Articles

Teresa Auch Schultz,
Elena Azadbakht,
Jonathan Bull, Rosalind
Bucy, and Jeremy Floyd

Assessing the Effectiveness of Open Access Finding Tools

ABSTRACT

The open access (OA) movement seeks to ensure that scholarly knowledge is available to anyone with internet access, but being available for free online is of little use if people cannot find open versions. A handful of tools have become available in recent years to help address this problem by searching for an open version of a document whenever a user hits a paywall. This project set out to study how effective four of these tools are when compared to each other and to Google Scholar, which has long been a source of finding OA versions. To do this, the project used Open Access Button, Unpaywall, Lazy Scholar, and Kopernio to search for open versions of 1,000 articles. Results show none of the tools found as many successful hits as Google Scholar, but two of the tools did register unique successful hits, indicating a benefit to incorporating them in searches for OA versions. Some of the tools also include additional features that can further benefit users in their search for accessible scholarly knowledge.

INTRODUCTION

The goal of open access (OA) is to ensure as many people as possible can read, use, and benefit from scholarly research without having to worry about paying to read and, in many cases, restrictions on reusing the works. However, OA scholarship helps few people if they cannot find it. This is especially problematic for green OA works, which are those that have been made open by being deposited in an open online repository even if they were published in a subscription-based journal. OpenDOAR reports more than 3,800 such repositories.[1] As users are unlikely to search each individual repository, an efficient search method is needed to find the OA items spread across so many locations.

In recent years, several browser extensions have been released that allow a user to search for an open version of an article while on a webpage for that article. The tools include:

- Lazy Scholar, a browser extension that searches Google Scholar, PubMed, EuropePMC, DOAI.io, and Dissem.in. It has extensions for both the Chrome and Firefox browsers.[2]
- Open Access Button, which uses both a website and a Chrome extension to search for OA versions.[3]
- Unpaywall, which also acts through a Chrome extension to search for open articles via the digital object identifier.[4]
- Kopernio, a browser extension that searches subject and institutional repositories and is owned by Clarivate Analytics. Kopernio has extensions for Chrome, Firefox, and Opera.[5]

Teresa Auch Schultz (teresas@unr.edu) is Social Sciences Librarian, University of Nevada, Reno.
Elena Azadbkaht (eazadbakht@unr.edu) is Health Sciences Librarian, University of Nevada, Reno.
Jonathan Bull (jon.bull@valpo.edu) is Scholarly Communications Librarian, Valparaiso University.
Rosalind Buch (rbucy@unr.edu) is Research & Instruction Librarian, University of Nevada, Reno.
Jeremy Floyd (jfloyd@unr.edu) is Metadata Librarian, University of Nevada, Reno.

Some of the tools offer other services, such as Open Access Button's ability to help the user email the author of an article if no open version is available, as well as integration with libraries' interlibrary loan workflows. Kopernio and Lazy Scholar offer to sync with a user's institutional library to see if an article is available through the library's collection.[6] Although other similar extensions might also exist, this article is focused on the four mentioned above based on the authors' knowledge of available OA finding tools at the time of the project.

LITERATURE REVIEW

As noted above, scholars have indicated for several years a need for reliable and user-friendly methods, systems, or tools that can help researchers find OA materials. Bosman et al. forwarded the idea of a scholarly commons—a set of principles, practices, and resources to enable research openness—that depends upon clear linkages between digital research objects.[7] Bulock notes that OA has "complicated" retrieval in that OA versions are often housed in various locations across the web, including institutional repositories (IRs), preprint servers, and personal websites.[8] There is no perfect search option or tool, although some have tried creating solutions, such as the Open Jericho project from Wayne State University, which is seeking to create an aggregator to search institutional repositories and eventually other sources as well.[9] However, this lack of a central search tool can lead to confusion among researchers.[10] Nicholas and colleagues found that their sample of early career scholars drawn from several countries relied heavily on Google and Google Scholar to find articles that interested them.[11] Many also turn to ResearchGate and other social media platforms and risk running afoul of copyright. The results of Ithaka S+R's 2015 survey of faculty in the United States reflect these findings to a certain extent, as variations exist between researchers in different disciplines.[12] A majority of the respondents also indicated an affinity for freely accessible materials. As more researchers become aware of and gravitate toward OA options, the efficacy of various discovery tools, such as the browser extensions evaluated in this study, will become even more pertinent.

