

Google Scholar, Sci-Hub and LibGen: Could they be our New Partners?

Louis Houle
McGill University, louis.houle@mcgill.ca

Louis Houle, "Google Scholar, Sci-Hub and LibGen: Could they be our New Partners?." *Proceedings of the IATUL Conferences*. Paper 3.
<http://docs.lib.purdue.edu/iatul/2017/partnership/3>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

GOOGLE SCHOLAR, SCI-HUB AND LIBGEN: COULD THEY BE OUR NEW PARTNERS?

Louis Houle
McGill University
Canada
louis.houle@mcgill.ca

Abstract

Since its debut 1 November 2004, librarians have raised several criticisms at Google Scholar (GS) such as its inconsistency of coverage and its currency and scope of coverage. It may have been true in the early years of Google Scholar but is this still through twelve years after? Is this sufficient to ignore it totally either in our information literacy programs or evaluate its value against the values of subscription-based abstracts and indexes?

In this era of severe budget constraints that libraries are facing, can we imagine of substituting most or all of our subject databases with the free access of Google Scholar for discoverability? How much overlap between our databases and Google Scholar? How reliable is Google Scholar? How stable is its content over time?

Open Access is getting to be 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. What can be said in terms of content and reliability of both Sci-Hub and LibGen?

This article reports on preliminary results of a one year study of Google Scholar where 2,750 random samples (peer review journal articles) coming from fifty-five different databases covering all disciplines (Arts & Humanities, Law, Music, Social Sciences and STM) are tested against GS. The samples have been searched against Google Scholar at four different intervals during the year. The same samples have been searched against both Sci-Hub and LibGen in order to see how much full-text content is available under these platforms. 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 Google Scholar, Sci-Hub and LibGen, research articles from both Nature and Science (from current issues, Nature Advance Online Publication and First Release from Science) were searched on a daily basis. Results are showing that most of the peer review articles are available in Google Scholar, Sci-Hub and LibGen.

Keywords

Open Access, OA, Google Scholar, GS, Google, Sci-Hub, Library Genesis, LibGen, Full-text, Peer Review Articles

Introduction

Since its debut in November 2004, Google Scholar is able to retrieve more and more scholarly journal article records from all the publicly accessible web sites and from subscription-based databases it is allowed to crawl (or harvest). GS can no longer be ignored since most students are starting their searches through this free internet search engine instead of using library catalogues and /or discovery tools. Searching Google Scholar facilitates the students by using an interface which is more familiar to them.

Few years after the launch of GS, many studies started to compare Google Scholar's performance with one or many databases (Chen, 2010a; Chen, 2010b; Falagas et al., 2008; Garcia-Perez, 2010; Jacso, 2005a; Jacso, 2005b; Levine-Clarck and Kraus, 2007; Lewandowski, 2010; Mayr and Walter, 2007; Mayr and Walter, 2008; Meier and Conkling, 2008; Neuhaus et al., 2006; Norris and Oppenheim, 2007; Pomerantz, 2006; Walters, 2007).

More recent studies (Asher et al., 2013; Bramer, 2016; Harzing, 2014; Harzing and Alakangas, 2016; Moed et al., 2016; Shariff et al., 2013) continued to compare GS on searchability, citations and comparisons with other searching databases.

Sixteen years after the initial launch of GS, are criticisms such as the inconsistency of coverage and the currency and scope of coverage raised by these studies still true?

Is this sufficient to ignore it totally either in information literacy programs or evaluate its value against the values of subscription-based abstracts and indexes (A&I)?

How much overlap between these A&I databases and Google Scholar? Are they still need all needed? Could Google Scholar be an alternative to subscription-based abstracting and indexing services? In this period of difficult budget situation facing academic institutions, these are important questions to ask ourselves and this is what this study will address. This study will also look at the stability of the GS content over a one year period.

