

Satellite Meeting: Serials and Other Continuing Resources Section &
Acquisition and Collection Development

[Open Access: Action Required](#)

Date: 16-17 August 2017

Location: European Solidarity Center (ESC), Gdańsk (Poland)

Sci-Hub and LibGen: what if... why not?

Louis Houle

McGill Library, McGill University, Montreal (Quebec), Canada.

E-mail address: louis.houle@mcgill.ca



Copyright © 2017 by Louis Houle. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

Open Access is becoming the predominant form of getting access to peer reviewed articles. Many new non-traditional tools (institutional repositories, social media, and peer to peer sites) are available out there to retrieve the full-text of peer reviewed articles.

Since the launch of Sci-Hub (2011) and Library Genesis (LibGen), several criticisms have been raised both from the library and publishing worlds. Some court decisions have also been rendered to block the access of such platforms in different countries. What can be said in terms of content, coverage, reliability, stability and currency of both Sci-Hub and LibGen?

In this era of severe budget constraints and economic recessions that libraries are facing, can we imagine of substituting some or most of our journal collection funds with the “open and free access” that Sci-Hub and LibGen is giving us? How much overlap between our collections and what is available through Sci-Hub and LibGen?

This article reports on preliminary results¹ of a one year study with Sci-Hub, LibGen and of Google Scholar (GS) where 2,750 random samples (peer reviewed journal articles) coming from fifty-five different databases covering all disciplines (Arts & Humanities, Law, Music, Social Sciences and STM) were tested against those platforms. The samples have been searched on all three platforms at four different intervals during the year in order to evaluate the stability of content.

Different data such as publication year, publishers, language of articles and OA are being looked at to see if content is affected by either or all of these parameters.

To verify the currency of information in Sci-Hub, LibGen and Google Scholar, research articles from both Nature and Science (from current issues, Nature Advance Online Publication and First Release from Science) were queried on a daily basis.

Results are showing overall retrieval rates of 70% in Sci-Hub and 69% in LibGen across all disciplines. Most of Nature and Science new research articles were found within the same day or 24 hours from their first release on their respective web sites. This is also true within Google but strangely enough, a delay of between 4-5 days is seen in Google Scholar.

One conclusion that could be drawn from the present results is that the content is there in Sci-Hub and LibGen. Should we continue to ignore them totally?

Keywords: Open Access, Google Scholar, Google, Sci-Hub, LibGen

Introduction

Discoverability is one important aspect of any information seeking needs but the retrieving of the full-text content is as much important and crucial for any researchers. Faculty and students are relying on their own institutions to give them access to scholarly articles via their own subscriptions where they need to pay large amount of money to many publishers. No one institution alone is able to subscribe to every journals that are being published in this twenty-first century. This is one of the reason as to why the Open Access movement became so important. Faculty and students are gaining access to more and more journal articles via Open Access. There has been many new non-traditional tools (institutional repositories, social media – e.g. Academia/Mendeley/ResearchGATE, peer to peer sites) created out there to retrieve the full-text of peer reviewed articles. Sci-Hub and LibGen are two of them which have drawn a lot of attention in the last few years. Bohannon (2016) reported that there are millions of articles downloaded every month from the Sci-Hub platform from all over the world.

To this day, only a few studies have looked at Sci-Hub (Gardner et al., 2017; Greshake, 2016, 2017; Himmelstein et al., 2017; Machin-Mastromatteo et al., 2016; Timus and Babutsidze, 2016) or LibGen (Cabanac, 2016) either for their content or usage patterns.

Most of these recent studies on Sci-Hub have been using a dataset of 28 million download requests between September 2015 and February 2016 which has been added to the Dryad Digital Repository - <http://datadryad.org/resource/doi:10.5061/dryad.q447c> (Elbakyan and Bohannon, 2016). Usage patterns were analyzed for different parts of the globe: Latin America (Machin-Mastromatteo et al., 2016), European countries (Timus and Babutsidze, 2016), United States (Gardner et al., 2017) or through the world in general along with different indicators (Greshake, 2016, 2017).