Previous studies on the findability of OA scholarship have focused primarily on Google and Google Scholar.[13] A few have assessed tools such as OAIster, OpenDOAR, and PubMed Central.[14] Norris, Oppenheim, and Rowland sought a selection of articles using Google, Google Scholar, OAIster, and OpenDOAR.[15] While OAIster and OpenDOAR found just 14 percent of the articles' open versions, Google and Google Scholar combined managed to locate 86 percent. Jamali and Nabavi assessed Google Scholar's ability to retrieve the full text of scholarly publications and documented the major sources of the full-text versions (publisher websites, institutional repositories, ResearchGate, etc.).[16] Google Scholar was able to locate full-text versions of more than half (57.3 percent) of the items included in the study. Most recently, Martin-Martin et al. likewise used Google Scholar to gauge the availability of OA documents across different disciplines.[17] They found that roughly 54.6 percent of the scholarly content for which they searched was freely available, although only 23.1 percent of their sample were OA by virtue of the publisher.

As of yet, no known studies have systematically evaluated the growing selection of open access tools' efficiency and effectiveness at retrieving OA versions of articles. However, several scholars and journalists have reviewed these new tools, especially the more established Open Access Button and Unpaywall.[18] These reviews were mostly positive, even as some acknowledged that the tools are not a wholescale solution for locating OA publications. Despite pointing out these tools'

limitations, reviewers voiced their hope that the OA finding tools could help disrupt the traditional scholarly publishing industry.[19]

At least one study has used the Open Access Button to determine the green OA availability of journal articles. Emery used the tool as the first step to identify OA article versions and then searched individual institutional repositories, followed by Google Scholar as the final steps.[20] Emery found that 22 percent of the study sample was available as green OA but did not say what portion of that was found by the Open Access Button. Emery did note that the Open Access Button returned 17 false positives (six in which the tool took the user to the wrong article or other content, and 11 in which it took the user to a citation of the article with no full text available). She also found at least 38 cases of false-negative returns from the Open Access Button, or articles that were openly available that the tool failed to find. The study did not count open versions found on ResearchGate or Academia.edu.

METHODOLOGY

OA Finding Tools
This study compared the Chrome browser extensions for Google Scholar and four OA finding tools: Lazy Scholar, Unpaywall, Open Access Button, and Kopernio. Each extension was used while in the Chrome browser to search for open versions of the selected articles and the success of each extension in finding any free, full version was recorded. The authors did not track whether an article was licensed for reuse. For the four OA finding tools, the occurrences of false positives (e.g., the retrieval of an error page, a paywalled version, or the wrong article entirely) were also tracked. False positives were not tracked for Google Scholar, which does not purport to find only open versions of articles. Data collection occurred over a six-week period in October and November 2018.

The authors used Web of Science to identify the test articles (N=1,000) with the aim of selecting articles that would give the tools the best chance for finding a high number of open versions. Articles selected were published in 2015 and 2016. These years were selected in order to try to avoid embargoes that might have prevented articles being made open through deposit. The articles were selected from two disciplines: Applied Physics and Oncology, both of which have a large share in Web of Science and come from a broader discipline with a strong OA culture.[21]

Each comparison began with searching the Google Scholar extension by article DOI or title if a DOI was not available. All versions retrieved by Google Scholar were examined until an open version was located or until the retrieved versions were exhausted. The remaining OA tools were then tested from the webpage for the article record on the journal's website (if available). If no journal page was available, the article PDF page was tested. All data were recorded in a shared Google Sheet according to a data dictionary.