Discoverability is one important aspect of any information seeking needs but the retrieving of the full-text content is as much important and necessary. 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, 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; Machin-Mastromatteo, 2016; Timus and Babutsidze, 2016) or LibGen (Cabanac, 2016) either for their content or usage patterns.

This study will use the same samples as the ones in Google Scholar in order to evaluate the peer review articles content availability in both Sci-Hub and LibGen and over a one year period also.

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 GS, Sci-Hub and LibGen. The following steps were used depending on the platform:

- a) In GS:
 1. Article title as a phrase between quotation marks
 2. If step one failed then by DOI or by the PubMed Central ID number (PMCID) whenever available
 3. If step 2 failed then article title as a phrase (no quotation marks) and the last name of one of the author
- b) 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
 - 3.
- c) 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 GS, Sci-Hub and LibGen. All 2,750 samples were queried quarterly between September 2016 and April 2017: October 2016, January 2017 and April 2017.

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

In order to measure the delay of coverage or the currency of the information in GS, Sci-Hub and LibGen, articles from two of the most prestigious journals were queried on a daily basis. From September 2016 to May 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 with 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 986 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 2,585 different journal titles
 - 2,439 journals supplying each only 1 article
 - 132 journals supplying each 2 articles
 - 10 journals supplying each 3 articles
 - 3 journals supplying each 4 articles
 - 1 journal supplying 5 articles
 -
 - 11% of samples (291) are Open Access articles

Results

As it can be seen from Figure 1, the overall retrieval rates for the 2,750 samples, as of April 2017, is pretty good. The discovery rate for Google Scholar is reaching a 95% ratio and the full-text retrieval rates for both Sci-Hub and LibGen are respectively 69% and 68%. This is not what one would have probably expected considering the breadth of the disciplines covered in this study. The majority of samples discovered in GS were done by article title (99.6%) and a very small proportion by DOI (0.4%). On the other hand, Sci-Hub and LibGen samples were mainly retrieved by article title (88% and 89% respectively) with a higher proportion than GS 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 (0.4%) and LibGen (0.1%).

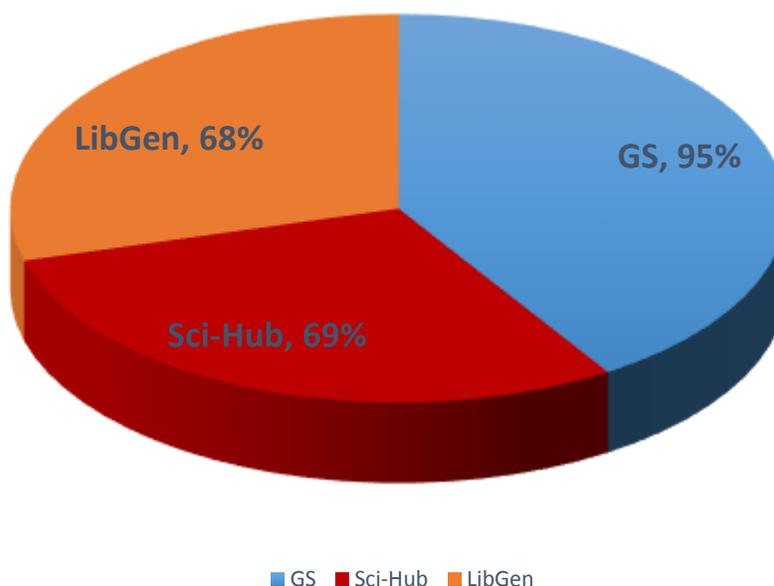


Figure 1. Overall Retrieval Rates for GS, Sci-Hub and LibGen samples in April 2017.

The distribution rates over all fifty-five databases ranges from 58% to 100% for GS, 20% to 100% for both Sci-Hub and LibGen. At the discipline level, the results are showing the lowest retrieval rate (58%) in GS for Music. Sci-Hub lowest retrieval rates are with 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 GS are seen with Agriculture/Environment, Education, Health/Biological Sciences, Multidisciplinary, Science and Social Sciences disciplines. For both Sci-Hub and LibGen, the highest retrieval rates are coming from the Multidisciplinary discipline.