The Science survey done in 2016 with 11,000 researchers (Travis, 2016) revealed that 60% of the respondents have used Sci-Hub and 88% said that there was no wrongdoing in getting pirated papers. The survey also gave the following reasons as to why one would use Sci-Hub: the lack of journal access was the primary reason given by 50% of the respondents, 17% for convenience and another 23% “because they objected to the profits publishers make”. About 37% who retrieved an article through Sci-Hub did it even though they had traditional forms of access.

During the last few years some publishers started to take some actions against Sci-Hub and LibGen. In October 2015, Elsevier received a court victory against Sci-Hub regarding its violation of the United States copyright law. More recently, on June 21, 2017, Elsevier won a suit of \$15 million in damages in the US court. On June 23, 2017, the American Chemical

Society (ACS) decided to file a suit against Sci-Hub in the District Court Eastern District of Virginia.

The main goal of this study is not to look at the legality or illegality of using Sci-Hub and LibGen but to evaluate the content availability on both platforms. Random peer review article samples have been selected from fifty-five databases covering all disciplines and queried against both Sci-Hub and LibGen over a one year period. This will allow to answer questions such as: how much content is there, is content stable over time and how current is the information?

Methodology

A total of fifty-five databases, representing all disciplines (Agriculture/Environment, Arts, Business/Management, Education, Engineering, Health/Biological Sciences, Humanities, Law, Multidisciplinary, Music, Science and Social Sciences) were surveyed throughout the year. These categories were taken from the many subject guides created by the different liaison librarians at McGill University. Five of those databases were full-text platforms from the top major journal title publishers: Elsevier, Sage, Springer, Taylor & Francis (T&F) and Wiley. Fifty randomly selected samples were retrieved from each database. In order to generate random numbers from this pool of databases, the free internet Research Randomizer (<https://www.randomizer.org/>) tool was used.

The database searches were performed by limiting the results to only journal articles or peer review articles whenever this was possible.

Once all of the 2,750 random samples were retrieved from the databases, they were individually queried in Sci-Hub and LibGen. The following steps were used depending on the platform:

- a) In Sci-Hub:
 1. Article title as a phrase
 2. If step one failed then by DOI or by the PubMed Central ID number (PMCID) whenever available
- b) In LibGen:
 1. Article title as a phrase
 2. If step one failed then by DOI or by the PubMed Central ID number (PMCID) whenever available
 3. If step two failed then by either the journal title and/or its ISSN

If a sample contained a non-English title then both the original title language and the English version, whenever available from the database, were searched in Sci-Hub and LibGen. All 2,750 samples were queried quarterly between September 2016 and July 2017: October 2016, January 2017, April 2017 and July 2017.

Methodology for the Delay of Coverage of Information in Sci-Hub, LibGen, Google and GS

In order to measure the delay of coverage or the currency of the information in Sci-Hub, LibGen and GS, articles from two of the most prestigious journals were queried on a daily basis. From September 2016 to June 2017, research articles from both Nature and Science were monitored to see how fast these titles were uploaded in those platforms. Here is a summary of what articles have been monitored throughout this study:

Nature

- Research articles coming out from new weekly issues (published on Thursdays)
- Research articles coming out as Advance Online Publication, known as Nature AOP (published on different weekdays but mainly Mondays and Wednesdays)

Science

- Research articles coming out from new weekly issues (published on Fridays)
- Research articles coming out as Science First Release papers (published generally on Thursdays but some are coming out on different weekdays)

To see if there are any discrepancies of the scholarly content between Google and GS, the samples were also searched against Google. The following steps were performed on a daily basis until the articles were found:

1. Articles were searched using their DOI provided by both Nature and Science
2. If step one failed then articles were searched using the title as a phrase search (in between quotes within GS)
3. In GS – if step 2 failed then clicking on the link “Try your query on the entire web” where it searched against Google

Samples Overview

Here is a summary of the 2,750 random research article samples retrieved from the fifty-five selected databases:

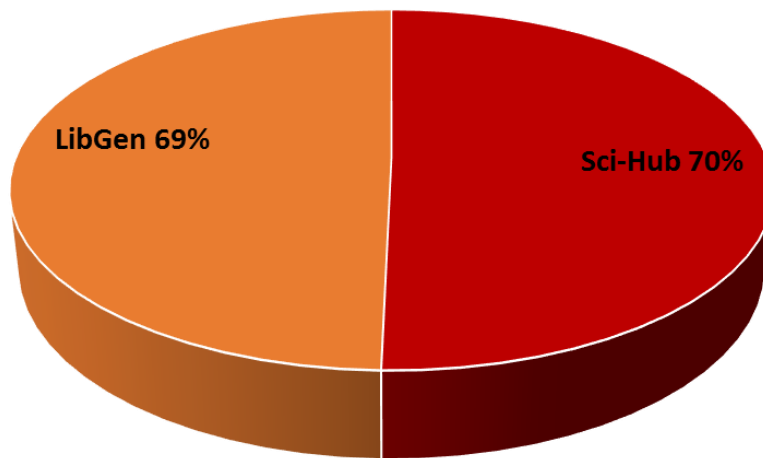
- **Publication Years**
 - 86 different years represented ranging from 1847 to 2017
 - 16 decades represented with some complete ones ('60s, '70s, '80s, '90s, '00s and 2010-2017)
 - Percentage distribution by decades:
 - 40% from the 2010s
 - 30% from the '00s
 - 16% from the '90s
 - 11% from the '80s
 - 4% from the '70s
 - 2% from the '60s
 - 1% from the '50s
 - Nine more decades with less than 1%
 - Percentage distribution of top 10 years:
 - 6% from 2012, 2014 and 2015
 - 5% from 2010, 2011 and 2013
 - 4% from 2005, 2007, 2008 and 2009
 - 3% from 2002, 2004, 2006 and 2016
- **Publishers**
 - 2,750 samples coming from 987 different publishers
 - Rank of Top 10 publishers by percentage of samples, representing just over 50% of all the samples:
 - Taylor & Francis – 12%
 - Elsevier – 10%
 - Wiley – 9%
 - Springer – 7%
 - Sage – 6%

- Oxford, Cambridge – 2%
 - MCB Emerald, IEEE, BMC and Brill – 1%
- Languages
 - Spread over 26 languages with English representing 91% of total samples followed by both German and French with 1.8% and Spanish with 1.6%. The remaining twenty-two languages having less than 1%. A total of 236 non-English samples (9%) are part of this study.
- Journals
 - Samples distributed over 2,572 different journal titles
 - 2,415 journals supplying each only 1 article
 - 141 journals supplying each 2 articles
 - 12 journals supplying each 3 articles
 - 3 journals supplying each 4 articles
 - 1 journal supplying 5 articles
 -
 - 11% of samples (292) are Open Access articles

Results

As it can be seen from Figure 1, the overall retrieval rates for the 2,750 samples, as of July 2017, is pretty good. The full-text retrieval rates for both Sci-Hub and LibGen are respectively 70% and 69%. This is not what one would have probably expected considering the breadth of the disciplines covered in this study. Himmelstein et al. (2017) reported that Sci-Hub's repository contains 69% of the 81.6 million scholarly articles with DOIs. The majority of samples discovered in Sci-Hub and LibGen were mainly retrieved by article title (88% and 89% respectively) with a smaller proportion through DOI (12% for Sci-Hub and 11% for LibGen). Some samples were retrieved by the PubMed Central reference number (PMCID) in both Sci-Hub (0.3%) and LibGen (0.2%).

Figure 1. Overall Retrieval Rates for Sci-Hub and LibGen samples in April 2017



The distribution rates over all fifty-five databases ranges from 20% to 100% for both Sci-Hub and LibGen. At the discipline level, the results are showing the lowest retrieval rates in Sci-Hub for Law (20%), Music (28%) and Business/Management (32%). As for LibGen, the lowest retrieval rates are with Law (20%), Music (28%) and Business/Management (30%). On the other hand, the highest retrieval rates (100%) in Sci-Hub and LibGen are coming from the Humanities and the Multidisciplinary disciplines. Himmelstein et al. (2017) found that Chemistry had the highest coverage at 92.8% while Computer Science the lowest with 76.3%.