Searches for open versions of paywalled articles were performed away from the authors' universities to ensure the institutions' subscriptions to various journals did not impact the results. Authors were limited in the number of articles they could search each day as some tools blocked continued use, presumably over concerns of illegitimate web activity, after as few as 15 searches.

Study Limitations
This methodology might have missed open versions of articles, even using these five search tools. Although studies have found Google Scholar to be one of the most effective ways of searching for

open versions, Way has shown that it is not perfect.[22] Therefore, it is possible that this study undercounted the number of OA articles.

The study tested the ability of OA finding tools to locate open articles from a journal's main article page, not other possible webpages (e.g., the Google Scholar results page). This design may have limited the effectiveness of some tools, such as Kopernio, which appear to work well with some webpages but not others.

RESULTS

Overall, the tools found open versions for just less than half of the study sample (490), whereas they found no open versions for 510 articles.

Although Lazy Scholar, Unpaywall, Open Access Button, and Kopernio all found open versions, Google Scholar returned the most with 462 articles (94 percent of all articles with at least one open version). Open Access Button, Lazy Scholar, and Unpaywall all found a majority of the open articles (62 percent, 73 percent, and 67 percent, respectively); however, Kopernio found open versions for just 34 percent of the articles (see figure 1).
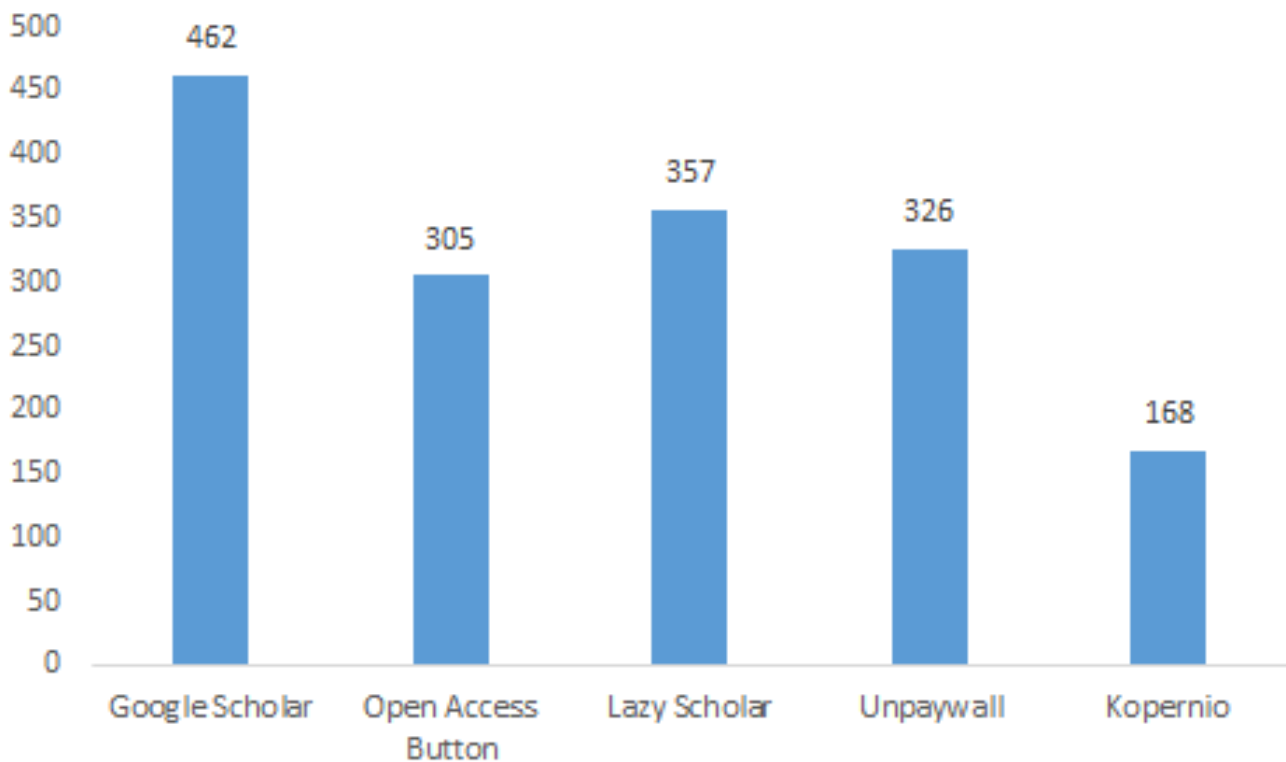

Figure 1. Number of open versions found by each tool.

