Diminishing the perceived need for black open access

Tina Baich
University Library, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA
E-mail address: cbaich@iupui.edu

Copyright © 2017 by Tina Baich. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0/

Abstract:
The attention garnered by unauthorized sharing and pirating of scholarly content has resulted in a new category on the open access spectrum – black open access. Though black open access attempts to solve the discovery problem inherent in the multitude of open access content sources, it does so in violation of copyright law. Tools have now been developed to combat this same problem legally, including the Open Access Button and Unpaywall.

Librarians can engage in several strategies to help diminish the need for black open access, including the promotion of these discovery tools through education and services. We can share the tools with our users and teach them why they should not engage in unauthorized sharing. We can use the tools to fulfill requests and capture the benefits of open access in interlibrary loan. There are also more general strategies related to infrastructure, policy, and education that are important to acknowledge. Librarians can and must move the open access conversation forward in a positive, and legal, direction.

This paper provides an overview of the black open access landscape, discusses the discovery tools for uncovering legal open access content, and highlights how librarians can improve systems, services, and education efforts related to open access and open access discovery tools.

Keywords: resource sharing, interlibrary loan, open access, discovery, copyright, research sharing

Introduction
The attention garnered by unauthorized sharing and pirating of scholarly content has resulted in a new category on the open access spectrum – black open access. Though black open access attempts to solve the discovery problem inherent in the multitude of open access content sources, it does so in violation of copyright law and publisher/database terms of service. Tools have now been developed to combat this same problem legally, including the Open Access Button and Unpaywall.

Librarians can engage in several strategies to help diminish the need for black open access, including the promotion of these discovery tools through education and services. We can share the tools with our users and teach them why they should not engage in unauthorized sharing. We can use the tools to fulfill requests and capture the benefits of open access in interlibrary loan. There are also more general strategies related to infrastructure, policy, and education that are important to acknowledge. Librarians can and must move the open access conversation forward in a positive, and legal, direction.
Open Access Landscape

Open access literature is that which is “digital, online, free of charge, and free of most copyright and licensing restrictions” (Suber, 2004). While this sounds simple enough, open access has been parsed into numerous models since Peter Suber introduced this definition. The first two models, gold and green open access, are based on the method of distribution rather than user rights or degrees of openness. Gold open access refers to open access publishing. In other words, peer-reviewed articles published in journals that make all content immediately open access. Gold open access is typically subsidized through membership fees, author fees, grant funding, or institutional support. Articles made available through open access publishing are fairly easily surfaced through discovery tools such as Google Scholar or library link resolvers.

Green open access, on the other hand, refers to materials self-archived by authors in some sort of open access repository. Green open access expands the scope of open access to include more than articles. Authors can deposit grey literature such as theses, conference presentations, or white papers. This necessarily means that not all materials found in open access repositories have passed through a peer review process. Green open access is available to authors at no cost and is sometimes required by research funders. Because of the wide variety of repositories and repository aggregators available and varying levels of interoperability with Google Scholar, discovery methods for self-archived open access materials are numerous.

Some publishers take a hybrid approach where they offer a mix of immediate open access (gold) and subscription content either within a single journal or across their journal portfolio. Authors are typically required to pay a fee to make their article part of the publisher’s immediate open access content. These article processing charges (APCs) can be quite high (e.g. Elsevier fees range from $500 to $5,000 USD (Elsevier, 2017)). Some publishers using this hybrid approach also charge additional or higher fees to allow the use of a Creative Commons Attribution license on the work. The interspersing of immediate open access content within subscription journals complicates discovery as the searcher must proceed all the way to the article webpage to determine availability.

Unauthorized Sharing

The complex open access landscape, discovery issues, and either ignorance or disregard of copyright law has led to an increasing amount of unauthorized sharing over the years. Björk (2017) notes that progress towards complete open access “has been slower than hoped” with “somewhere between 35 and 50 percent of all research articles from recent years...found as either gold, hybrid, or self-archived repository copies” (p. 1). This slow adoption rate is perhaps the cause for the emergence of other simpler, yet legally and ethically questionable, sharing mechanisms, which Björk dubs “black open access” and others have called “guerilla open access” (Swartz, 2014; Bodó, 2016; Crissinger, 2017; Lawson, 2017).