Looking down at the database level, see Table 1, there are some disparities between databases and interestingly enough also between databases from within their own disciplines. For both Sci-Hub and LibGen, the lowest retrieval rates are coming from the Index to Legal Periodicals & Books Full Text (20%), RILM Abstracts of Music Literature (28%), the Index to Foreign Legal Periodicals (30%) and the Canadian Business & Current Affairs (32% for Sci-Hub and 30% for LibGen) databases.

At the other end of the spectrum, for both Sci-Hub and LibGen, five databases (9%) are showing a 100% retrieval rate (British Humanities, Elsevier ScienceDirect, Sage, Springer and Wiley) coming from the following disciplines – Humanities and Multidisciplinary. Only six databases resulted with retrieval rates lower than 50% for both Sci-Hub (20% - 44%) and LibGen (20% - 42%).

Table 1. Overall retrieval rates (%) by database for Sci-Hub and LibGen in April 2017

Database	Discipline	Platform	Sci-Hub	LibGen
ABI Inform Global 1905- Academic Search Complete	Business/Management	ProQuest	86	86
Agricola 1970 - Applied Science & Technology Full Text 1983 -	Multidisciplinary	EBSCO	76	76
Art Full Text 1984 - ATLA Religion Database with ATLASerials 1908- BIOSIS Previews + Archive 1926- British Humanities (BHI) Index 1962 - Business Source Complete	Agriculture/Environment	Ovid	78	78
CAB Abstracts + Archive 1910 - Canadian Business & Current Affairs Database (CBCA) 1971- CINAHL Plus with Full Text 1937 - Communication Abstracts 1972 - Compendex 1884 - Econlit 1886- Education Full text (Wilson/Ebsco)	Science	EBSCO/Wilson	84	82
Embse Classic + Embase 1947- ERIC 1966 - (also via United States Dept. Education open access) Expanded Academic ASAP	Arts	EBSCO/Wilson	62	62
Francis 1972- GeoRef 1693 - Historical Abstracts 1955 -	Humanities	EBSCO	44	42
	Health/Biological Sciences	Ovid	78	78
	Humanities	ProQuest	100	100
	Business/Management	EBSCO	72	72
	Agriculture/Environment	Ovid	64	64
	Business/Management	ProQuest	32	30
	Health/Biological Sciences	EBSCO	84	80
	Humanities	EBSCO	78	78
	Engineering	EI Village	74	74
	Business/Management	EBSCO	68	66
	Education	EBSCO/Wilson	56	56
	Health/Biological Sciences	Ovid	56	54
	Education	EBSCO	70	70
	Multidisciplinary Humanities/Social Sciences (foreign language)	Gale	86	84
	Science	EBSCO	64	64
	Humanities	ProQuest	70	70
	Humanities	EBSCO	52	52

