

Submitted for publication 7/23/2017 with added late-breaking reference 8/3/2017 – last revision 8/4/2017 eic

Abstract: Piracy of subscription-based scholarly literature has reached a new peak with the advent of Sci-Hub and other sites like it. Sci-Hub is a collection of over 68 million items led by and compiled by a researcher from Kazakhstan. This interview was conducted to gain insights from an academic library director, Michael Levine-Clark, who with other colleagues is doing research on how scholars discover and access research materials. Peter Katz, who works for Elsevier, discusses what Sci-Hub means for him as someone responsible for identifying and blocking activity that sets off alerts in their usage system indicating that the activity may be being undertaken by unauthorized users. These two individuals provide thoughtful observations about what Sci-Hub may mean for the future of article discovery in academic libraries and how open access models will impact and influence the dynamic between pirated materials and those found behind a paywall.

Key words: Sci-Hub; illegal downloading; academic libraries

Uncommon Conversations

Sci-Hub - what is it, and why does it matter to academic libraries? An Interview with Michael Levine-Clark and Peter Katz

Edited by Eleanor I. Cook

Eleanor I. Cook, J.Y. Joyner Library, East Carolina University, Greenville, NC cooke@ecu.edu
& Michael Levine-Clark, Dean of Libraries, University of Denver & Peter Katz, Vice President, Customer Lifecycle Solutions Support, Elsevier

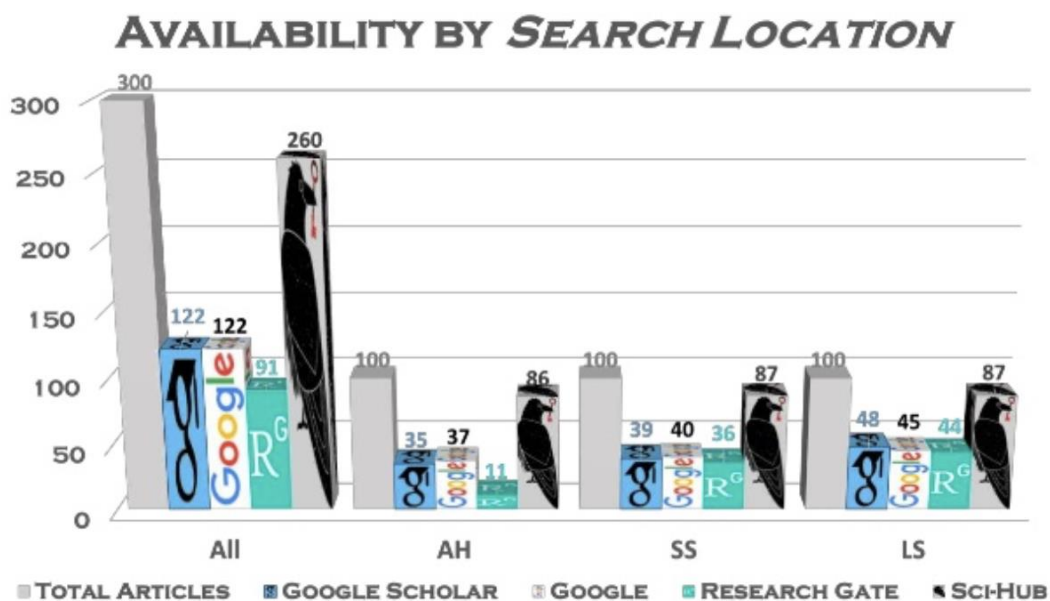
The following questions were posed to two members of the scholarly communications community.

EIC: First, a brief introduction about what Sci-Hub is for our readers. Sci-Hub is a huge (reportedly over 68 million items) collection of research articles and other materials compiled by a group led by Alexandra Elbakyan, a researcher from Kazakhstan. The materials available in Sci-Hub are for the most part copyrighted and therefore otherwise not accessible except by those who have access to a paid subscription. The purpose of Sci-Hub is to provide availability of published content to disadvantaged researchers who live in parts of the world where legitimate access is unaffordable. However, the methods employed to compile the database include use of -compromised user credentials and other hacking strategies whereby access is gained to IP-authenticated content from academic institutions. There are several good overviews of Sci-Hub that readers can refer to (assuming they have access to these articles!) (1)

This interview should not be read as an endorsement or support for Sci-Hub or similar products.

1. **EIC:** Michael and Peter, I have asked you two to reflect on some questions related to the entity known as Sci-Hub. First please introduce yourselves in this context and tell us what you have learned about Sci-Hub in the course of your research and/or work responsibilities.