The veritable explosion of unauthorized sharing via social networks and internet sites like Sci-Hub is a relatively recent phenomenon. In fact, a study of UK researchers published in 2010 fails to even mention social media as a vehicle for sharing articles. Instead, the focus was on Google Scholar and personal networks. One researcher is quoted as saying “Certainly a lot of the articles that I pick up in journals are through verbal face to face recommendations” (author’s emphasis)” (Procter et al., 2010: 4046). One year later, a study of the use of social media tools in the research process found that 27 percent of the active social media users surveyed (only 13 percent of all surveyed) used social networking sites; however, the study also found that such sites were used mostly to disseminate research findings. When asked about
discovery preferences, placing a general call for information on the internet or through another
electronic mechanism was the least favored method. The study’s authors concluded that “dissemination
and discovery of content through social media is now established as complementary to the well-
established dissemination and discovery services delivered by publishers and other vendors” (Rowlands,
et al., 2011). Even in 2013, Tenopir, Volentine, and King (2013) found that while UK researchers use
social media, “their use is more often occasional” (p. 198) and no mention is made of social media as a
discovery mechanism.

A current longitudinal study by CIBER Research provides interesting insight into the practices of early
career researchers with regard to the use of social media for unauthorized sharing. The study is
following 116 early career researchers from seven countries over the course of three years. In year one
(2016), the study found that ECRs “do use social media for communication and findings and passing
information around” and that “ResearchGate (14 mentions) and Twitter (8) are clearly the tools of
choice” (CIBER Research, 2016: 38). The year one report also notes that “Obtaining PDFs and connecting
with their colleagues are the main activities undertaken by ECRs on social network platforms” such as
ResearchGate (CIBER Research, 2016: 38). The interim 2017 report states that use of social media is only
growing among the study population, but, encouragingly, “ethical behavior is also becoming a matter of
concern” (CIBER Research, 2017: 2).

Though researchers in informatics and other fields began to take note of #icanhazpdf as a mode of
set the resource sharing world alight with their ACRL conference paper “Bypassing Interlibrary Loan via
Twitter: An Exploration of #icanhazpdf.” Around the same time Swab and Romme (2015) raised the
#icanhazpdf alarm in a poster at the Canadian Health Libraries Association. Both sets of authors have
continued their research into the Twitter hashtag specifically and crowdsourcing research sharing
generally (Swab & Romme, 2016; Gardner & Gardner, 2017). In 2016, the attention around
unauthorized sharing grew with Bohannon’s Science pieces (Bohannon, 2016a; Bohannon, 2016b) on the
illegal aggregator, Sci-Hub. Many authors, including myself, have begun to look more closely at Sci-Hub
and the seemingly unstoppable trend of unauthorized research sharing, or black open access (e.g.
Banks, 2016; Lewis, 2016b; Crissinger, 2017; Badke, 2017; and Novo & Onishi, 2017).

Black Open Access Landscape
Jennifer Herron’s predatory journal scale provides model that can be extended to describe the various
platforms through which unauthorized sharing, or black open access, is conducted. She divides
predatory journals into three categories: the “oblivious offender,” the “phisher,” and the “hijacker”
(Herron, 2017). Sources of black open access materials can be similarly categorized. The following is not
a comprehensive survey of unauthorized sharing platforms; it is instead meant to be representative.

Herron (2017) defines “the oblivious offender” as one who may be legitimate, but is oblivious to best
practices or rules. The oblivious offender who shares their work in an unauthorized manner likely does
not intend to violate author agreements or copyright law. They merely want to disseminate their work
to a broad audience. The unintentional sharer uses academic social networks such as Academia.edu or
ResearchGate.net to share their own research with their peers without realizing they signed away their
right to do so when they published. Academic researchers are enticed to post their works here by the
sheer number of their colleagues already present on these platforms, which also offer their own citation
issue takedown notices for content that they find illegally posted on academic social networks. Björk
(2017) notes an instance described by Jennifer Howard (2013) in the Chronicle of Higher Education
where Elsevier issued such notices to Academia.edu. A more recent incident saw the American Psychological Association (APA) sending takedown notices to all sorts of unauthorized sharing sites as well as directly to universities and authors who had posted content in violation of copyright (Mika, 2017). Following negative reactions from authors, the APA issued a news release stating they would refocus their efforts on piracy sites rather than targeting individual authors for unauthorized sharing of content from their journals (Mills, 2017).

The “phisher” takes unauthorized sharing further by requesting or supplying someone else’s work through means that violate copyright law. An example of such a phishing effort is using #icanhazpdf on Twitter to obtain copies of materials normally only available with a subscription. Individuals using #icanhazpdf to share such materials are likely aware they are circumventing paywalls and violating copyright notices based on the normal #icanhazpdf protocol described by Gardner and Gardner (2015).