It was most common for three or more of the tools to find an open version for an article, with just 48 found by two tools and 98 found by only one tool (see figure 2).
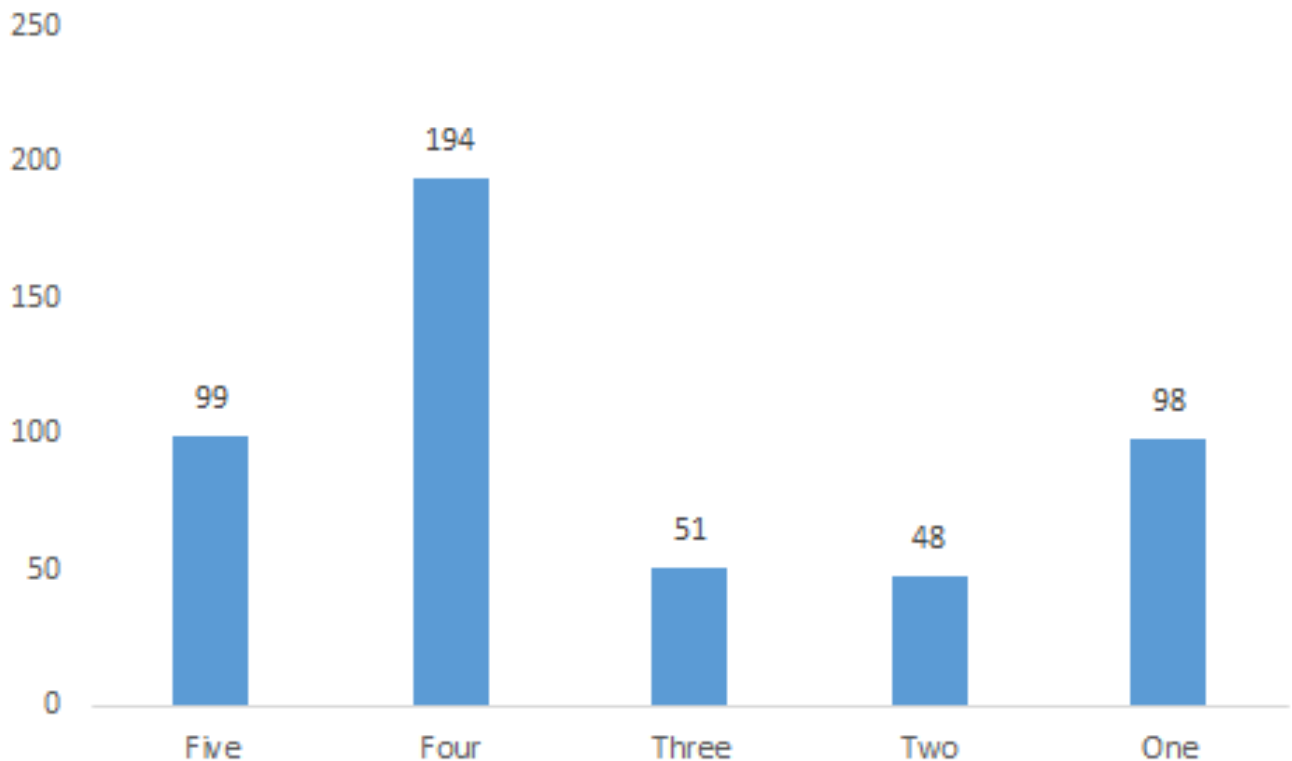
Figure 2. Number of articles where X number of OA finding tools found an open version.

