr Benutzende, während die Problematiken mit den wissenschaftlichen Verlagen stiegen. Umso mehr Gewicht sollte auf die Unterstützung durch die bereits vorhandenen Informationsexpert*innen an den Universitätsbibliotheken gelegt werden, die Möglichkeiten und Methoden der Schöpfung von Ressourcen aus dem Deep Web zu vermitteln.

6 Fazit

Anhand der quantitativen Methode der Befragung dreier verschiedener Gruppen und der Zusammentragung relevanter Literatur verfolgte diese Bachelorarbeit das Ziel, den Stand des Deep Webs im bibliothekarischen Kontext zu ergründen und daraus Aussagen zur Relevanz von Bibliotheken in diesem Zusammenhang zu ziehen. Die Forschungsfrage dabei lautete:

«Wie gross ist die Diskrepanz im Kenntnisstand zwischen Bibliothekar*innen und Studierenden/Forschenden bezüglich des Deep Webs, und inwiefern beeinflusst dies die Aktualität und Relevanz von Bibliotheken?»

Die Resultate ergaben eine hohe grundlegende Kenntnis des Deep Webs bei Bibliothekar*innen schweizweit und Studierenden der FHGR, wobei letztere befragte Gruppe hauptsächlich aus Informationswissenschaftsstudierenden bestand und die Bibliothekar*innen zu hohen Anteilen einen informationswissenschaftlichen Hintergrund aufwiesen oder aus wissenschaftlichen Bibliotheken stammten. Demnach kann die Repräsentativität beider Gruppen hinterfragt werden. Der Mittelbau gab nur zur Hälfte an, über das Deep Web Bescheid zu wissen und es regelmässig zu nutzen – allerdings wurde die Definition des Begriffs im Fragebogen nicht ausführlich genug dargelegt und hätte somit mit hoher Wahrscheinlichkeit zu anderen Ergebnissen zugunsten des Deep Webs geführt.

Die Fragebögen waren darauf ausgelegt, grundlegende Kenntnisfragen zu klären und weniger in die Tiefe zu gehen. Dennoch fehlten einige Punkte, die zu einem vollständigeren Ergebnisbild geführt hätten: Die Fragen zum Deep Web und zu Google (insbesondere Google Scholar) waren zu wenig klar definiert und abgegrenzt. Eine Erklärung, dass Sci-Hub und Library Genesis nicht zum Deep Web gehören, aber darauf zugreifen, hätte vermutlich zum Verständnis beigetragen. Aufgrund der hohen Zugriffszahlen des Dark Webs bei den Studierenden wären eine Kommentarspalte oder Fragen zum Tor-Browser hilfreich gewesen, die Einblicke in die Nutzung geliefert hätten. Ausserdem kann hinterfragt werden, ob die vorgegebenen Optionen zu den Gründen der Bevorzugung von Schattenbibliotheken gegenüber lizenzierten Inhalten eine vollständige Abbildung möglicher Antworten war oder eine freie Texteingabe zu anderen Resultaten geführt hätte. Bezüglich der Statistiken von Sci-Hub und LibGen der Universität Bern wäre ein direkter Vergleich mit den E-Book-Downloads im gleichen Zeitraum sinnvoll gewesen, um genauere Aussagen zur Bedeutung dieser Zahlen zu treffen.