MLC: With Jason Price (SCELC) and John McDonald (EBSCO) I have been working on a series of studies looking at relative discoverability and availability (meaning full-text access) of articles through Google and Google Scholar, the major library discovery systems, scholarly sharing sites such as ResearchGate, and Sci-Hub. There's a great deal of nuance in the findings, so I encourage readers to look at the slides we've presented for more detail, but one of the key findings is that Sci-Hub presents the full text of the highest percentage of articles of any of these resources. (2)



Out of a random sample of 300 articles, with 100 each from the arts & humanities, the social sciences, and the life sciences, 260 (87%) were available via Sci-Hub. Though we did not focus on this, an even higher percentage of articles from major STM publishers is available through Sci-Hub. This high level of full-text availability makes Sci-Hub a compelling option for *all* users, whether they have access through a legitimate source or not; it's understandable (though not excusable) that so many people make such high use of this site.

PK: For the past several years, I have been responsible for identifying and blocking usage activity that sets off alerts in our usage system indicating that the activity may be being undertaken by unauthorized users. In many such cases the purpose is to steal the content to make it available in sites like Sci-Hub, which provide access at no charge, or a variety of other sites located in China, Iran and other countries that charge users for content at rates lower than the single article price charged by commercial and society publishers on their branded websites. Ironically, many of the articles hosted on the illegal sites are available at no cost on publisher sites since they are published under various Open Access models.

The fact that Sci-Hub has already built such a large corpus of journal articles and books might lead one to believe that piracy would be slowing--but that is not the case. Each day we see suspicious activity, which when shared with our customers, results in their discovering that student and staff IDs have been compromised and used to download articles.

This activity seems to be centered in China; IPs owned by Chinese Internet service providers (ISP's) are most typically used to initiate the downloads, however the individuals undertaking these efforts often rent IPs from ISPs in Europe and the US to mask their whereabouts. It's my view that there are a number of entities building Sci-Hub copycat sites in China and Iran. I believe that Sci-Hub's activity is now more discreet as they need to download fewer articles. It is important to note that Elbyakan has admitted under oath that she used stolen credentials to steal content from publishers and that she was found guilty of copyright violation in a court of law.

2. **EIC:** Based on your experiences, what observations can you make about scholars' attitudes regarding Sci-Hub and similar "rogue" databases of research materials?

PK: I have not had any direct conversations with researchers regarding attitudes about use of these sites but I have read postings from those who argue that there is a moral argument in favor of using Sci-Hub. If I were in a conversation with someone holding this view my response would be, "Sorry, but there is no such thing as a moral reason for either stealing published content or using material you know has been stolen."

All of the librarians I interact with are as concerned as I am about the breaches and theft of content and do their best to cure the breaches.

MLC: Similarly to Peter, I have not spoken directly with anyone who uses Sci-Hub. But a large survey of researchers, published on the *Science* website, found that almost 88% of 10,839 respondents did not believe it was wrong to use Sci-Hub. (3) There is clearly a widespread attitude among researchers that it's ok to use (and even support) Sci-Hub.

I have talked to a number of faculty members at my own institution who are regular users of ResearchGate and Academic.edu -- resources that John and Jason and I have characterized as "rogue"-- in comparison with Sci-Hub, which we characterize as "pirate". ResearchGate and similar sites have plenty of content that has been uploaded legitimately, but also include lots of articles for which there is no evidence that the version available can legitimately be shared.

We as librarians should educate our users about how rogue and pirate sites operate, how to determine what versions of their own authored articles they should be able to upload, and how to find versions of other articles legitimately.

Clearly scholars are using tools like Sci-Hub because they feel some animosity toward publishers (and the *Science* survey validates that), but they are also using these tools because we (librarians, publishers, and vendors) have created systems for discovering and accessing articles that are complicated and confusing. Wherever we can, we need to make it so easy to get to the full text of an article that there's no reason to use Sci-Hub.

3. **EIC:** There was a court ruling recently awarding Elsevier monetary damages for piracy of content. Seeing that it may be impossible to collect on this, what other methods might publishers employ to protect their intellectual property?

PK: Many publishers are implementing more robust security for their online services. Since IP recognition is the most common form of authentication, blocking IPs that are experiencing suspicious usage is the most common remedy. This can be done automatically using tools that trigger blocks when certain thresholds are reached or manually after the fact. Some publishers watermark articles in order to be able to identify stolen content after it is downloaded.

Publishers should work more closely with proxy and link resolver providers to improve the ability of their services to identify suspicious activity, and also work closely with libraries to ensure that the customers know how to implement improved security measures.

New modes of authentication are being developed that provide better protection for publishers' IP without limiting the needs of libraries to provide broad access to their communities of licensed content. The RA21 project, (Research Access for the 21st Century) ra21.org a joint project of the International Association of Scientific, Technical and Medical Publishers (STM) and National Information Standards Organization (NISO) is currently running several pilots to test new methods of authentication. I urge readers of this interview who are not yet aware of RA21 to visit the site and consider taking part in one of the pilots.

