Walking the Plank: How Scholarly Piracy Affects Publishers, Libraries and Their Users

Laurie Morrison, Carol Stephenson, and Elizabeth Yates*

Introduction
The arrival of technology supporting peer-to-peer (P2P) file sharing in scholarly communication has, until recently, had minimal impact on libraries. However, threats posed by pirate sites including Library Genesis Project (LibGen) and Sci-Hub are now impacting both library users and library licensing agreements with publishers. Publishers are nervous as they witness their proprietary content leaking out of paywalled systems—not just hundreds of thousands of articles, but millions. Accordingly, publishers are monitoring activities in licensed products very closely for any behavior that they deem suspicious. When a user’s activities cause a publisher to question whether materials are being pirated, the outcomes can vary. Consequences can range from relatively minor inconvenience for blocked users, who must find workarounds to access scholarly content—to the potential for major disruption of a centuries-old proprietary publishing system. This article uses a case study involving a student at Brock University to highlight significant challenges facing libraries and the rights of their users in the current environment of piracy-wary academic publishers.

Case Study: Access Denied
“I feel like I’m being penalized for my honesty.” That’s how a graduate student at Brock University felt in January 2016, after her legitimate quest to download several hundred articles for a meta-analysis project turned into a protracted—and ultimately unsuccessful—negotiation with the American Psychological Association. Sarah† had downloaded about 20 articles from the PsycINFO database when she received the following screen prompt:

The APA PsycNET Terms and Conditions prohibit “Systematic downloading of content, whether done manually or by technological means.” Please contact Permissions@apa.org if you are interested in data mining or wish to conduct a systematic review or meta analysis with PsycINFO data. If you need help with access to resources on APA PsycNET, please contact psycinfo@apa.org.

Sarah obeyed the prompt and was then contacted by an APA analyst, who asked her to sign a restrictive legal agreement before downloading any more articles. Alarmed by the terms of a contract that she didn’t understand, Sarah contacted library staff for help. That triggered a months-long challenge to key issues around user rights, scholarly piracy, licensing and the Brock Library systems.

* Laurie Morrison is Associate University Librarian (Interim) at Brock University, lmorrison@brocku.ca; Carol Stephenson is Licensing Coordinator at Council of Prairie and Pacific University Libraries, carol@coppul.ca; Elizabeth Yates is Liaison/Scholarly Communication Librarian at Brock University, eyates@brocku.ca.

† The student’s name has been changed for privacy reasons.
After Sarah reached out, her liaison librarian followed up with several library departments related to collections, licensing and copyright, and to the university’s copyright and legal advisor. The problem became a hot potato, requiring diligence by the library to work toward a resolution for the student. It was a novel situation which exposed weaknesses internally for the library, and the university. Externally, the Ontario Council of University Libraries (OCUL)—a university library consortium which licenses PsycINFO to provincial libraries including Brock—assisted by contacting APA representatives. OCUL’s position was that Sarah’s objective was covered by terms of the OCUL APA license, which states:

Consistent with the Fair Use provisions of Section 107 of the U.S. Copyright Act or all applicable copyright law and fair use conventions in Canada, and specific provisions in this Agreement, Licensee and its Member Sites may display, print, download, extract or use a reasonable amount of content contained in the Licensed Materials for educational, scientific, or research purposes, including extraction and manipulation of information for the purpose of illustration, explanation, example, comment, criticism, teaching, research, or analysis.

However, APA maintained its stance that Sarah would have to sign a contract before mass downloading articles for her meta-analysis. This prompted a conversation between Brock’s copyright and legal advisor, the liaison librarian, an associate university librarian, and a lawyer for the APA. Brock reiterated that the student’s uses were covered by the “reasonable amount” stated in the OCUL APA license and that was it inappropriate for APA to be in direct contact with students. The lawyer stated the association’s position that it had to block mass downloading as a defense against piracy. The APA refused to describe what behavior looks suspicious to them or how they determine a potential piracy risk, as they are extremely worried about staying one step ahead of pirates.

The conversation ended with an agreement that:

1. The Library should be contacted when APA is concerned about downloading and that users should not be directly contacted
2. The Library and the APA would collaborate on the terms of an agreement to be signed by users who want to download large quantities of content

However, neither of these agreed-upon goals came to fruition. And at least one more Brock user has since been blocked by the APA.