Index to Foreign Legal Periodicals 1985- Index to Legal Periodicals & Books Full Text	Law (foreign language)	Hein Online	30	30
Inspec	Law	EBSCO/Wilson	20	20
Library & Information Science Abstracts (LISA) 1969 - Library Literature & Information Science Full Text 1980- Linguistics and Language Behavior Abstracts (LLBA) 1973 - MathSciNet 1800- Medline 1946 - Meteorological & Geostrophysical Abstracts 1974 - MLA International Bibliography 1962- Music Periodicals Database 1874 - OmniFile Full Text Mega PAIS Index 1914- Periodicals Archive Online Philosopher's Index 1940 - ProQuest Research Library PsycINFO 1806 - PubMed 1946 - RILM Abstracts of Music Literature 1800- Sage ScienceDirect SciFinder Scholar 1907 - Scopus Social Services Abstracts 1979 - Social Work Abstracts 1968 - SocINDEX with Full	Engineering/Science	EI Village	78	78
	Education	ProQuest	62	60
	Education	EBSCO/Wilson	44	42
	Social Sciences	ProQuest	72	72
	Science	AMS	54	54
	Health/Biological Sciences	Ovid	56	56
	Science	ProQuest	90	90
	Humanities	ProQuest	56	56
	Music	ProQuest	54	54
	Multidisciplinary	EBSCO/Wilson	64	64
	Humanities	ProQuest	88	88
	Humanities/social Sciences	ProQuest	54	54
	Humanities	ProQuest	86	86
	Multidisciplinary	ProQuest	70	70
	Social Sciences	Ovid	78	78
	Health/Biological Sciences	NIH	60	50
	Music	EBSCO	28	28
	Multidisciplinary	Sage	100	100
	Multidisciplinary	Elsevier	100	100
	Science	ACS	84	84
	Multidisciplinary	Elsevier	62	62
	Social Sciences	ProQuest	92	92
	Social Sciences	Ovid	72	72
	Social Sciences	EBSCO	64	64

Text 1908 - Sociological Abstracts				
1952 -	Social Sciences	ProQuest	92	92
SPORTDiscus	Multidisciplinary	EBSCO	60	60
Springer	Multidisciplinary	Springer- Nature	100	100
Taylor & Francis	Multidisciplinary	T&F Thomson	96	96
Web of Science	Multidisciplinary	Reuters	80	80
Wiley	Multidisciplinary	Wiley	100	100
Zoological Record 1864 -	Health/Biological Sciences	Ovid	52	52

Table 2 is listing the distribution of retrieval rates by decades for Sci-Hub and LibGen. Looking at the decades where there are more than 100 samples (1970s, 1980s, 1990s, 2000s and 2010s), there is a steady increase (except for the 1980s where there is a slight decrease) in the retrieval rates in both Sci-Hub and LibGen: ranging from 65% (1970s) to 76% (2010s) for Sci-Hub and from 63% (1970s) to 75% (2010s). Although the number of samples is small, Sci-Hub and LibGen are also successful with older articles except for three samples published in 1847, 1869 and 1886 where the retrieval was not successful. Of course there are too few very old samples to draw any conclusions but one can confirm that there are older peer review articles covered in Sci-Hub and LibGen. Cabanac (2016) found that the oldest article available from the Philosophical Transactions of the Royal Society of London published in 1665 (“Epistle Dedicatory” by Oldenburg, H.) can be retrieved by LibGen. Greshake (2017) found a publication in Sci-Hub dating back to 1619 published by Elsevier (by W. Snell in *Descriptio Cometae*). After verification, these two items can be retrieved in both Sci-Hub and LibGen.

For the publication year with 3% or more of the samples (see section on Samples Overview for the list), Sci-Hub retrieval rates are ranging from 60% (2009) to 89% (2016) and for LibGen from 60% (2009) to 87% (2016).

Himmelstein et al. (2017) reported that since 1850, the annual coverage is between 60-80% for most of the years. They also found a drop in the coverage starting in 2010: e.g. 56% for 2016 and 45.3% for 2017.

Table 2. Overall distribution of retrieval rates by decades for Sci-Hub and LibGen

Decades	Total # samples	# Sci-Hub	Sci-Hub % of total	# LibGen	LibGen % of total
1840s	1	0	0	0	0
1860s	1	0	0	0	0
1880s	1	0	0	0	0
1890s	1	1	100	1	100
1900s	3	2	67	2	67
1910s	1	1	100	1	100
1920s	3	3	100	3	100
1930s	8	6	75	6	75
1940s	10	3	30	3	30
1950s	21	9	43	9	43
1960s	48	26	54	26	54
1970s	102	66	65	64	63
1980s	292	177	61	176	60
1990s	436	286	66	284	65
2000s	836	589	70	585	70
2010s	986	748	76	742	75

Different languages are covered by many abstract and index databases. Non-English articles are represented in forty-two of the databases (76%) of this study. More than half of the languages (14) are represented with only three or less samples and another four between five and eight samples. So, it is difficult to draw any conclusions with so few samples for these eighteen languages. For Sci-Hub and LibGen, only ten out of the twenty-six languages (38%) were retrieved with a wide range of success rate ranging from 5% to 75%. The most successful retrieval rates were for English (75%), German (31%), Chinese (25%) followed by French (19%). Strangely enough, Spanish success rate was only at 5% (two out of forty-three samples).