First, a requestor tweets a link or partial citation to a pay-walled article with the hashtag #icanhazPDF and their e-mail address. Second, sympathetic users then use their institutional subscriptions or personal memberships to download the desired PDF and email it to the requestor, off of Twitter. Once in possession of the desired PDF, diligent requestors delete their tweet containing the original request. Thanking a user who fulfills the request is discouraged (Gardner & Gardner, 2015: 96).

However, the scale of sharing in this way is small and unsystematic in comparison to the “hijacker.” Various researchers have found relatively small #icanhazpdf samples for their studies. Gardner and Gardner harvested 824 tweets over six months for use in their study; Swab and Romme found 302 requests for health sciences literature in a three-month period; and Liu found 1,314 #icanhazpdf tweets over the course of a year (Gardner & Gardner, 2015; Swab & Romme, 2016; Liu, 2013).

The “hijacker” engages in large scale illegal aggregation of content and may see themselves as doing a public service or as engaging in civil disobedience in the face of unethical publishing practices. The obvious example of this is Sci-Hub, the most recent unauthorized sharing mechanism to be in the spotlight. The site was created in 2011 by a Kazakhstani graduate student, Alexandra Elbakyan, to address her “frustration with the barriers that scientists face,” especially in developing parts of the world (Bohannon, 2016a). Elsevier sued Elbakyan and succeeded in procuring a judicial order for Sci-Hub to cease operations in October 2015 (Banks, 2016). However, Sci-Hub continues to operate as it moves from domain to domain (Banks, 2016). The data behind Bohannon’s (2016b) piece, “Who’s Downloading Pirated Papers? Everyone,” reveals the truth of that title. Server log data from September 2015 through February 2016, a mere six months, show “28 million download requests, from all regions of the world and covering most scientific disciplines,” though “Sci-Hub users concentrate where academic researchers are working” in the U.S. And Europe (Bohannon, 2016b). A recent study (Himmelstein, et al., 2017) found that Sci-Hub contains 68.9 percent of all scholarly articles, which translates into more than 56.2 million articles. The study also estimates that Sci-Hub provided access to 99.3 percent of valid incoming requests during a six-month period in 2015-2016. Hoy (2017) concisely describes the process of using Sci-Hub.

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 symbiotic 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 downloaded and can then be shared over and over again (p. 74).

It may be useful to keep this scale in mind when providing research support to library users.

**Discovery Problems and Solutions**

The varying legal open access models and the multitude of corresponding discovery options present obstacles to users. Library users want ease of access (Connaway, et al., 2011), and while access to these materials is open, their discovery is not always straightforward. Many users begin with Google and while repositories are improving in their ability to play well with search engines, Google is not a comprehensive discovery mechanism for green open access. Yet when “information consumers” were asked by OCLC Research where they begin their information search, 84 percent indicated beginning in a search engine while not a single person began their search on a library website (De Rosa, et al., 2010). As Kroll and Forsman note, “researchers find Google and Google Scholar to be amazingly effective in finding isolated bits of information or getting to publications or findings of interest to them” (Kroll and Forsman, 2010). As a result, users are unlikely to search multiple resources for the information they seek both out of convenience and the possible perception that what they seek has been found.

Users’ desire for convenience and ease of use have contributed to the rise of black open access. Users want to find content where they are and they are obviously on academic social networks. Academia.edu, for example, had “approximately 30 million registered users who have uploaded approximately 8.5 million articles” as of January 2016 (Niyazov, et al., 2016: 1). Studies of academic social networks as a whole indicate that use of ResearchGate is even more prevalent (e.g. CIBER Research, 2016; Muscanell & Utz, 2017). These academic social networks have also worked to enhance discoverability by pushing content to followers of people and/or subjects rather than relying on users to conduct their own searches (Niyazov, et al., 2016: 2). They are also willing to use other social media venues and one-click pirate sites such as Sci-Hub because of their convenience and ease of use. A Science survey found that “17% picked simple convenience as their top motive” for using Sci-Hub (Travis, 2016). Additionally, “37% of those who had obtained a pirated journal article through Sci-Hub or other means said they did have traditional forms of access” (Travis, 2016). For some, it is only after these unauthorized sharing methods fail that they turn to what users see as the more cumbersome options of interlibrary loan or even library subscriptions.