While this was an internal wake-up call for Brock Library staff, one student’s crisis has ramifications for all academic libraries and their users.

**Pirates and Peer-To-Peer Sharing**

The underlying motivation for APA to block a Brock student stems from a very real fear publishers are facing: large-scale piracy of licensed content. Before addressing the most recent developments to attack a near monopoly of academic publishing, we will a look at how we got here.

The advent of distributed networked communication technology has brought massive disruption in many industry sectors. The now-defunct Napster, for example, was among the first sites to challenge the entertainment industry by providing bootlegged copies of music through peer-to-peer (P2P) file sharing. One can argue that this, as well as others such developments in the entertainment industry (eg. Piratebay, Kodi), reflect consumers’ reactions to what they perceived as egregious business practices and outrageous profits. Sound familiar? Indeed, the enormous profit margins in scholarly publishing have sparked similar ire among consumers including researchers and their institutional libraries. Illegal article-sharing sites Sci-Hub and LibGen have emerged as the Napster or PirateBay of the academic publishing industry.

Piracy in the entertainment industry is a logical comparison to recent developments in academic publishing. Technology has provided the tools for P2P file sharing to circumvent a proprietary industry. Because of
these events, some entertainment companies responded with innovative and successful new business models while others tried to litigate their hold on an outdated (though hugely profitable) business model. However, the major difference between the entertainment industry and scholarly publishing centers on the notion of public good. The creators of music, movies and other forms of entertainment (individual or commercial entities) are compensated for their work by consumer royalties protected by copyright. In contrast, scholars generally sign away their copyright when publishing journal articles. They are not directly compensated by consumers, while the publishers reap the financial rewards for scholars’ published research. The intellectual property created by researchers, held behind paywalls, is a public good and some might argue that it is an act of civil disobedience to set it free. As Willinsky, founder and director of the Public Knowledge Project puts it, “... research represents a different order of intellectual property. This is not only because of the public and tax-exempt funding involved in its production and publication. It is because of how this work’s value and benefit is realised through others’ access to and use of this work.”

Before technology afforded the current sharing economy, academic products (articles, books etc.) were shared through informal social networks of peers, passing along photocopied articles sent by mail or fax. Among the first channels to demonstrate the potential of social media was Twitter. In January 2011, scientist Andrea Kuszewski is credited with morphing the popular “I Can Haz Cheezburger” meme when she tweeted “OMG, that should be a new ‘I’m requesting a paper’ hashtag! #icanhazPDF...” The hashtag is used most heavily in the life sciences and biomedicine and largely by anglophones. Through Reddit Scholar, created in 2009, users post requests to a forum where other users download content and provide it either via non-paywalled link or cloud storage. At least four Facebook groups are devoted to sharing articles via either Facebook messaging or posting copy in comments on an original post.

While sharing articles is a longstanding practice among scholars, pirating articles and building open repositories of licensed content is a new phenomenon with a massive scale and scope. In 2015, Library Genesis (LibGen) contained over 25 million documents totaling more than 42 terabytes of data. Most of LibGen’s content is from Elsevier, Springer, and Wiley journals.

The Disrupter

At the age of 22, Alexandra Elbakyan, a neuroscience Phd candidate and native of Kazakhstan, created the website Sci-Hub which now hosts an estimated 50 million papers. After studying at institutions in the United States and Germany, where she had easy access to realms of subscription-based scientific literature, Elbakyan experienced frustration upon returning to Kazakhstan. She then began circumventing paywalls to download articles she and other researchers needed. Elbakyan easily automated the piracy process, made the content available to anyone, and thus created Sci-Hub. At its core, Sci-Hub is a search interface for known items only, offering greater simplicity than even Google, providing access to scientific academic literature. Sci-Hub bypasses paywalls and institutional proxy servers by (allegedly) using compromised user authentication credentials. The assumption is that credentials are either donated by users who support Sci-Hub or, though unproven and denied by Elbakyan, are obtained by phishing. Regardless of the method, once Sci-Hub gets past a paywall, copyright material is downloaded and stored for future use. In 2013, Sci-Hub began working with LibGen, a repository hosted in Russia which was also providing access to licensed content.

