Sci-Hub: What Librarians Should Know and Do about Article Piracy

1 author:

Matthew B Hoy

Mayo Foundation for Medical Education and Research

25 PUBLICATIONS   **1,277** CITATIONS

Routledge
Taylor & Francis Group

EMERGING TECHNOLOGIES

Matthew B. Hoy and Tara Brigham, Column Editors

Sci-Hub: What Librarians Should Know and Do about Article Piracy

Matthew B. Hoy

Mayo Clinic Health System–Eau Claire, Eau Claire, Wisconsin, USA

ABSTRACT

The high cost of journal articles has driven many researchers to turn to a new way of getting access: "pirate" article sites. Sci-Hub, the largest and best known of these sites, currently offers instant access to more than 58 million journal articles. Users attracted by the ease of use and breadth of the collection may not realize that these articles are often obtained using stolen credentials and downloading them may be illegal. This article will briefly describe Sci-Hub and how it works, the legal and ethical issues it raises, and the problems it may cause for librarians. Librarians should be aware of Sci-Hub and the ways it may change their patrons' expectations. They should also understand the risks Sci-Hub can pose to their patrons and their institutions.

Introduction

Prices for academic journals, particularly science journals, have been on the rise for decades.[1] For researchers without access to an institutional library, these cost increases have come to mean paying $30 or more for each journal article they need to read. When they need to read hundreds of papers, these costs quickly become unmanageable. In 2011, a researcher in Kazakhstan created a website call Sci-Hub that allows anyone to access the full text of millions of scientific articles for free. Librarians should be aware of Sci-Hub because many of their patrons are already using it, and it will undoubtedly affect their expectations about what their library should provide and how quickly they should be able to get articles. Librarians should also understand the risks that Sci-Hub can pose to their patrons, their institutions, and themselves.

What is SCI-HUB?

The tagline on the Sci-Hub website reads "to remove all barriers in the way of science," and the site's "about" section describes it as "the first pirate website

in the world to provide mass and public access to tens of millions of research papers."[2] The page goes on to list three key ideas that guide it: "knowledge to all," "no copyright," and "open access."[2] Sci-Hub was created by Alexandra Elbakyan, a computer science graduate student from Kazakhstan. Elbakyan has stated that her reasons for creating Sci-Hub came out of the frustrations she faced as a researcher; she needed to read thousands of articles but was unable to access them legally without paying around $30 for each article.[3] She and other researchers in her situation had previously relied on colleagues with better access to supply the articles they needed via web forums or Twitter, but she wanted to find a solution that could be automated and would scale to meet demand.[4,5] Elbakyan built Sci-Hub to solve this problem; it allows users to anonymously find specific journal articles and download them for free, using donated or stolen access credentials.

Sci-Hub is not a discovery tool; users need to know exactly what article they are looking for when they use it. Users search for articles using URLs, DOI numbers, PubMed ID numbers, complete titles, or other unique identifiers. Once the site has determined the specific article that the user is trying to locate, it queries the Library Genesis database <http://gen.lib. rus.ec/> to see if there is a copy available. The Library Genesis database is a separate "pirate" entity, but it works together with Sci-Hub in a sym-biotic relationship. If a copy of the article already exists in Library Genesis, Sci-Hub sends a copy to the user. If Library Genesis does not have a copy, Sci-Hub begins cycling through its list of proxy credentials until it finds one that has access to that article. It uses that proxy to access the article, serves a copy to the user, and uploads a copy into the Library Genesis database. If that article is requested again in the future, Sci-Hub will be able to get it directly from Library Genesis without needing to use proxy credentials. This automated system means that even if an unauthorized proxy login is discovered quickly, thousands of articles may have already been down-loaded and can then be shared over and over again. Using this "crowdsourc-ing" method, Sci-Hub and Library Genesis have built up a huge collection of over 58,00,000 journal articles.[6] More are added each day, and the more popular a paper is, the more quickly it will be added to the database. Library Genesis also makes copies of its entire database available for download with instructions on how to create "mirror" sites. This means that even if Library Genesis is eventually shut down, copy-cat sites can quickly pop up to replace it.

How Sci-Hub gained access to so many institutional proxies is not entirely clear. Elbakyan has repeatedly claimed that academics frustrated with the status quo willingly donate their proxy logins.[7] At least one Marquette University professor claims to have had his login credentials phished and used on Sci-Hub.[8] Whether Sci-Hub is actively phishing or not is moot: as Smith points out "there's nothing to stop supporters taking matters in their own

hands and passing on acquired username and password combinations."[9] Librarians report seeing huge spikes in proxy activity from compromised logins, so these accounts should be readily apparent when viewing proxy logs. Library staff should contact their institution's information technology (IT) and legal departments if they suspect a login is being used to provide Sci-Hub with proxy access.