Unauthorized sharing is obviously a problematic solution for these discovery problems in that it pushes the bounds of legal and ethical behavior. However, it is unrealistic to expect the average user to search multiple channels for what they need. Those that haven’t turned to unauthorized sharing often turn to interlibrary loan to help them navigate the open access landscape. In a study of interlibrary loan requests placed between July 2011 and June 2013, the author (Baich, 2015) found an increase in the number of requests for materials available via open access while overall borrowing requests held steady.
Regardless of how library users currently obtain open access materials, libraries must work to make discovery and delivery easier for them.

In response to these discovery issues, two tools have been developed outside libraries to try to legally address the needs for simplified discovery of open access content, the Open Access Button and Unpaywall. The Open Access Button (OAB) was created by British researchers David Carroll and Joseph McArthur with the support of a team of international volunteers. Launched in November 2013, the OAB "tracks how often researchers hit pay walls and attempts to connect users with freely accessible copies of articles" (SPARC, 2015). OAB claims to use "all of the aggregated repositories in the world," specifically citing oaDOI, Share, CORE, OpenAIRE, Dissem.in, Europe PMC, and BASE on its website (Open Access Button, n.d.). Users of the OAB can either go directly to the website openaccessbutton.org or download the browser extension. When an OAB user hits a paywall, she can click the OAB to search for a legal open access copy. If no such copy is found, the user has the option to email the author with a request that they deposit their work in a repository. In March 2017, OAB announced that they have been working to integrate with library catalogs and interlibrary loan (ILL) systems and began looking for libraries to participate in discussions and pilots (Open Access Button, 2017).

At the same time OAB made their announcement about ILL integration, another browser extension intended to locate legal open access copies of scholarly articles launched. Developed by Impactstory, the Unpaywall extension displays a lock icon (color-coding optional) on the right side of the browser window indicating whether or not a gold or green open access copy is available. According to the Unpaywall FAQ (n.d.), the tool locates 65 to 85 percent of articles. If it is unable to locate full text, a gray lock appears. Unpaywall gathers content from numerous sources, including PubMed Central, DOAJ, Crossref, DataCite, Google Scholar, and BASE, and then makes that data available for reuse via the oaDOI API, another Impact Story project (Unpaywall, n.d.). The extension relies on accurate DOI information to make a match.

Both efforts are nonprofit and open source with the mission of improving access to scholarly research. The proximity of their recent announcements led to some coverage in the press, namely The Chronicle of Higher Education. While Jason Priem, one of Unpaywall’s founders, sees what they are doing as challenging the scholarly publishing status quo, others interviewed for the article are more circumspect as to how successful the challenge will be. For instance, Martin P. Eve, a University of London professor, is quoted as saying “Unpaywall is dependent upon the uptake of green OA. That is, it is only ever effective if an academic has deposited a copy of a paper in a repository. At present, there is no evidence that green OA leads to subscription cancellations” (McKenzie, 2017). While both OAB and Unpaywall locate gold open access articles as well, it is true that publisher embargoes prevent much of the most recent, and thus most wanted, content from being found by these tools.

Library Action Plan
Libraries cannot ignore the potential impact that the trend in unauthorized sharing could have on them. Scientist and savvy library user Adam Bond (2013) recognizes how this trend can hurt libraries and ultimately users by steering traffic away from interlibrary loan and existing journal and database subscriptions leading to a false impression of user wants and needs. It will be hard, if not impossible, for libraries to justify funding for services and materials that aren’t being used or for new resources if they do not have the usage data that is instead going to unauthorized sharing platforms. Additionally, mass copyright infringement will likely lead publishers to increase subscription prices to account for the revenue lost to unauthorized sharing and can result in the suspension of access if institutional user
accounts are involved in mass infringement (Russell & Sanchez, 2016). All of this adversely affects both libraries and their users.

There are, however, things that libraries can and should do to diminish the need for black open access. Libraries can work to improve everything from collection strategies and education efforts to discovery and delivery. Let’s first look at discovery and delivery. Libraries can start by utilizing the tools previously discussed, the Open Access Button (OAB) and Unpaywall, to improve discovery for their users and resource sharing staff. This should involve promoting the use of the browser extensions to library users as an alternative to black open access and working to integrate these tools into both the user and staff interfaces of our existing systems. The OAB has efforts underway to integrate with interlibrary loan (Open Access Button, 2017) and oaDOI, the engine behind Unpaywall, can already be integrated into the SFX link resolver (Piwowar, 2017). Librarians should be a part of the efforts to more fully integrate these tools in both discovery and delivery systems and encourage widespread adoption.