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. The lowest retrieval rates in GS are coming from the RILM Abstracts of Music Literature (58%) and from the Index to Foreign Legal Periodicals (76%) databases. 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 and the Index to Foreign Legal Periodicals with both 28% and the Canadian Business & Current Affairs (32% for Sci-Hub and 30% for LibGen) databases. At the other end of the spectrum, there are fifteen databases (27%) showing a 100% retrieval rate represented in seven different disciplines (Agriculture/Environment, Business/Management, Education, Health/Biological Sciences, Multidisciplinary, Science and Social Sciences) in GS. More impressive is the fact that fifty-three databases (96%) are showing retrieval rates of 80% or higher

in GS. For both Sci-Hub and LibGen, three databases (5%) are showing a 100% retrieval rate (Elsevier ScienceDirect, Springer and Wiley) coming from only one discipline – Multidisciplinary. Only six databases resulted with retrieval rates lower than 50% for both Sci-Hub (20% - 42%) and LibGen (20% - 40%).

There does not seem to be any more major discrepancies in of the content availability by discipline as it was seen in the past few years after the launch of GS. What can be seen here is a lower coverage of peer review articles from the Music and Law disciplines. But GS has improved greatly its coverage from disciplines such as the Humanities, Chemistry, and some publishers such as Elsevier and ACS which were prone to have a low visibility few years after the launch of Google Scholar.

Database	Discipline	Platform	GS	Sci-Hub	LibGen
ABI Inform Global 1905- Academic Search Complete	Business/Management Multidisciplinary	ProQuest EBSCO	98 98	86 74	86 74
Agricola 1970 - Applied Science & Technology Full Text 1983 -	Agriculture/Environment Science	Ovid EBSCO/Wilson	100 100	76 82	74 82
Art Full Text 1984 - ATLA Religion Database with ATLASerials 1908- BIOSIS Previews + Archive 1926- British Humanities (BHI) Index 1962 - Business Source Complete	Arts Humanities Health/Biological Sciences Humanities Business/Management	EBSCO/Wilson EBSCO Ovid ProQuest EBSCO	94 80 88 98 100	62 42 78 96 72	62 40 78 96 72
CAB Abstracts + Archive 1910 - Canadian Business & Current Affairs Database (CBCA) 1971- CINAHL Plus with Full Text 1937 - Communication Abstracts 1972 -	Agriculture/Environment Business/Management Health/Biological Sciences Humanities	Ovid ProQuest EBSCO EBSCO	98 92 100 96	64 32 82 76	64 30 78 76
Compendex 1884 - Econlit 1886- Education Full text (Wilson/Ebsco) Embase Classic + Embase 1947- ERIC 1966 - (also via United States Dept. Education open access) Expanded Academic ASAP	Engineering Business/Management Education Health/Biological Sciences Education Multidisciplinary Humanities/Social Sciences (foreign language)	EI Village EBSCO EBSCO/Wilson Ovid EBSCO Gale EBSCO	96 100 98 96 100 98 96	72 68 56 58 70 86 64	72 66 52 54 70 84 64
Francis 1972- GeoRef 1693 -	Sciences (foreign language) Science	EBSCO ProQuest	96 82	64 60	64 60