Trotzdem ergab sich eine Diskrepanz nicht im Kenntnisstand seitens der Bibliotheken, jedoch in der Vermittlung jenes Wissens. Bibliotheken sind also Informationszentren und Anlaufstellen, die sich in Zukunft einer sich ändernden Verantwortung stellen könnten. Die Frage nach der Relevanz ist unumstritten, jedoch bedarf es einer stärkeren Positionierung in den Bereichen des Webs. Es fehlt nicht unbedingt am oberflächlichen Fachwissen über das Deep Web selbst, zumindest nicht bei den Informationsspezialisten, sondern an der Kommunikation vom Angebot und Vermittlung des Wissens zur Kundschaft. Es muss genauer untersucht werden, was Benutzende vom Deep Web wirklich brauchen und wie sie es für sich gewinnbringend nutzen können. Dazu müssen weiterführende spezifische Studien durchgeführt werden. Ausserdem sollte das Bibliothekspersonal nicht dazu angewiesen werden, nicht über Schattenbibliotheken zu reden, sondern darüber aufzuklären und gegebenenfalls selbst an Schulungen über das Thema teilzunehmen – schliesslich sind es Bibliotheken, die im Zwiespalt zwischen den Verlagen und der Kundschaft stehen. Das Deep Web steht schlussendlich in direktem Zusammenhang mit den Schattenbibliotheken und im etwas weiteren Sinne mit den Dark-Web-Technologien; Informationen über diese Themen zur Verfügung zu stellen ist der direkteste Weg, nachhaltige Informationskompetenzen bei Benutzenden aufzubauen. Bibliotheken sind nicht nur die Navigatoren der Information, sondern auch diejenigen, die die Tools zur Navigation aushändigen.

Es braucht eine effektivere Ermittlung des Informationsbedarfs der Kundschaft und entsprechende Konzipierungen von Angeboten und Schulungen, die auf diesen Bedarf abgestimmt sind. Ebenso sollte das Bibliothekspersonal in den Thematiken der Schattenbibliotheken und Cyber Security geschult werden, um die rege Nutzung der Kundschaft dieser Phänomene nachzuvollziehen und die Relevanz der Bibliotheken darin zu positionieren. Schulungen in diese Richtung werden im Oktober 2023 durch die Verfasserin an der Universitätsbibliothek Bern durchgeführt.

7 Literaturverzeichnis

- Abbate, J. E. (1994). From ARPANET to Internet. A history of ARPA-sponsored computer networks, 1966-1988. Ph.D. University of Pennsylvania, Pennsylvania. Retrieved from <https://www.proquest.com/dissertations-theses/arpamet-internet-history-arpa-sponsored-computer/docview/304104775/se-2?accountid=17231>
- Alshammari, F. (2017). Towards an Evaluation of a Recommended Tor Browser Configuration in Light of Website Fingerprinting Attacks. Université d'Ottawa / University of Ottawa. <https://doi.org/10.20381/ruor-20806>
- Armstrong, M. (2021). How Many Websites Are There?, Statista. Verfügbar unter: <https://www.statista.com/chart/19058/number-of-websites-online/>
- Banks, M. (2016). Sci-Hub: What It Is and Why It Matters, American Libraries Magazine. Verfügbar unter: <https://americanlibrariesmagazine.org/2016/05/31/why-sci-hub-matters/>
- BBC (2019, 23. Oktober). BBC News launches 'dark web' Tor mirror. BBC News. Verfügbar unter: <https://www.bbc.com/news/technology-50150981>
- Bendezú-Quispe, G., Nieto-Gutiérrez, W., Pacheco-Mendoza, J. & Taype-Rondan, A. (2016). Sci-Hub and medical practice: an ethical dilemma in Peru. *The Lancet. Global Health*, 4(9). [https://doi.org/10.1016/S2214-109X\(16\)30188-7](https://doi.org/10.1016/S2214-109X(16)30188-7)
- Bergman, M. K. (2000). *The Deep Web: Surfacing Hidden Value* (Smither, M., Hrsg.). BrightPlanet.com LLC. <https://doi.org/10.3998/3336451.0007.104>
- Bodó, B. (2018a). The Genesis of Library Genesis. The Birth of a Global Scholarly Shadow Library. In J. Karaganis (Ed.), *Shadow Libraries. Access to knowledge in global higher education* (S. 25–51). The MIT Press.
- Bodó, B. (2018b). Library Genesis in Numbers. Mapping the Underground Flow of Knowledge. In J. Karaganis (Ed.), *Shadow Libraries. Access to knowledge in global higher education* (S. 53–78). The MIT Press.
- Brooks, D. (2015). Lebanon library at center of internet privacy debate in shutting off its Tor server. *Concord Monitor*. Verfügbar unter: <https://www.concordmonitor.com/Archive/2015/09/LibraryTor-cm-091115>
- Çalışkan, E., Minárik, T. & Osula, A.-M. (2015). Technical and legal overview of the Tor Anonymity Network. Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence. Verfügbar unter:

http://goodtimesweb.org/surveillance/2015/tor_anonymity_network.pdf

Chertoff, M. (2017). A public policy perspective of the Dark Web. *Journal of Cyber Policy*, 2(1), 26–38. <https://doi.org/10.1080/23738871.2017.1298643>

Coffey, M. L. (2020). Library application of Deep Web and Dark Web technologies. *School of Information Student Research Journal*, 10(1). <https://doi.org/10.31979/2575-2499.100108>

Cohen-Almagor, R. (2011). Internet History. *International Journal of Technoethics*, 2(2), 45–64. <https://doi.org/10.4018/978-1-4666-2931-8.ch002>

Devine, J. & Egger-Sider, F. (2014). Going beyond Google again. Strategies for using and teaching the invisible web. Chicago: Neal-Schuman. Retrieved from <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=800350>

Dougherty, S. (2023). Freenet renamed to Hyphanet, Hyphanet. Verfügbar unter: <https://www.hyphanet.org/freenet-renamed-to-hyphanet.html>

Elbakyan, A. (2017). Some facts on Sci-Hub that Wikipedia gets wrong. Verfügbar unter: <https://engineering.wordpress.com/2017/07/02/some-facts-on-sci-hub-that-wikipedia-gets-wrong/>

District Court, S.D. New York (21.07.2017) 1:15-cv-04282. Retrieved from <https://www.courtlistener.com/docket/4355308/elsevier-inc-v-sci-hub/>

Esteve, A. (2017). The business of personal data. Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law*, 7(1), 36–47. <https://doi.org/10.1093/idpl/ipw026>

Facebook. (2023). Facebook over Tor. Verfügbar unter: <https://www.facebook.com/facebookcorewwi/>

Fachhochschule Graubünden. (2020, 30. April). Modulübersicht. Bachelorstudium Information Science. Verfügbar unter: https://www.fhgr.ch/fileadmin/studium/bachelorangebot/information_science/Bachelorstudium-Information_Science-Moduluebersicht.pdf

Fernández-Marcial, V., Costa, L. & González-Solar, L. Top Universities, Top Libraries. Do Research Services in Academic Libraries Contribute to University Output? In *Proceedings of the IATUL Conferences* (S. 1–9). Verfügbar unter: <https://docs.lib.purdue.edu/iatul/2016/spaces/2>