Elbakyan sees the site as a natural extension of her dream of helping humans share good ideas. “Journal paywalls are an example of something that works in the reverse direction,” she says, “making communication less open and efficient.” There is a philosophical ideal underpinning Elbakyan’s motives: “knowledge should be common and not intellectual property; research articles are used for communication in science: communication
implies common ownership. As a devout pirate, I think that copyright should be abolished.” Indeed, Elbakyan points to Article 27 of the United Nations’ Universal Declaration of Human Rights, which states “Everyone has the right to freely participate in the cultural life of the community, the enjoy the arts and to share in the scientific advancement and its benefits.” By rooting illegal pirating in a social justice narrative, it is easy to see why Elbakyan might be referred to as the “Robin Hood of Science.”

Publishers, however, are unlikely to take this perspective as they witness large swaths of their content being illegally downloaded at a rate of thousands per day. They might have been willing to look the other way with informal P2P sharing given the small scale, but Sci-Hub has kicked this up into unequivocal theft. Elbakyan found a way to transform what was informal and decentralized (Twitter, Reddit etc.) and to an effective centralized system. Publishers can no longer ignore a threat to the very foundation of their business practice.

Not surprisingly, in 2015 Elsevier and Wiley launched a lawsuit against both Sci-Hub and LibGen, citing copyright infringement. New York District Court Judge Robert Sweet supported the publishers’ claim that Sci-Hub infringes on their rights and that the Sci-Hub website be taken down. Subsequently, the site was promptly remounted with a new domain. Legal recourse for publishers is beyond the reach of the US judicial system as long as Sci-Hub servers holding all the illegally obtained content are in Russia. In the end, Sci-Hub continues to serve up millions of pirated articles daily. Given that the legal action failed to stop piracy, the lawsuit may, in the end, have served Sci-Hub’s agenda better than that of the publishers. The case’s publicity garnered more attention for Sci-Hub and the challenges surrounding academic publishing than any one individual person could have gained. Elbakyan was ready to seize the moment. In addition to being technologically capable, Elbakyan is quite clever in using communication avenues to promote her cause. For example, she very savvily took the lawsuit as an opportunity to leverage her mission. This included writing a letter to Judge Sweet, articulating her argument that knowledge should be free and that she has received no complaints from authors or researchers.

In April 2016, Elbakyan provided Science writer, John Bohannon, with Sci-Hub server data for a six-month period between September 1, 2015 and February 28th, 2015. The data revealed that Sci-Hub provided 28 million downloads to users around globe, including those who could legally access the same content via their institutions. Notably, “some of the most intense use of Sci-Hub appears to be happening on the campuses of U.S. and European universities.” In fact, the US is the fifth largest downloader of Sci-Hub content. While the publishers were successful with the lawsuit, in reality, the US courts have no jurisdiction over Sci-Hub’s servers that reside in Russia. One could argue that it only served to increase Sci-Hub’s profile, bringing it to the attention of more potential users.

As 2017 began, Sci-Hub operator’s notoriety was given another boost: Nature named Elbakyan one of the Top 10 people who mattered in 2016. “Few people support the fact that she acted illegally, but many see Sci-Hub as advancing the cause of the open-access movement, which holds that papers should be made (legally) free to read and reuse.”

Implications for Libraries
Who is winning from Sci-Hub and other instances of scholarly piracy? Certainly, researchers can celebrate short-term gains from quickly getting an article they might otherwise have either had no or delayed access in obtaining. The research itself - and its authors - will benefit from greater exposure via illegal access systems: while piracy is not open access per se, it could trigger similarly positive increases in readership and citations for illegally shared articles. And perhaps there will be a major, long-term win if the issue of scholarly piracy succeeds in reshaping a broken publishing ecosphere into a system that is fair, sustainable, and easy to navigate. But it is not clear how or when those changes will ever occur or how big a part piracy could play in a potential transformation of scholarly publishing.
How should libraries respond to piracy? Librarians are in a unique position as the major market for scholarly publishers; the traditional gateway for access and consumer use of the content; and primary critics of the currently dysfunctional system of pricing, organization, and distribution of scholarly communication. The authors suggest three foci for addressing piracy issues: 1) Educate Ourselves and Our Users, 2) Make Content Easy to Find and Use, 3) Support a Diverse and Equitable Publishing Ecosystem.