Historical Abstracts 1955 -	Humanities	EBSCO	80	52	52
Index to Foreign Legal Periodicals 1985- Index to Legal Periodicals & Books Full Text	Law (foreign language)	Hein Online	76	28	28
Inspec	Law	EBSCO/Wilson	90	20	20
Library & Information Science Abstracts (LISA) 1969 -	Engineering/Science	EI Village	96	78	76
Library Literature & Information Science Full Text 1980- Linguistics and Language Behavior Abstracts (LLBA) 1973 -	Education	ProQuest	100	62	58
MathSciNet 1800- Medline 1946 -	Education	EBSCO/Wilson	92	42	40
Meteorological & Geostrophysical Abstracts 1974 -	Social Sciences	ProQuest	94	70	70
MLA International Bibliography 1962- Music Periodicals Database 1874 -	Science	AMS	90	54	54
OmniFile Full Text Mega PAIS Index 1914- Periodicals Archive Online	Health/Biological Sciences	Ovid	100	54	54
Philosopher's Index 1940 -	Science	ProQuest	96	90	90
ProQuest Research Library	Humanities	ProQuest	92	56	56
PsycINFO 1806 -	Music	ProQuest	82	54	54
PubMed 1946 - RILM Abstracts of Music Literature 1800- Sage	Multidisciplinary	EBSCO/Wilson	96	62	60
ScienceDirect	Humanities	ProQuest	100	88	88
SciFinder Scholar 1907 -	Humanities/social Sciences	ProQuest	94	52	52
Scopus	Humanities	ProQuest	98	86	86
Social Services Abstracts 1979 -	Multidisciplinary	ProQuest	100	70	70
Social Work Abstracts 1968 -	Social Sciences	Ovid	98	78	78
SocINDEX with Full Text 1908 -	Health/Biological Sciences	NIH	98	60	50
Sociological Abstracts 1952 -	Music	EBSCO	58	28	28
SPORTDiscus	Multidisciplinary	Sage	98	94	94
	Multidisciplinary	Elsevier	100	100	100
	Science	ACS	100	84	84
	Multidisciplinary	Elsevier	92	62	62
	Social Sciences	ProQuest	100	88	88
	Social Sciences	Ovid	96	72	72
	Social Sciences	EBSCO	96	64	64
	Social Sciences	ProQuest	96	88	88
	Multidisciplinary	EBSCO	94	60	58

Springer+BMC+Adis	Multidisciplinary	Springer-Nature	100	100	100
Taylor & Francis	Multidisciplinary	T&F Thomson Reuters	100	96	96
Web of Science	Multidisciplinary	Wiley	100	80	80
Wiley	Multidisciplinary Health/Biological Sciences	Ovid	100	100	100
Zoological Record 1864 -			86	52	52

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

Regarding Sci-Hub and LibGen it is hard to compare the retrieval data by discipline/database since there has not been studies similar to this one for Sci-Hub and LibGen. But 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 better 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 (94% for Sage and 96% for T&F).

Table 2 is listing the distribution of retrieval rates by decades for GS, 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 in the retrieval rates in GS: ranging from 89% (1970s) to 97% (2010s). A similar pattern is observed in both Sci-Hub and LibGen except for the 1980s where there is a slight decrease in the retrieval rates. Although the number of samples is small, GS is also successful with older articles such as the ones from the nineteenth and early twentieth centuries. Sci-Hub and LibGen are showing similar patterns 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 GS, 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. After verification, this article is also discoverable and retrievable by both GS and Sci-Hub.

For the publication year with 3% or more of the samples (see section on Samples Overview for the list), GS retrieval rates range from 91% (2002) to 99% (2011). For Sci-Hub retrieval rates are ranging from 60% (2009) to 89% (2016) and for LibGen from 59% (2009) to 86% (2016).

Decades	Total # samples	# GS	GS % of total	# Sci-Hub	Sci-Hub % of total	# LibGen	LibGen % of total
1840s	1	1	100	0	0	0	0
1860s	1	1	100	0	0	0	0
1880s	1	0	0	0	0	0	0
1890s	1	1	100	1	100	1	100
1900s	3	3	100	2	67	2	67
1910s	1	1	100	1	100	1	100
1920s	3	3	100	3	100	3	100
1930s	8	7	88	6	75	6	75
1940s	10	8	80	3	30	3	30

1950s	21	19	90	9	43	9	43
1960s	48	45	94	24	50	24	50
1970s	102	91	89	64	63	63	62
1980s	292	266	91	168	58	167	57
1990s	436	401	92	282	65	279	64
2000s	836	798	95	584	70	577	69
2010s	986	953	97	744	75	734	74

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

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 GS, twenty out of twenty-six languages (77%) were retrieved with different success rate ranging from 40% to 100%. For the four most represented languages, GS retrieval rates are very good: English (96%), French and German (75%) and Spanish (88%).