- Finklea, K. (2017). Dark Web (Congressional Research Service, Hrsg.) (CRS Report R44101). Verfügbar unter: [https://a51.nl/sites/default/files/pdf/r44101%20\(1\).pdf](https://a51.nl/sites/default/files/pdf/r44101%20(1).pdf)
- Ford, N. & Mansourian, Y. (2006). The invisible web: an empirical study of “cognitive invisibility”. *Journal of Documentation*, 62(5), 584–596. <https://doi.org/10.1108/00220410610688732>
- Gardner, G. J., McLaughlin, S. R. [Stephen R.] & Asher, A. D. (2017). Shadow libraries and you: Sci-Hub usage and the future of ILL. ACRL. Verfügbar unter: <https://alair.ala.org/bitstream/handle/11213/17772/shadowlibrariesandyou.pdf?sequence=1>
- Giampietro, M. (2013). Twenty years of a free, open web, CERN. Verfügbar unter: <https://home.cern/news/news/computing/twenty-years-free-open-web>
- González-Solar, L. & Fernández-Marcial, V. (2019). Sci-Hub, a challenge for academic and research libraries. *El Profesional de la Información*, 28(1). <https://doi.org/10.3145/epi.2019.ene.12>
- Google (Autor). (2020). Trillions of Questions, No Easy Answers. A (home) movie about how Google Search works.
- Google. (2023). Umfassender Leitfaden zur Funktionsweise der Google Suche. Verfügbar unter: <https://developers.google.com/search/docs/fundamentals/how-search-works?hl=de>
- Greshake, B. (2017). Looking into Pandora's Box: The Content of Sci-Hub and its Usage. *F1000Research*, 6(541), 1–17. <https://doi.org/10.12688/f1000research.11366.1>
- Hatta, M. (2020). Deep web, dark web, dark net. A taxonomy of "hidden" internet. *Annals of Business Administrative Science*, 19(6), 277–292. <https://doi.org/10.7880/abas.0200908a>
- Hawkins, B. (2016). Under the ocean of the Internet. The Deep Web. GIAC Gold Certification. SANS Institute. Verfügbar unter: <https://www.sans.org/white-papers/37012/>
- Himmelstein, D. S., Romero, A. R., Levernier, J. G., Munro, T. A., McLaughlin, S. R. [Stephen Reid], Greshake Tzovaras, B. et al. (2018). Sci-Hub provides access to nearly all scholarly literature. *ELife*, 7. <https://doi.org/10.7554/eLife.32822>
- Hoang, N. P., Kintis, P., Antonakakis, M. & Polychronakis, M. (2018). An Empirical Study of the I2P Anonymity Network and its Censorship Resistance. In *Proceedings of the Internet Measurement Conference 2018 (ACM Conferences, S. 379–392)*. New York, NY: ACM.

- Hoy, M. B. (2017). Sci-Hub: What librarians should know and do about article piracy. *Medical Reference Services Quarterly*, 36(1), 73–78. <https://doi.org/10.1080/02763869.2017.1259918>
- I2P. (2023a). I2P Compared to Freenet. Verfügbar unter: <https://geti2p.net/en/comparison/freenet>
- I2P. (2023b). I2P Compared to Tor. Verfügbar unter: <https://geti2p.net/en/comparison/tor>
- I2P. (2023c). The Invisible Internet Project (I2P). About. Verfügbar unter: <https://geti2p.net/en/about/intro>
- International Telecommunication Union. (2022). Internet use. Facts and Figures 2022. Verfügbar unter: <https://www.itu.int/itu-d/reports/statistics/2022/11/24/ff22-internet-use/>
- Karaganis, J. (Ed.). (2018). *Shadow Libraries. Access to knowledge in global higher education: The MIT Press*. Retrieved from <https://library.open.org/handle/20.500.12657/26038>
- Kavallieros, D., Myttas, D., Kermitsis, E., Lissaris, E., Giataganas, G. & Darra, E. (2021). Understanding the Dark Web. In B. Akhgar, M. Gercke, S. Vrochidis & H. Gibson (Eds.), *Dark Web Investigation (Security informatics and law enforcement, 1st ed., S. 3–26)*. Cham: Springer. https://doi.org/10.1007/978-3-030-55343-2_1
- Kumar, H., Akash, Arun & Dixit, P. (2022). Achieving Anonymity with the Help of TOR. TOR: A Review. *International Journal of Innovative Science and Research Technology*, 7(11), 2214–2220. Verfügbar unter: <https://ijisrt.com/assets/upload/files/ijisrt22nov1595.pdf>
- Lacson, W. & Jones, B. (2016). The 21st Century Darknet Market: Lessons From The Fall Of Silk Road. *International Journal of Cyber Criminology*, 10(1), 40–61. <https://doi.org/10.5281/zenodo.58521>
- Lawson, S. (2017). Access, ethics and piracy. *Insights. The UKSG journal*, 30(1), 25–30. <https://doi.org/10.1629/uksg.333>
- Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C. et al. (2009). A brief history of the internet. *ACM SIGCOMM Computer Communication Re-view*, 39(5), 22–31. <https://doi.org/10.1145/1629607.1629613>
- Leyden, J. (2019, 29. Oktober). ARPANET anniversary: The internet's first transmission was sent 50 years ago today. *The Daily Swig*. Verfügbar unter:

<https://portswigger.net/daily-swig/arpanet-anniversary-the-internets-first-transmission-was-sent-50-years-ago-today>

Library Freedom Project. (2023). What does Library Freedom Project do? Verfügbar unter: <https://libraryfreedom.org/>

Lund, B. (2021). The Brave browser: a monetary opportunity for libraries in the cryptoverse. *Library Hi Tech News*, 38(6), 15–16. <https://doi.org/10.1108/LHTN-05-2021-0023>

Lund, B. & Beckstrom, M. (2021). The Integration of Tor into Library Services: An Appeal to the Core Mission and Values of Libraries. *Public Library Quarterly*, 40(1), 60–76. <https://doi.org/10.1080/01616846.2019.1696078>

Lyman, P. & Varian, H. R. (2003). How much information? School of Information Management and Systems, University of California. Verfügbar unter: <https://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>

Macrina, A. & Phetteplace, E. (2015). The Tor browser and intellectual freedom in the Digital Age. *Reference & User Services Quarterly*, 54(4), 17–20. Verfügbar unter: <https://www.jstor.org/stable/10.2307/refuseserq.54.4.17>

MDN Web Docs. (2023, 13. Juni). How does the Internet work?, Mozilla Foundation. Verfügbar unter: https://developer.mozilla.org/en-US/docs/Learn/Common_questions/Web_mechanics/How_does_the_Internet_work

Negi, N. (2017). Comparison of anonymous communication networks. Tor, I2P, Freenet. *International Research Journal of Engineering and Technology (IRJET)*, 4(7), 2542–2544.

Ngo, F. T., Marcum, C. & Belshaw, S. (2023). The Dark Web: What Is It, How to Access It, and Why We Need to Study It. *Journal of Contemporary Criminal Justice*, 39(2), 160–166. <https://doi.org/10.1177/10439862231159774>

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B. et al. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, 1-23. <https://doi.org/10.7717/peerj.4375>

Prasad, M. (2017). Deep Web: Librarian's Perspective. *Pearl: A Journal of Library and Information Science*, 11(4), 418–423. <https://doi.org/10.5958/0975-6922.2017.00054.7>

Ruiz-Martínez, A. (2012). A survey on solutions and main free tools for privacy enhancing

- Web communications. *Journal of Network and Computer Applications*, 35(5), 1473–1492. <https://doi.org/10.1016/j.jnca.2012.02.011>
- Sandvik, R. (2017, 27. Oktober). The New York Times is Now Available as a Tor Onion Service. *NYT Open*. Verfügbar unter: <https://open.nytimes.com/https-open-nytimes-com-the-new-york-times-as-a-tor-onion-service-e0d0b67b7482>
- Schollmeier, R. (2002). A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In R. L. Graham & N. Shahmehri (eds.), *Proceedings / First International Conference on Peer-to-Peer Computing*. 27 - 29 August 2001, Linköping, Sweden (S. 101–102). Los Alamitos, Calif.: IEEE Computer Soc.
- Sci-Hub. (2023a). About. Verfügbar unter: <https://www.sci-hub.se/about>
- Sci-Hub. (2023b). Datenbank. Verfügbar unter: <https://www.sci-hub.se/>
- Sci-Hub. (2023c). Statistics. Verfügbar unter: <https://www.sci-hub.se/stats>
- Steinebach, M., Schäfer, M., Karakuz, A., Brandl, K. & Yannikos, Y. (2019). Detection and Analysis of Tor Onion Services. In *Proceedings of the 14th International Conference on Availability, Reliability and Security (ACM Digital Library)*, S. 1–10. New York: Association for Computing Machinery.
- The Tor Project. (2015). Tor Exit Nodes in Libraries. Pilot (phase one). Verfügbar unter: <https://blog.torproject.org/tor-exit-nodes-libraries-pilot-phase-one/>
- The Tor Project, Inc. (2023a). About Tor Browser. Verfügbar unter: <https://tb-manual.torproject.org/about/>
- The Tor Project, Inc. (2023b). History. Verfügbar unter: <https://www.torproject.org/about/history/>
- The Tor Project, Inc. (2023c). Onion Services. Verfügbar unter: <https://tb-manual.torproject.org/onion-services/>
- The Tor Project, Inc. (2023d). Tor Metrics. Onion Services. Verfügbar unter: <https://metrics.torproject.org/hidserv-dir-v3-onions-seen.html>
- The Tor Project, Inc. (2023e). Tor Metrics. Users. Verfügbar unter: <https://metrics.torproject.org/userstats-relay-country.html>
- Weimann, G. (2016). Going Dark: Terrorism on the Dark Web. *Studies in Conflict & Terrorism*, 39(3), 195–206. <https://doi.org/10.1080/1057610X.2015.1119546>