**Educate Ourselves and Our Users**

*Openly talk about the issues.* Librarians need to be informed and able to inform our users about the dramatic implications of piracy for scholarly communication, without fear of repercussions. When Gabriel J. Gardner praised the beauty and simplicity of Sci-Hub at an American Library Association panel and said “You should all try it” in the context of showing library systems “what they’re up against”, the Association of American Publishers complained to Gardner’s employer, the Dean of Library Services at California State University Long Beach, about Gardner’s comments. Gardner’s Dean defended him and clarified that the librarian had not been condoning Sci-Hub.17 Russell and Sanchez state that libraries should take action through advocating for open access with organizations including ALA, ACRL, and SPARC and international organizations such as the World Summit on the Information Society (WSIS) and the World Intellectual Property Organization’s (WIPO) Development Agenda.18

Sci-Hub is not Open Access. Sci-Hub’s website proclaims three main ideas: knowledge for all; no copyright; and open access. However, this is a misuse of the term “open access”, which refers to research being liberated by its creators and not stolen by pirates. Librarians should be concerned that Sci-Hub clouds our users’ views of open access.19 Of 11,000 researchers surveyed by Science, 88% considered it ‘not wrong’ to use Sci-Hub.20 Users need to be aware that piracy does more than affect the profit margins of major publishers: illegal downloading has the potential to harm scholarly societies, not-for-profit publishers, and university presses which may suffer from lost subscription revenues. Piracy may end up causing the demise of publishers that support advocacy for science, science diplomacy, and science education.21 But these conflicts create opportunities to re-educate our users about legitimate paths to open access and to foster support for alternative publishers. As a counterpoint to pirate repositories, libraries can promote institutional/subject repositories, preprint servers and other legitimate online archives which, to this point, have met with limited enthusiasm from faculty.22 Liaison librarians should be competent in scholarly communication issues, able to educate faculty about author copyrights in publisher agreements and publishing in “good” open access non-predatory journals. Librarians can actively advocate for open access journals to be recognized as legitimate publications which count towards tenure and promotion.

Credential sharing is much more than content sharing. Library users probably don’t realize that sharing their login credentials constitutes “illegal activity” and violates the terms of institutional licenses for publishers’ content.23 When users share library credentials with people not affiliated with their institutions, they potentially open online gateways to other university services such as campus email and course management systems. Personal sensitive personal information such as income tax forms, students’ grades or job applications could also be at risk.24

**Make Content Easy to Find and Use**

*Remove barriers to access.* Sci-Hub’s success in providing easier access to scholarly content should be viewed as a failure of both the library and the publisher communities. Bohannon’s analysis, cited earlier, showed that many Sci-Hub users had library access but chose to use Sci-Hub instead, citing speed and ease of use as benefits.25 Gardner and Gardner also found that speed and ease of access were primary motivators for scholars using social
media and pirated sites rather than interlibrary loan to obtain articles.\textsuperscript{26} For academic libraries “this is a good moment … to step back to reconfirm (or reconsider) their vision for discovery, to ensure that their visions connect with information-seeking practices and preferences, and to determine whether they have a viable strategy in place.”\textsuperscript{27}

**Relax licensing controls.** Most research articles are still controlled by the restrictions of publishers’ licenses. It is critical that librarians understand licensing clauses which could restrict the usability of content. Libraries must ensure they do not sign licenses which impose barriers to use. “Access is not just about removing the price to the user, but about allowing the user to do work, dissemination, augmentation, analysis with the content, legally.”\textsuperscript{28} We suggest the following clauses be examined closely:

- **Nondisclosure clauses:** License clauses should not be confidential. Nondisclosure licenses inhibit librarians from sharing and working collaboratively. Libraries shouldn’t be in competition with each other.
- **Liability clauses:** While Libraries must take reasonable measures to inform users how they can/cannot use licensed content, publishers should not hold libraries’ responsible for users’ copyright contraventions.
- **Authorized users:** Some publishers are now restricting types of users covered by licenses, removing distance education students, walk-in users, and multiple-campus users (without additional payment).
- **Interlibrary loan:** Licenses should not remove any rights granted under national copyright laws and the traditions of worldwide interlibrary loan. New technologies used by libraries, such as patron-direct web delivery, should be permitted.
- **Scholarly sharing:** Most faculty are unaware that some licenses restrict or forbid peer-to-peer single article sharing. Removing the right for peers to legitimately share content will not prevent piracy.
- **Scholarly use:** In the Brock University example, APA did not consider downloading articles for systematic reviews or meta-analyses as valid scholarly uses within the context of the library license and required that the student sign an additional license for this purpose. Recognized scholarly uses need to be clearly stated to ensure access for current patterns of scholarship. Text and data mining is another area that needs to be covered by library licenses. Licenses must ensure that publishers cannot shut down entire library sites over concerns about downloading. Publishers must work with libraries to resolve potential threats.
- **User privacy and confidentiality:** At Brock, APA contacted a student directly without the library’s knowledge. Licenses must clearly state that the library is the contact point for licensed content and that publisher concerns over use of that content will be communicated to the library first. Licenses should clearly state that publishers must not directly contact users.