When looking at articles where only one tool returned an open version, Google Scholar had the highest results (84). Open Access Button (4) and Lazy Scholar (10) also returned unique hits, but Unpaywall and Kopernio did not.

Open Access Button returned the most false positives with 46, or nearly 5 percent of all 1,000 articles. Lazy Scholar returned 31 false positives (3 percent), Unpaywall returned 14 (1 percent), and Kopernio returned 13 (1 percent).

DISCUSSION

The results for the OA search tools show that while all four options met with some success, none of them performed as well as Google Scholar. Three of the tools—Lazy Scholar, Open Access Button, and Unpaywall—did find at least half or more of the open versions that Google Scholar did. It is important to note that Open Access Button, which found the second fewest open versions, does not search ResearchGate and Academia.edu because of legal concerns over article versions that are likely infringing copyright.[23] This could have affected Open Access Button's performance.

Likewise, Kopernio's lower percentage of finding OA resources might relate to concerns over article versions as well. When creating an account on Kopernio, the user is asked to affiliate themselves with an institution so that the tool can search existing library subscriptions at that institution. For this study, the authors did not affiliate with their home institutions when setting up Kopernio to get a better idea of which content was open as opposed to content being accessible because of the tool connecting to a library's subscription collection. If the authors were to identify

with an institution, the number of accessible articles would likely increase, but this access would not be a true representation of what open content is discoverable.

In addition, some tools might work better with certain publishers than others. For instance, Kopernio did not appear to work with Spandidos Publications, a leading biomedical science publisher that publishes much of its content as gold OA, meaning the entire journal is published as OA. Kopernio found just one open version of a Spandidos article, compared to 153 by Google Scholar. This could be an unintentional malfunction either with Spandidos or Kopernio, which if fixed, could greatly increase the efficacy of this finding tool. However, Open Access Button, Lazy Scholar, Unpaywall, and Google were able to find OA publications from Spandidos at similar rates (135, 138, and 139, respectively) with no false positives.

While none of the tools performed as well as Google Scholar, some of the tools were easier to use compared to Google Scholar. Google Scholar does not automatically show an open version first; instead, users often have to first select the "All X Versions" option at the bottom of each record and then open each version until they find an open version. Lazy Scholar and Unpaywall appear (for the most part) automatically, meaning users can see right away if an open version is available and then click a button once to be taken to that version. Although Open Access Button and Kopernio do not show automatically if they have found an open version, users need to click a button on their toolbar once to activate each tool and see if the tool was able to find an open version. Open Access Button also provides the extra benefit of making it easy for users to email authors to make their works open if an open version is not already available. Relying on Lazy Scholar, Unpaywall, or Open Access Button first causes users no harm, and they can always rely on Google Scholar as a backup.

Whether all four tools are needed is questionable. For instance, a few of the authors found Kopernio difficult to work with as it seemed to be incompatible with at least one publisher's website and it introduced extra steps in downloading a PDF file. The fact that it also returned by far the fewest open versions—just 36 percent of the ones Google Scholar found and no unique hits—does not argue well for users to include it in their OA finding toolbox. Also, while Lazy Scholar, Unpaywall, and Open Access Button all performed better on their own, the authors wonder what improvements could be created by combining the resources of the individual tools.

CONCLUSION

The growth of OA finding tools is encouraging to see as far as helping to make OA works more discoverable. Although the study showed that Google Scholar uncovered more articles than any of the other tools, the utility of at least two of the tools—Lazy Scholar and Open Access Button—can still be seen in that both found articles not discovered by the other tools, including Google Scholar. Indeed, using the tools in conjunction with one another appears to be the best method. And although Open Access Button found the second fewest articles, the tool's effort to integrate with interlibrary loan and discovery workflows, as well as its concern about legal issues are all promising for its future. Likewise, Kopernio might be a better tool for those interested in combining access to a library collection—which likely has a large number of final, publisher versions of scholarship—with their search for openly available scholarship.