Mukhopadhyay (2017) has suggested a broader, and more controversial, framework for integrating open access discovery, including black (or guerrilla) open access, with existing library discovery systems. Creating a streamlined discovery experience is the only way libraries can compete with Google. However, while it is important to improve library discovery systems, it is equally, if not more, important for librarians to meet users where they are. Library content can be surfaced in Google Scholar searches now, and Zepheira is actively working with a variety of partners to more fully expose library data on the web (Zepheira, n.d.). We should support and participate in efforts to use linked data to push library content more fully into search engines rather than relying solely on library discovery systems.

At the same time, librarians have to think beyond technical infrastructure. Collection development policies are fundamental to libraries, yet the time has come to rethink our traditional collecting strategies. We cannot continue to decry high subscription prices and paying for content created by our own professors without taking action. The time has come to adjust our collecting strategy from a “just in case” to a “just in time” model. Subscribing to an endless stream of journals and databases that are rarely, if ever, used is not sustainable. Begin the transition and tell your users if they want it, you will get it. They shouldn’t worry about whether the getting involves purchasing or borrowing. If we truly want to change scholarly communication, we have to begin putting our money towards open access publishing rather than continuing to pay traditional subscription fees (see Lewis, 2016a for a fuller discussion of this concept).

Finally, user education is key. In order to successfully implement the infrastructure and policy changes described here, library users must understand why they should not engage in unauthorized sharing; why open access is an important information policy issue; and how to use library systems effectively. As you speak about unauthorized sharing and open access with your constituents, consider what frame those constituents would connect with most strongly. Will they see unauthorized sharing as a violation of the honor code? Will they see open access as a social justice issue? An answer to information inequity (Crissinger, 2017)? Will they only consider the benefit providing their work in an open access platform brings to their citation counts? Use the frame that will create the greatest level of support at your institution.

Instruction and scholarly communication librarians can also use unauthorized sharing platforms like Sci-Hub “as a case study for asking our faculty and students larger questions about responsibility and sustainable change” (Crissinger, 2017: 86). Real life examples will show that unauthorized sharing is “a topic ripe for conversations about ethics, technology, copyright, and inequality” (Crissinger, 2017: 86).
Using case studies of unauthorized sharing also helps librarians to address information literacy concepts such as “creating new knowledge” and “participating ethically” in communities as suggested by Crissinger (2017: 88). These more philosophical discussions are a complement to more practical education such as how to comply with funder’s public access mandates or how to create a data management plan.

In addition to advocating infrastructure enhancements, resource sharing librarians can play a role in user education. If using open access materials to fulfill requests is part of your borrowing workflow, you can craft notification messages that create a greater awareness of open access among your users and teach them how to search for open access materials in future. You can also use your website and request forms to convey information about open access and copyright. Ultimately though, I think it is most important for resource sharing librarians to advocate for better library systems that are intuitive and convenient for users. This will help us meet users where they are and will likely be more effective in the long run than educating from interlibrary loan.

Conclusion
This may seem to be an unrealistic vision, but I truly believe in our ability to improve our services in ways that make unauthorized sharing unnecessary. If we can harness the technological advances around us, we should be able to achieve a one button service for our users. Click “Get It” and it will come. Our goal should be to defy the assumptions of researchers like Björk (2017) who see interlibrary loan as an “archaic and slow method” of obtaining materials (p. 2). In the US, the Big Ten Academic Alliance has a similar vision and has issued a series of reports outlining what they see as the future of resource discovery and delivery (BTAA, 2016; BTAA, 2017). Locating and connecting similar efforts in other countries merits further research. I’d like to close with a quote from Crissinger (2017). “As librarians, we need to recognize that discussing Sci-Hub with our communities can provide an opportunity to promote our vision for the future of scholarly communication while intentionally, collaboratively building a future where Sci-Hub is no longer needed or relevant” (p. 86). That is true of all unauthorized sharing platforms. We cannot give in to a future of users pirating content and publishers demanding higher ransoms. We must seize the opportunities to improve our systems, services, and education efforts in order to diminish the need for black open access and bring our users to library-provided and true open access content.

References


Elsevier (2017) Open access journals. Available at: https://www.elsevier.com/about/open-science/open-access/open-access-journals (accessed 8 August 2017).


Lewis D.W (2016b) Will Sci-Hub kill the open access citation advantage and (at least for now) save toll access journals? Available at: https://scholarworks.iupui.edu/handle/1805/11161 (accessed 13 October 2016).