For Sci-Hub and LibGen, only ten out of twenty-six languages (38%) were retrieved with lower success rate than GS and ranging from 5% to 74%. The most successful retrieval rates were for English (74%), German (33%) 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 986 different ones, only the top 10 will be looked at (see Samples Overview section) since they count for 51% of the total samples. In GS, the retrieval rates are ranging between 96% (MCB-Emerald) to 100% (Brill, BMC, Cambridge, IEEE, Oxford, Springer and T&F). The remaining three publishers (Elsevier, Sage and Wiley) reached the 99% level.

Sci-Hub and LibGen retrieval rates are ranging from 85% (Oxford) to 100% (BMC, Cambridge and IEEE) with levels between 93% to 98% 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 95% and 100% in both Sci-Hub and LibGen.

Some of the samples (11%) were also Open Access articles. It would be expected that all of them should be retrieved by GS, Sci-Hub and LibGen. Curiously, this is not quite the case. Google Scholar is showing a retrieval rate of 97% - only nine articles were not found. Sci-Hub is showing a retrieval rate of only 40% while LibGen also has a low rate at 36%.

Currency of Information

One important element in any searching and full-text retrieving tools, is how up to date are they. How do GS, 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. 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 8% 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, 97% for LibGen and only 5% for GS. Most of the Nature articles were made available in GS within a 4-5 days (76%) period with a range of a few hours to 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 10% of the articles were available within the 24 hour time frame. Most of the Science articles were made available in GS within the 4-5 days period with a range from a few hours 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, 95% 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 period.

Conclusion

Sixteen years after the launch of Google Scholar, it is clear that the scholarly content is there and has increased tremendously from the first few years of activity. This is no surprise since the quantity of articles has exploded ever since along with the increase of mergers amongst the different publishers (e.g. Kluwer Academic with Springer in Spring 2004, Blackwell with Wiley in February 2007, CSA with ProQuest in February 2007, BMC with Springer in 2008, Wilson with EBSCO in June 2011 and MacMillan Education/Nature/Palgrave with Springer in 2015), which created bigger packages (more journal titles, addition of e-journal backfiles and many new content databases from aggregators such as EBSCO and ProQuest) more visible and discoverable via Google Scholar. The fact that, since the launch of GS in November 2004, there has been some publishers who had been reluctant in the past to make their content harvested by GS, have either changed their mind over the years or had no choice because of the merger with another publisher whose policy was exactly the opposite.

Let's not forget also, when negotiating e-resource licenses, librarians are asking the publishers to make their content harvested by Google and Google Scholar since these platforms are the preferred discovery tools that academic communities are using.

Based on the data gathered here, it is safe to say that an institution could go ahead and cancel some, many or all of its A&I databases and rely mainly on GS for its discovery of peer review articles. The wide coverage of disciplines, languages, publication years, publishers along with the timeliness of the information are proving that GS is a tool that should be a high priority in an academic environment.

On the other hand, regarding Sci-Hub and LibGen, 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.

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.? Using Sci-Hub or LibGen will be a personal decision.