Legal and Ethical Issues

Sci-Hub is enabling copyright law violation on a massive scale, and publishers have noticed. Members of the American Association of Publishers, led by Elsevier, successfully sued to have Sci-Hub shut down in 2015. The site was migrated to a new domain name hosted outside U.S. jurisdiction and resumed service just 18 days after the sci-hub.org domain name was seized.[10] Elbakyan justifies this open defiance of copyright law by noting that "we never received any complaints from authors or researchers, only Elsevier is complaining about free distribution of knowledge."[3] The Association of American Publishers fired back in a statement on its website, noting that Sci-Hub "should not be equated with any legitimate open access publishing practices, and its efforts have nothing to do with the real innovation going on in publishing to advance access to published material."[11]

Sci-Hub itself may be infringing copyright by gathering, storing, and sharing articles, but the legality of its actions depends on jurisdiction and is too complex to be easily solved. What is clear is that many Sci-Hub users are violating copyright law. Some users are doing nothing wrong according to the laws of their home country. According to server logs released by Sci-Hub, users in Iran were the most prolific article downloaders, and Iran is currently not a signatory to any copyright treaties with the United States.[4,12] However, users in the United States ranked fifth on the list of most frequent downloaders, and U.S. copyright laws clearly apply to them.

Despite the questionable legality of Sci-Hub's service, there are ethical issues that help justify its existence. Several physicians in Peru recently published a letter in *The Lancet Global Health* noting that Sci-Hub has become an important source of medical literature there due to lack of access via other methods.[13] Services such as HINARI and HathiTrust can help alleviate access issues in poor and developing countries, but they also have barriers to entry and lack the immediacy of Sci-Hub. 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?

Many authors and reviewers have also noted that publisher's high profit margins are built on work provided to them for free. Some view this as justification for Sci-Hub; they consider it a method for reclaiming research

that has been locked behind paywalls by "greedy" publishers.[14] When research has been paid for with public funding and provided to the publisher at no cost, it is hard to find fault with their argument that something must change, but using Sci-Hub to reform the current publishing model may have unintended consequences. For every high-profit mega-publisher, there are dozens of smaller publishers that Sci-Hub could drive out of business.[8] It is also important to note that Sci-Hub is not offering an alternative to traditional publishing; all of the articles contained there are created by and dependent on the traditional model. Users are searching via PubMed ID numbers, DOIs, and other unique identifiers created and maintained by the traditional publishing model, and are downloading files generated by existing publishers. Sci-Hub offers no editorial or peer-review services and is not likely to start these services. If the traditional publishing model changes or goes away, Sci-Hub will have nothing to offer but a significant number of aging articles. Dissatisfaction with current publishing models may also explain why some researchers share their login credentials with Sci-Hub. What better way to subvert the current publishing model than by intentionally or "accidentally" sharing proxy logins, making their institution's entire collection available to Sci-Hub?

Despite these issues, most users are not turning to Sci-Hub because of legal or ethical issues; they simply want the fastest and easiest access to articles. McNutt noted that "much traffic to Sci-Hub is from researchers who already have access to the articles they seek through mechanisms such as site licenses, open access, or other means."[15] Bohannon's analysis of server logs showed "users in the United States seem to congregate near universities and likely have institutional access to the articles they request."[4] For most of these users, Sci-Hub is a matter of convenience, not access. As librarians know, patrons usually follow Zipf's "principle of least effort," whether it yields the best results or not.[16]

What Should Librarians Do About SCI-HUB?

Librarians are caught in the middle of the Sci-Hub issue. On one side are authors and researchers and on the other, publishers. Authors want their work disseminated as widely as possible, and researchers want easy access to as much literature as possible. Publishers want to control access to their copyrighted material to maximize profit. Librarians have to balance the interests of both groups; as Ruff noted, librarians "understand the challenges scholars face in gaining access to information, [but] they are also bound by their contracts with publishers."[17]

Copyright infringement is not a new issue. Librarians have dealt with this for decades by educating patrons when infringement occurs and setting policies in place to discourage it. If a patron is determined to violate copyright

law, there is little that can be done to stop them. Sci-Hub is the same problem, automated and on a massive scale. Librarians can do little to stop patrons from using Sci-Hub, but there is always room for education and explanation of why other sources may be a better choice. Actively banning the use of pirate sites is ineffective. It alerts more patrons to their existence and creates a negative power dynamic between librarians and their patrons. Blocking pirate sites with Internet filters does not work either, as determined users can always find a way around filters, and domain names change frequently. As always, the best course of action for librarians is to educate their patrons when they see infringement happening.

While librarians cannot stop their patrons from using pirate article sites, they *can* stop their institution from being party to that infringement. Frequently checking proxy logs for excessive use and quickly disabling compromised logins can lower the chances of an institution's access being used to seed content into Sci-Hub. Librarians may also want to notify their institutions' IT and legal departments about the potential risks of compromised proxy logins. For many institutions, these logins can grant access to more than just library resources.