Wood, J. (2010). The Darknet: A Digital Copyright Revolution. *Richmond Journal of Law & Technology*, 16(4), 14. Verfügbar unter: <https://scholarship.richmond.edu/jolt/vol16/iss4/4>

Wouters, P., Reddy, C. & Aguillo, I. (2006). On the visibility of information on the Web. an exploratory experimental approach. *Research Evaluation*, 15(2), 107–115. <https://doi.org/10.3152/147154406781775931>

Zegarac, N. (13.07.2023a). ID0115279 - Zugriffsstatistiken (E-Mail).

Zegarac, N. (13.07.2023b). Zugriffsstatistiken Mirror-Seiten (Chat).

Bisher erschienene Schriften

Ergebnisse von Forschungsprojekten erscheinen jeweils in Form von Arbeitsberichten in Reihen.
Sonstige Publikationen erscheinen in Form von alleinstehenden Schriften.

Derzeit gibt es in den Churer Schriften zur Informationswissenschaft folgende Reihen:
Reihe Berufsmarktforschung

Weitere Publikationen

Churer Schriften zur Informationswissenschaft – Schrift 152

Herausgegeben von Wolfgang Semar

Lisa Heller

Zur Genese eines nationalen Bibliotheksprojekts: Swiss Library Service Platform (SLSP)

Chur, 2022

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 153

Herausgegeben von Wolfgang Semar

Antonin Friberg

Die Effektivität von Social Norms Nudging in der Customer Journey

Chur, 2022

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 154

Herausgegeben von Wolfgang Semar

Curdin Marxer

«Drug Repurposing»

Wie können unstrukturierte Textdaten für die Ermittlung neuer «Drug Repurposing» Kandidaten nutzbar gemacht werden und wie können sie Datenbanken ergänzen?

Chur, 2022

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 155

Herausgegeben von Wolfgang Semar

Samir Limani

Sicht der administrativen Mitarbeitenden von Bündner Spitälern und Kliniken auf den

Digitalisierungsstand ihres Unternehmens

Chur, 2022

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 156

Herausgegeben von Wolfgang Semar

Marina Lea Schürmann

Deep Learning für Part-of-Speech-Tagging

Vergleich eines auf Transformers basierenden POS-Taggers mit bestehenden Modellen

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 157

Herausgegeben von Wolfgang Semar

Yannick Mireille Kaufmann

Einsatz von Unternehmenswikis als Wissensmanagement-Tool in einer Netzwerkorganisation

Evaluationsstudie zu «wikimia», eine Wissensdatenbank in der schweizerischen Berufs-,

Studien- und Laufbahnberatung Masterthesis 2022

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 158

Herausgegeben von Wolfgang Semar

Franco Malacrida

Standortfindung von Schweizer Start-ups

Welche Standortfaktoren sind für Schweizer Start-ups am wichtigsten?