**Support a Diverse and Equitable Publishing Ecosystem**

There is consensus that scholarly publishing is broken. Journal publishers collectively earned $10 billion in 2016, much of that revenue coming from ever-dwindling library coffers. Profit margins for the big five academic publishers are consistently over 30%—higher than giant consumer corporations including Apple and Disney.\textsuperscript{29} On the open access front, 15 years have passed since the Budapest Initiative set lofty goals for publicly accessible research.\textsuperscript{30} Support for legal open access has grown slowly but steadily: efforts “to wrestle the knowledge of the world from the hands of the publishers, one article at a time,” has resulted in about 27 million (24%) of about 114 million English-language articles becoming publicly accessible by 2014.\textsuperscript{31} In contrast, Elbakyan single-handedly provided “free” access to 48 million articles since 2011.\textsuperscript{32}
There is some hope that piracy could spur the publishing industry to adopt new—less usurious—business models. Librarians are already developing and promoting alternative publishing models, including SCOAP 3, Knowledge Unlatched, the Public Knowledge Project and the Library Publishing Coalition. One of the more promising models is offered by Open Library of the Humanities, an open access humanities mega-journal and growing stable of disciplinary journals, all supported by pooled funds from member libraries. Preprint servers and subject repositories also provide positive potential for open access. While preprint sharing has been a long-standing cultural norm in math and physics (e.g. ArXiv) and some branches of social sciences including economics (e.g. SSRN), other disciplines are starting to embrace this practice and/or develop new sharing hubs such as SocArXiv and bioRXiv. Murphy suggests that wider use of bioRXiv could potentially disrupt the death grip of flagship journals such as Nature and Science on communicating major scientific discoveries. While humanities scholars have been slow to participate in open scholarship, the recent launch of MLA Commons is a hopeful sign that attitudes are changing. Unlike Sci-Hub, these legitimate venues for sharing and accessing scholarship offer the unique benefit of building scholarly and career networks with colleagues. Librarians, who have a broad perspective on the scholarly communication ecosystem, are well-positioned to highlight new developments and educate their researchers and campus communities about publishing issues. And certainly, many faculty members are pioneering new methods of shareable scholarship—ranging from open lab notebooks to publication credits for peer review. Lasting transformational change, however, may require champions from within the ranks of senior university administrators and prominent researchers before others will follow suit.

Conclusion
Currently, more questions than answers can be found when pondering the issues of scholarly piracy. While publishers are currently in fight-back mode, it is unclear whether Sci-Hub’s growing popularity will eventually force them to adopt less restrictive and more open publishing models. Perhaps interest in innovative open access publishing/sharing models will render subscription publishing obsolete. In the meantime, the status of libraries as both promoters of open access and guardians of copyright law can force library staff to adopt contradictory and uncomfortable positions on piracy. Ultimately, we must prioritize our users’ rights to privacy and full access to licensed content for legitimate purposes. Back in January 2016, that drive to help a student pursue what should have been a straightforward quest—downloading articles for a meta-analysis—prompted a deep dive into these complex and murky issues. Sarah faced stress and frustration as Brock’s library attempted to resolve the downloading problem, but eventually found workarounds and spread her downloading over several weeks. “At the time, it was very problematic,” she said. “Are there not bigger issues that APA has to deal with?”

Notes


11. Ibid.


32. Ibid.

33. https://scoap3.org/

34. http://www.knowledgeunlatched.org/

35. https://pkp.sfu.ca/

36. https://www.librarypublishing.org/

37. https://www.openlibhums.org/

38. Murphy, “Should All Research Papers Be Free?”

39. Ibid.