Conclusion

Pirate article databases like Sci-Hub represent a major threat to traditional publishing models. Users love the simple interface and broad access to articles, but these sites violate copyright laws on a massive scale, and publishers are working hard to shut them down. Current legal proceedings have been unsuccessful in turning off Sci-Hub, and future attempts are likely to meet similar roadblocks. Librarians should know about sites like Sci-Hub and work to ensure their patrons are aware of all the legal mechanisms they have to obtain articles. When patrons are actively using Sci-Hub, librarians can educate them about the legal issues surrounding the site. Institutions may choose to block the Sci-Hub domain(s), but filtering is usually ineffective as domain names can change and users are persistent at routing around filtering. Library staff should watch their proxy logs for users with excessive download traffic, as that may indicate a stolen proxy login. Although unlikely, these stolen logins could open the library to future liability. Whether Sci-Hub is a pathway to "true" open access or just a haven for pirates and thieves is still open to debate. Either way, pirate article sites are likely here to stay and librarians should prepare accordingly.

Notes on Contributor

Matthew B. Hoy, MLIS, AHIP (hoy.matt@mayo.edu) is Supervisor of the Medical Library, Mayo Clinic Health System – Eau Claire, 1221 Whipple Street, Eau Claire, WI 54701.

References

1. Rose-Wiles, Lisa M. "The High Cost of Science Journals: A Case Study and Discussion." *Journal of Electronic Resources Librarianship* 23 no. 3 (September 20, 2011): 219–241. doi:10.1080/1941126X.2011.601225.

2. "Sci-Hub: Removing Barriers in the Way of Science." Accessed September 6, 2016. http://sci-hub.bz/.

3. "*Elsevier Inc. et al. v. Sci-Hub et al.* "Docket Item 50 | United States Courts Archive." September 15, 2015. https://www.unitedstatescourts.org/federal/nysd/442951/50-0.html.

4. Bohannon, John. "Who's Downloading Pirated Papers? Everyone." *Science* 352, no. 6285 (April 29, 2016): 508–512. doi:10.1126/science.352.6285.508.

5. Bohannon, John. "The Frustrated Science Student behind Sci-Hub." *Science* 352, no. 6285 (April 29, 2016): 511. doi:10.1126/science.352.6285.511.

6. Cabanac, Guillaume. "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, no. 4 (March 27, 2015): 874–884. doi:10.1002/asi.23445.

7. "Does Sci-Hub Phish for Credentials?" Sauropod Vertebra Picture of the Week. February 25, 2016. https://svpow.com/2016/02/25/does-sci-hub-phish-for-credentials/.

8. Banks, Marcus. "Sci-Hub: What It Is and Why It Matters." *American Libraries Magazine*, May 31, 2016. https://americanlibrariesmagazine.org/2016/05/31/why-sci-hub-matters/.

9. Smith, David. "Sci-Hub: How Does It Work?" *The Scholarly Kitchen*. February 25, 2016. https://scholarlykitchen.sspnet.org/2016/02/25/sci-hub-how-does-it-work/.

10. Waddell, Kaveh. "The Research Pirates of the Dark Web." *The Atlantic*, February 9, 2016. http://www.theatlantic.com/technology/archive/2016/02/the-research-pirates-of-the-dark-web/461829/.

11. Association of American Publishers. "AAP Statement on Sci-Hub." May 10, 2016. http://newsroom.publishers.org/aap-statement-on-sci-hub.

12. Library of Congress, Copyright Office. "International Copyright Relations of the United States." Accessed October 20, 2016. http://purl.access.gpo.gov/GPO/LPS101162.

13. Bendezú-Quispe, Guido, Wendy Nieto-Gutiérrez, Josmel Pacheco-Mendoza, and Alvaro Taype-Rondan. "Sci-Hub and Medical Practice: An Ethical Dilemma in Peru." *The Lancet Global Health* 4, no. 9 (September, 2016): e608. doi:10.1016/S2214-109X(16)30188-7.

14. Dobbs, David. "Testify: The Open-Science Movement Catches Fire." *WIRED*. January 30, 2012. https://www.wired.com/2012/01/testify-the-open-science-movement-catches-fire/.

15. McNutt, Marcia. "My Love-Hate of Sci-Hub." *Science* 352, no. 6285 (April 29, 2016): 497. doi:10.1126/science.aaf9419.

16. Zipf, George Kingsley. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology.* Cambridge, MA: Addison-Wesley Press, 1949.

17. Ruff, Corinne. "Librarians Find Themselves Caught Between Journal Pirates and Publishers." February 18, 2016. *The Chronicle of Higher Education.* http://www.chronicle.com/article/Librarians-Find-Themselves/235353/.