References

- Asher, A. D., Duke, L. M., & Wilson, S. (2013). Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources. *College & Research Libraries*, 74(5), 464-488. doi: 10.5860/crl-374
- Bramer, W. M. (2016). Variation in number of hits for complex searches in Google Scholar. *Journal of the Medical Library Association : JMLA*, 104(2), 143-145. doi: 10.3163/1536-5050.104.2.009
- 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
- Chen, X. (2010a). The Declining Value of Subscription-based Abstracting and Indexing Services in the New Knowledge Dissemination Era. *Serials Review*, 36(2), 79-85. doi: 10.1080/00987913.2010.10765288
- Chen, X. (2010b). Google Scholar's Dramatic Coverage Improvement Five Years after Debut. *Serials Review*, 36(4), 221-226. doi: 10.1080/00987913.2010.10765321
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), 338-342. doi: 10.1096/fj.07-9492LSF
- García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. *Journal of the American Society for Information Science and Technology*, 61(10), 2070-2085. doi: 10.1002/asi.21372
- 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).
- Harzing, A.-W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. [journal article]. *Scientometrics*, 98(1), 565-575. doi: 10.1007/s11192-013-0975-y
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. [journal article]. *Scientometrics*, 106(2), 787-804. doi: 10.1007/s11192-015-1798-9
- Jacso, P. (2005). Comparison and Analysis of the Citedness Scores in Web of Science and Google Scholar. In E. A. Fox, E. J. Neuhold, P. Premismit & V. Wuwongse (Eds.), *Digital Libraries: Implementing Strategies and Sharing Experiences: 8th International Conference on Asian Digital Libraries, ICADL 2005, Bangkok, Thailand, December 12-15, 2005. Proceedings* (pp. 360-369). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Jacsó, P. (2005). Visualizing overlap and rank differences among web-wide search engines: Some free tools and services. *Online Information Review*, 29(5), 554-560. doi: doi:10.1108/14684520510628927
- Levine-Clark, M., & Kraus, J. (2007). Finding Chemistry Information Using Google Scholar. *Science & Technology Libraries*, 27(4), 3-17. doi: 10.1300/J122v27n04_02
- Lewandowski, D. (2010). Google Scholar as a tool for discovering journal articles in library and information science. *Online Information Review*, 34(2), 250-262. doi: doi:10.1108/14684521011036972
- 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: doi:10.1177/0266666916671080

- Mayr, P., & Walter, A.-K. (2008). Studying Journal Coverage in Google Scholar. *Journal of Library Administration*, 47(1-2), 81-99. doi: 10.1080/01930820802110894
- Mayr, P., & Walter, A. K. (2007). An exploratory study of Google Scholar. *Online Information Review*, 31(6), 814-830. doi:doi:10.1108/14684520710841784
- Meier, J. J., & Conkling, T. W. (2008). Google Scholar's Coverage of the Engineering Literature: An Empirical Study. *The Journal of Academic Librarianship*, 34(3), 196-201. doi: <http://dx.doi.org/10.1016/j.acalib.2008.03.002>
- Moed, H. F., Bar-Ilan, J., & Halevi, G. (2016). A new methodology for comparing Google Scholar and Scopus. *Journal of Informetrics*, 10(2), 533-551. doi: <http://dx.doi.org/10.1016/j.joi.2016.04.017>
- Neuhaus, C., Neuhaus, E., Asher, A., & Wrede, C. (2006). The depth and breadth of Google Scholar: An empirical study. *portal: Libraries and the Academy*, 6(2), 127-141.
- Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics*, 1(2), 161-169. doi: <http://dx.doi.org/10.1016/j.joi.2006.12.001>
- Pomerantz, J. (2006). Google Scholar and 100 Percent Availability of Information. [Article]. *Information Technology & Libraries*, 25(2), 52-56.
- Shariff, S. Z., Bejaimal, S. A., Sontrop, J. M., Iansavichus, A. V., Haynes, R. B., Weir, M. A., & Garg, A. X. (2013). Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *Journal of medical Internet research*, 15(8), e164.
- Timus, N., & Babutsidze, Z. (2016). Pirating European Studies. [Commentary]. *Journal of Contemporary European Research*, 12(3), 783-791.
- Walters, W. H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, 43(4), 1121-1132. doi: <http://dx.doi.org/10.1016/j.ipm.2006.08.006>