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 159
Herausgegeben von Wolfgang Semar
Josip Spec
From ISAD(G) to Records in Contexts – A new era
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 160
Herausgegeben von Wolfgang Semar
Loris Haller
Gemeinwohl fördern als Geschäftsmodell
Kriterien für die Entwicklung eines Frameworks für gemeinwohlorientierte Geschäftsmodelle
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 161
Herausgegeben von Wolfgang Semar
Céline Graf
«Ghostbusters Münstergasse»
Vermittlung von regionalen Onlinere Ressourcen und Recherchekompetenzen mit einem digitalen
Educational Escape Room an der Bibliothek Münstergasse der Universitätsbibliothek Bern
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 162
Herausgegeben von Wolfgang Semar
Mahmoud Hemila
Qualitätsanalyse von inhaltsbasierten Empfehlungssystemen für Journals
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 163
Herausgegeben von Wolfgang Semar
Nicolas Brauchli
Inwiefern unterscheiden sich die Online-Plattformen der Legacy-Medien von den Digital Born
Plattformen in der Deutschschweizer Medienlandschaft?
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 164
Herausgegeben von Wolfgang Semar
Flurin Böni
Das verborgene Gold am Ende des Rainbow-Washing
Eine Analyse der Vereinbarkeit sozialen Engagements mit unternehmerischen Zielen
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 165
Herausgegeben von Wolfgang Semar
Alina Viert
Herausforderungen in der Aufbewahrung von Videospiele n und ihrer Peripherie
Fragen und Antworten insbesondere zur Peripherie und zur Emulation als Lösungsansatz
Chur, 2023
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 166
Herausgegeben von Wolfgang Semar
Susanne Knöpfel
Wissenslandkarten als Grundlage für Visualisierungen im Wissensmanagement
Chur, 2023
ISSN 1660-945X

Über die Informationswissenschaft der Fachhochschule Graubünden

Die Informationswissenschaft ist in der Schweiz noch ein relativ junger Lehr- und Forschungsbereich. International weist diese Disziplin aber vor allem im anglo-amerikanischen Bereich eine jahrzehntelange Tradition auf. Die klassischen Bezeichnungen dort sind Information Science, Library Science oder Information Studies. Die Grundfragestellung der Informationswissenschaft liegt in der Betrachtung der Rolle und des Umgangs mit Information in allen ihren Ausprägungen und Medien sowohl in Wirtschaft und Gesellschaft. Die Informationswissenschaft wird in Chur integriert betrachtet.

Diese Sicht umfasst nicht nur die Teildisziplinen Bibliothekswissenschaft, Archivwissenschaft und Dokumentationswissenschaft. Auch neue Entwicklungen im Bereich Medienwirtschaft, Informations- und Wissensmanagement und Big Data werden gezielt aufgegriffen und im Lehr- und Forschungsprogramm berücksichtigt.

Der Studiengang Informationswissenschaft wird seit 1998 als Vollzeitstudiengang in Chur angeboten und seit 2002 als Teilzeit-Studiengang in Zürich. Seit 2010 rundet der Master of Science in Business Administration das Lehrangebot ab.

Der Arbeitsbereich Informationswissenschaft vereinigt Cluster von Forschungs-, Entwicklungs- und Dienstleistungspotenzialen in unterschiedlichen Kompetenzzentren:

- Information Management & Competitive Intelligence
- Collaborative Knowledge Management
- Information and Data Management
- Records Management
- Library Consulting
- Information Laboratory
- Digital Education

Diese Kompetenzzentren werden im Swiss Institute for Information Science (SII) zusammengefasst.

Impressum

Impressum

FHGR - Fachhochschule
Graubünden
Information Science
Pulvermühlestrasse 57
CH-7000 Chur

www.informationsscience.ch

www.fhgr.ch

ISSN 1660-945X

Institutsleitung

Prof. Dr. Ingo Barkow
Telefon: +41 81 286 24 61
Email: ingo.barkow@fhgr.ch

Sekretariat

Telefon: +41 81 286 24 24
Fax: +41 81 286 24 00
Email: clarita.decurtins@fhgr.ch