At the publisher's level, because the samples are distributed through 987 different ones, only the top 10 will be looked at since they count for 51% of the total samples (see table 3). Sci-Hub and LibGen retrieval rates are ranging from 88% (Oxford) to 100% (BMC, Cambridge and IEEE) with levels between 95% and 99% for the remaining publishers.

As it can be seen, the five major e-journal publishers (Elsevier, Sage, Springer, T&F and Wiley) are all well represented and have high retrieval rates between 97% and 99% in both Sci-Hub and LibGen.

Table 3. Overall distribution of retrieval rates by Publisher (Top 10 only) for Sci-Hub and LibGen (July 2017)

Publisher	# samples	# Sci-Hub	% Sci-Hub	# LibGen	% LibGen
Taylor & Francis	326	315	97	315	97
Elsevier	262	254	97	254	97
Wiley	243	238	98	238	98
Springer	208	204	98	204	98
Sage	166	165	99	164	99
Oxford	66	58	88	58	88
Cambridge	40	40	100	40	100
MCB-Emerald	29	28	96	28	96
IEEE	25	25	100	25	100
BMC	21	21	100	21	100
Brill	21	20	95	20	95

Cabanac (2016) who studied the LibGen content in 2014 reported that “LibGen hosts 36% of all articles with DOI” and was higher for articles published by Elsevier (64%), Springer (53%) and Wiley (59%) with an average of 68%. Two years after this study, the results here are not only in line with those of Cabanac (2016) but are showing higher retrieval rates for the same three publishers: i.e. a 100% rate for both Sci-Hub and LibGen. The other two major publishers in this study are also showing very good identical retrieval rates in both Sci-Hub and LibGen (100% for Sage and 96% for T&F).

Himmelstein et al. (2017) evaluated the coverage of Sci-Hub by publisher and found the following results: 97.3% for Elsevier, 95.3% for Sage, 89.4% for Springer Nature, 96.2% for Taylor & Francis and 94.8% for Wiley.

Some of the samples (11%) were also Open Access articles. It would be expected that all of them should be retrieved by Sci-Hub and LibGen. Curiously, this is not quite the case. Sci-Hub is showing a retrieval rate of only 41% while LibGen also has a low rate at 39%.

Himmelstein et al. (2017) reported that open access journals had a coverage of only 49.1%.

The 2,750 samples are distributed within 2,572 different journal titles each having between 1 and 5 articles (94% with 1 article, 5.5% with 2 articles, 0.5% with 3 articles, 0.12% with 4 articles and 0.04% with 5 articles).

8

Similar to Himmelstein et al. (2017), where they had few journals with a coverage between 5-75%, only 0.54% (14 titles) had a coverage ranging from 25% to 50%. Journals with a zero coverage represented 31% (798 titles) in Sci-Hub and 32% (812 titles) in LibGen. These levels are higher than the one found by Himmelstein et al. (2017): 11% (2,342 titles).

At the other end, journals with a 100% coverage represented 69% (1,761 titles) in Sci-Hub and 68% (1,746 titles) in LibGen.

Currency of Information

One important element in any searching and full-text retrieving tools, is how up to date they are. How do Sci-Hub and LibGen do keep up and maintain their content up to date? Both Nature and Science research articles (new issues and pre-pub ones) were monitored on a daily basis for the period of September 2016 to June 2017. Searches were done mainly by the DOIs supplied by both publishers but a very small proportion were retrieved using the title of the article with GS (6% for Nature and 10% for Science). Most of the Nature research articles were retrieved within 24 hours after they came out from the Nature web site: 99% for Sci-Hub, 98% for LibGen and only 8% for GS. Most of the Nature articles were made available in GS within a 4-5 days (77%) period with a range of a few hours and up to eight days.