Future studies can include newer OA finding tools that have entered the market, as well as evaluate the user experience of the tools. Another study can also look at how well Open Access

Button's author email feature works. Also, as Open Access Button and Unpaywall continue to move into new areas, such as interlibrary loan support, research could explore if these are more effective ways of connecting users to OA material as well as measure users' understanding of OA versions they find.

Overall, the emergence of OA finding tools offers much potential for increasing the visibility of OA versions of scholarship, although no tool is perfect. However, if scholars wish to support OA through their research practices or find themselves unable to purchase or legally acquire the publisher's version, each of these tools can be valuable additions to their work.

DATA STATEMENT

The data used for this study has been shared publicly in the Zenodo database under a CC-BY 4.0 license at https://doi.org/10.5281/zenodo.2602200.

ENDNOTES

[1] Jisc, "Browse by Country and Region," accessed February 15, 2019, http://v2.sherpa.ac.uk/view/repository_by_country/countries=5Fby=5Fregion.html.

[2] Colby Vorland, "Extension," accessed March 14, 2019, http://www.lazyscholar.org/; Colby Vorland, "Data Sources," *Lazy Scholar* (blog), accessed March 14, 2019, http://www.lazyscholar.org/data-sources/.

[3] "Avoid Paywalls, Request Research," Open Access Button, accessed March 14, 2019, https://openaccessbutton.org/.

[4] Unpaywall, "Browser Extension," accessed March 14, 2019, https://unpaywall.org/products/extension.

[5] Kopernio, "FAQs," accessed March 14, 2019, https://kopernio.com/faq.

[6] Colby Vorland, "Features," *Lazy Scholar* (blog), accessed March 14, 2019, http://www.lazyscholar.org/category/features/.

[7] Jeroen Bosman et al., "The Scholarly Commons—Principles and Practices to Guide Research Communication," *Open Science Framework*, September 15, 2017, https://doi.org/10.17605/OSF.IO/6C2XT.

[8] Chris Bulock, "Delivering Open," *Serials Review* 43, no. 3–4 (October 2, 2017): 268–70, https://doi.org/10.1080/00987913.2017.1385128.

[9] Elliot Polak, email message to author, June 4, 2019.

[10] Bulock, "Delivering Open."

[11] David Nicholas et al., "Where and How Early Career Researchers Find Scholarly Information," *Learned Publishing* 30, no. 1 (January 1, 2017): 19–29, https://doi.org/10.1002/leap.1087.

[12] Christine Wolff, Alisa B Rod, and Roger C. Schonfeld, "Ithaka S+R US Faculty Survey 2015," 2015, 83, https://sr.ithaka.org/publications/ithaka-sr-us-faculty-survey-2015/.

[13] Mamiko Matsubayashi et al., "Status of Open Access in the Biomedical Field in 2005," *Journal of the Medical Library Association* 97, no. 1 (January 2009): 4–11, https://doi.org/10.3163/1536-5050.97.1.002; Michael Norris, Charles Oppenheim, and Fytton Rowland, "The Citation Advantage of Open-Access Articles," *Journal of the American Society for Information Science and Technology* 59, no. 12 (October 1, 2008): 1963–72, https://doi.org/10.1002/asi.20898; Doug Way, "The Open Access Availability of Library and Information Science Literature," *College & Research Libraries* 71, no. 4 (2010): 302–09; Charles Lyons and H. Austin Booth, "An Overview of Open Access in the Fields of Business and Management," *Journal of Business & Finance Librarianship* 16, no. 2 (March 31, 2011): 108–24, https://doi.org/10.1080/08963568.2011.554786; Hamid R. Jamali and Majid Nabavi, "Open Access and Sources of Full-Text Articles in Google Scholar in Different Subject Fields," *Scientometrics* 105, no. 3 (December 1, 2015): 1635–51, https://doi.org/10.1007/s11192-015-1642-2; Alberto Martín-Martín et al., "Evidence of Open Access of Scientific Publications in Google Scholar: A Large-Scale Analysis," *Journal of Informetrics* 12, no. 3 (August 1, 2018): 819–41, https://doi.org/10.1016/j.joi.2018.06.012.