MLC: The University of Denver has signed on to pilot RA21, mostly because we want to make the process of accessing our licensed resources easier for our users. In addition to making it harder for piracy to occur, RA21 could have the additional benefit of making library-based discovery and access to resources more appealing to users. So a side benefit may be that students and scholars at our institutions won't see any need to use Sci-Hub.

4. **EIC:** Assuming that the massive amounts of content that have been released into the public sphere cannot ever be returned to a region behind a paywall, how will this affect future publishing? Might this be the tipping point for a new business model that embraces more open access principles?

PK: I wouldn't connect article theft to the growth of open access, or a tipping point, as they are completely different things. But yes, many publishers have added OA to their business models including Elsevier, which is the second largest publisher of Open Access articles. And I certainly believe that OA will continue to grow. But more importantly, the future of publishing is not tied to "the PDF found by 'the DOI,'" it's about accessing content in a variety of formats, and extracting and delivering the knowledge contained within them. Publishers will always continue to find new and innovative ways to deliver content to scholars, the PDFs being just one of them. Sites like Sci-Hub will not be able to compete with many of the innovative new tools publishers are creating and delivering to customers.

MLC: That's an interesting question. Assuming that publishers manage to block the flow of articles to Sci-Hub, then Sci-Hub becomes an increasingly older collection of articles as time goes by -- content that often will flip to OA post-embargo anyway. Perhaps publishers will be more likely to make older content OA because it's already freely available. But the more important question is whether it will have any impact on new publications -- and this seems to be dependent on whether publishers can find a way to prevent articles from ending up on Sci-Hub in the future. If Sci-Hub continues to grow, and continues to

have the most current articles, then it might be one of a number of factors that convince a publisher that flipping a particular journal to OA would make good financial sense -- but it would have to be just one of many factors.

5. **EIC:** A library could conceivably cancel all of its subscriptions and provide access through Sci-Hub. Why shouldn't librarians embrace Sci-Hub and similar sites as a solution to budget woes?

PK: Sci-Hub hosts content that has been stolen from publishers and has been convicted in a court of law of doing so. As I said earlier, there is no moral justification for using content you know has been stolen.

MLC: Librarians should be concerned about ethical uses of information, and regardless of how librarians think about publishers and publisher profits Sci-Hub is based upon theft. As we work with our patrons, we should be teaching about how to access information responsibly, and that means obeying the laws that are in place while also advocating for open access. We should also be teaching our patrons that sharing their login information places their personal information at risk, even though they may see as a way of helping others get access to content.

EIC: Even as this column entry goes to press, a startling new study reported in *Science* (4) suggests the possibility that utter doom for the subscription journal may be at hand due to the volume of content available through Sci-Hub. If trends like this are not somehow counterbalanced, we may be entering a whole new era for libraries and for the information industry.

(1) Here are some introductory background readings on Sci-Hub - just a few selections:
Banks, M. (2016). What sci-hub is and why it matters. *American Libraries*, 47(6), 46-48. Retrieved from <https://search.proquest.com.jproxy.lib.ecu.edu/docview/1793930437?accountid=10639>

Russell, C. and Sanchez, E. (2016). Sci-Hub unmasked: Piracy, information policy, and your library. *College & Research Libraries News*, 77 (3), 122-125. Retrieved from <https://doi.org/10.5860/crln.77.3.9457>

Anderson, R. (2016). Sci-Hub and Academic Identity Theft: An Open Letter to University Faculty Everywhere. *Scholarly Kitchen*, May 19. Retrieved from <https://scholarlykitchen.sspnet.org/2016/05/19/sci-hub-and-academic-identity-theft-an-open-letter-to-university-faculty-everywhere/>

(2) Levine-Clark, Michael, John McDonald, and Jason Price, "Availability of Freely Available Articles from Gold, Green, Rogue, and Pirated Sources: How do Library Knowledge Bases Stack Up?" Electronic Resources & Libraries, Austin, April 4, 2017 (<https://www.slideshare.net/MichaelLevineClark/discovery-of-oa-articles>) and:

Levine-Clark, Michael, John McDonald, and Jason Price, "Access to Freely Available Journal Articles: Gold, Green, and Rogue Open Access Across the Disciplines," The Charleston Conference, November 3, 2016 (<https://www.slideshare.net/jpricein/access-to-freely-available-journal-articles-gold-green-and-rogue-open-access-across-the-disciplines>)

(3) Travis, John (2016). "In Survey, Most Give Thumbs-Up to Pirated Papers," Science Insider, *Science*. May 6, 2016. (<http://www.sciencemag.org/news/2016/05/survey-most-give-thumbs-pirated-papers>).

(4) McKenzie, Lindsay (2017). "Sci-Hub's cache of pirated papers is so big, subscription journals are doomed, data analyst suggests," *Science*. July 27, 2017. (<http://www.sciencemag.org/news/2017/07/sci-hub-s-cache-pirated-papers-so-big-subscription-journals-are-doomed-data-analyst>)