All of the Science research articles (100%) were retrieved within 24 hours in both Sci-Hub and LibGen. Similar to Nature, only 9% of the articles were available within the 24 hour time frame in GS. Most of the Science articles were made available in GS within the 4-5 days (86%) period with a range from a few hours and up to seven days.

Because of the higher numbers of days that both Nature and Science research articles took to be made available in GS, Google was also searched along with the other platforms.

Surprisingly, most Nature and Science research articles were made available many days before they were in GS. For Nature articles, 96% of them were available within 24 hours with a range of a few hours to two days.

For Science articles, 100% of them were made available within that 24 hour time frame.

Conclusion

This study proved that there is a lot of full-text content available covering all disciplines at different levels and, surprisingly, up to date within usually the same day the article is published/released.

It is true that disciplines such as Law, Music and to some extent Business/Management are not covered as well as all of the other ones.

The breath of coverage in terms of publication year, language, publisher and disciplines is quite impressive in both Sci-Hub and LibGen.

More and more institutions and library patrons are retrieving their articles via their OA versions from different sources like institutional/subject repositories, hybrid journals, OA journals and social media. By looking at the large number of downloaded articles from both Sci-Hub and LibGen it is obvious that more and more people from all over the world, from both developing and developed countries are using these “new” platforms in order to get the articles they need. They are bypassing the more traditional forms of access that they have locally at their disposal: access through their own journal subscriptions and Interlibrary Loan.

Is it any different to get full-text content from Sci-Hub/LibGen than any other social media/networking platforms like Academia, Facebook, Mendeley, ResearchGATE, Twitter, etc.?

Are articles made available through these media more legitimate than the ones in Sci-Hub and LibGen? Using Sci-Hub/LibGen or not should remain a personal decision.

References

- Bohannon, J. (2016). Who's downloading pirated papers? Everyone. *Science*, 352(6285), 508-512. doi: 10.1126/science.352.6285.508
- Cabanac, G. (2016). Bibliogifts in LibGen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing. *Journal of the Association for Information Science and Technology*, 67(4), 874-884. doi: 10.1002/asi.23445
- Elbakyan, A., & Bohannon, J. (2016). Data from: Who's downloading pirated papers? Everyone. In: Dryad Digital Repository. doi: 10.5061/dryad.q447c
- Gardner, G. J., McLaughlin, S. R., & Asher, A. D. (2017, March 22-25). *Shadow Libraries and You: Sci-Hub Usage and the Future of ILL*. Paper presented at the ACRL 2017, Baltimore, Maryland.
- Greshake, B. (2016). Correlating the Sci-Hub data with World Bank Indicators and Identifying Academic Use. *The Winnower*. doi: 10.15200/winn.146485.57797
- Greshake, B. (2017). *Looking into Pandora's Box: The Content of Sci-Hub and its Usage [version 1; referees: 2 approved, 2 approved with reservations]* (Vol. 6).
- Himmelstein, D.S., Romero, A.R., McLaughlin, S.R., Greshake Tzovaras, B., Greene, C.S. (2017). Sci-Hub provides access to nearly all scholarly literature. *PeerJ Preprints*, 5:e3100v1, 10.7287/peerj.preprints.3100v1
- Machin-Mastromatteo, J. D., Uribe-Tirado, A., & Romero-Ortiz, M. E. (2016). Piracy of scientific papers in Latin America. *Information Development*, 32(5), 1806-1814. doi: 10.1177/02666666916671080
- Timus, N., & Babutsidze, Z. (2016). Pirating European Studies. [Commentary]. *Journal of Contemporary European Research*, 12(3), 783-791.
- Travis, J. (2016). In survey, most give thumbs-up to pirated papers. *Science*. doi: 10.1126/science.aaf5704