[14] Norris, Oppenheim, and Rowland, "The Citation Advantage of Open-Access Articles"; Micahel Norris, Fytton Rowland, and Charles Oppenheim, "Finding Open Access Articles Using Google, Google Scholar, OAIster and OpenDOAR," *Online Information Review* 32, no. 6 (November 21, 2008): 709–15, https://doi.org/10.1108/14684520810923881; Maria-Francisca Abad-García, Aurora González-Teruel, and Javier González-Llinares, "Effectiveness of OpenAIRE, BASE, Recolecta, and Google Scholar at Finding Spanish Articles in Repositories," *Journal of the Association for Information Science and Technology* 69, no. 4 (April 1, 2018): 619–22, https://doi.org/10.1002/asi.23975.

[15] Norris, Rowland, and Oppenheim, "Finding Open Access Articles Using Google, Google Scholar, OAIster and OpenDOAR."

[16] Jamali and Nabavi, "Open Access and Sources of Full-Text Articles in Google Scholar in Different Subject Fields."

[17] Martín-Martín et al., "Evidence of Open Access of Scientific Publications in Google Scholar."

[18] Stephen Curry, "Push Button for Open Access," *The Guardian*, November 18, 2013, sec. Science, https://www.theguardian.com/science/2013/nov/18/open-access-button-push; Bonnie Swoger, "The Open Access Button: Discovering When and Where Researchers Hit Paywalls," *Scientific American Blog Network*, accessed May 30, 2017, https://blogs.scientificamerican.com/information-culture/the-open-access-button-discovering-when-and-where-researchers-hit-paywalls/; Lindsay Mckenzie, "How a Browser Extension Could Shake Up Academic Publishing," *Chronicle of Higher Education* 68, no. 33 (April 21, 2017): A29–A29; Joyce Valenza, "Unpaywall Frees Scholarly Content," *School Library Journal* 63, no. 5 (May 2017): 11–11; Barbara Quint, "Must Buy? Maybe Not," *Information Today* 34, no. 5 (June 2017): 17–17; Michaela D. Willi Hooper, "Product Review: Unpaywall [Chrome & Firefox Browser Extension]," *Journal of Librarianship & Scholarly Communication* 5

(January 2017): 1–3, https://doi.org/10.7710/2162-3309.2190; Terry Ballard, "Two New Services Aim to Improve Access to Scholarly Pdfs," *Information Today* 34, no. 9 (November 2017): Cover-29; Diana Kwon, "A Growing Open Access Toolbox," *The Scientist*, accessed December 11, 2017, https://www.the-scientist.com/?articles.view/articleNo/51048/title/A-Growing-Open-Access-Toolbox/; Kent Anderson, "The New Plugins — What Goals Are the Access Solutions Pursuing?," The Scholarly Kitchen, August 23, 2018, https://scholarlykitchen.sspnet.org/2018/08/23/new-plugins-kopernio-unpaywall-pursuing/.

[19] Curry, "Push Button for Open Access"; Swoger, "The Open Access Button"; Mckenzie, "How a Browser Extension Could Shake Up Academic Publishing"; Kwon, "A Growing Open Access Toolbox."

[20] Jill Emery, "How Green Is Our Valley?: Five-Year Study of Selected LIS Journals from Taylor & Francis for Green Deposit of Articles," *Insights* 31, no. 0 (June 20, 2018): 23, https://doi.org/10.1629/uksg.406.

[21] Anna Severin et al., "Discipline-Specific Open Access Publishing Practices and Barriers to Change: An Evidence-Based Review," *F1000Research* 7 (December 11, 2018): 1925, https://doi.org/10.12688/f1000research.17328.1.

[22] Way, "The Open Access Availability of Library and Information Science Literature."

[23] Open Access Button, "Open Access Button Library Service FAQs," Google Docs, accessed February 19, 2019, https://docs.google.com/document/d/1_HWKrYG7Qj7ff05-cx8Kw40mL7ExwRz6ks5Fb10GEGg/edit?usp=embed_facebook.