Bypassing Interlibrary Loan Via Twitter: An Exploration of #icanhazpdf Requests

Carolyn Caffrey Gardner and Gabriel J. Gardner

Introduction

Twitter has emerged as a popular social media platform for many scientists and scholars.

Priem and Costello estimate that one out of every forty scholars, defined as faculty, postdoc, or doctoral student, in the United States and the United Kingdom is a registered Twitter user.[1]

What's more, they are using the platform to share scholarly material. Moriano et al. have demonstrated that the number of tweets containing links to scholarly publications increased substantially from 2011 to 2013; similarly, they also increased as a percentage of overall twitter traffic.[2] Their analysis also noted that certain publishers' content is shared more than others. The top domain names tweeted in their sample were: nature.com, arxiv.org, sciencemag.org, wiley.com, and sciencedirect.com. While the interdisciplinary nature of these domains makes it difficult to draw conclusions about Twitter use rates by specific academic discipline, Priem and Costello concluded that no one academic discipline was significantly overrepresented on Twitter.[3]

Apart from everyday social use of the microblogging service, scholars are clearly using Twitter to increase their professional networks, organize prepublication review of working papers and manuscript drafts, offer post-publication critique,[4] disseminate published research,[5] and share pre-prints.[6]

Twitter is also used to facilitate access to scholarly articles that would otherwise be denied to users behind a paywall or obtained using interlibrary loan. Like peer-to-peer sharing in the music industry, this peer-to-peer access to scholarly material is ethically dubious, and may run afoul of copyright laws, but it is easy to accomplish. The Twitter user simply appends the metadata label, or "hashtag", #icanhazPDF in the tweet rendering it discoverable through traditional linking and search functions.

This peer-to-peer access, while ethically dubious, is coordinated by the use of the Twitter hashtag #icanhazPDF. The hashtag is included as a tag in the tweet, which renders it discoverable through traditional linking and search functions.

Literature Review

The modern interlibrary loan (ILL) office has been likened to that of a detective's office, assisting scholars of all stripes track down materials not held in the library. ILL offices can often find materials when only partial citation information is available.[7] In spite of alternatives and freely available content online, the average ILL request per ARL library has increased in 31 of the 35 years ARL has kept such statistics.[8] Users expect that these ILL requests will happen as easily and instantaneously as they do using popular interfaces, such as Google[9] and many ILL services have made strides, sometimes filling requests in as little as 24 hours. Also, as electronic journals have become an established part of the scholarly communications land-

Carolyn Caffrey Gardner is Information Literacy & Educational Technology Librarian, University of Southern California, e-mail: ccgardne@usc.edu; Gabriel J. Gardner is Librarian for Romance, German, Russian Languages & Literatures, and Linguistics, California State University, Long Beach, e-mail: gabriel.gardner@csulb.edu

scape, fewer lending agreements contain ILL restrictions. Lamoreux and Stemper found that at University of Minnesota, 89% of licenses allowed for lending and it was primarily small scholarly associations that restricted lending.[10] The scholarship practices that have resulted in increased ILL requests are bleeding over into non-library spaces.

Usage of the #icanhazPDF hashtag to facilitate the sharing of scholarly articles dates back to 2011 with the suggestion from Andrea Kuszewski on the hashtag language, a riff on a popular internet cat meme.[11] The hashtag and the social sharing networks it facilitates have received passing mentions in some medical literature[12,13] but little direct study. Dunn et al. characterize #icanhazPDF as a form of "guerrilla open access" through subversion of publisher agreements, in the tradition of internet activist, Aaron Swartz's, manifesto of the same name.[14,15] Since its inception, the hashtag has been a controversial topic of discussion on science blogs. Michael Eisen, co-founder of the Public Library of Science (PLoS), portrayed the use of the hashtag as an act of civil disobedience in opposition to the current copyright regime that governs scientific publishing.[16]

The sharing mechanics follow a simple protocol. 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.[17] This allows the fulfilling user, who likely violated a copyright or license agreement, to maintain anonymity.

The small amount of research on #icanhazPDF has focused on demographic data. In a blog post, Jean Liu collected tweets using the hashtag over the course of a year beginning in May 2012. Her analysis revealed that overall use of the hashtag slowly increased over the period of study to an average of 3.6 #icanhazPDF tweets per day.[18] Using #icanhazPDF

represents a small percentage of Twitter-user behavior compared to sharing scholarly research by sending links to pay-walled papers. In addition to initial quantitative analysis, Liu also sampled location and profile data. These were then used to determine occupation and country location of #icanhazPDF users. Usage was overwhelmingly an Anglophone phenomenon with the almost half of the tweets sampled coming from the United States; the United Kingdom produced the second highest number of tweets. Occupation data revealed that academics and students, despite being the most likely to have institutional access to scholarly research, were the most frequent users of the hashtag. Finally, Liu's category of communicators which encompassed journalists and bloggers, had the third highest use of #icanhazPDF.[19]

Though scholars and scientists have been the primary focus regarding #icanhazPDF, librarians have also taken note of the phenomenon and begun to grapple with how it might affect their institutions and workflows. Greenhill and Wiebrands argue that libraries should view #icanhazPDF and other copyright-violating (or license-breaching) methods of content sharing as competition and not ignore the black market transactions.[20] When such peer-to-peer access is viewed as a competitor, libraries are at a disadvantage because they must adhere to copyright and intellectual property laws, which may take more time and/or financial resources. To differentiate themselves from crowdsourced methods they might emphasize the local or niche content they provide, and the physical space they provide, while advocating for "more open and fair" publishing models.[21]

One apparent impact that #icanhazPDF sharing has on libraries, is in the area of interlibrary loan (ILL). Each request fulfilled through Twitter represents one side of a possible ILL transaction.[22] The institutions of the fulfilling users will record downloads of the requested files. Any libraries that requestors might have used however, are left without any record of user demand. Thus, #icanhazPDF and other methods of peer-to-peer sharing distort library use statistics: libraries serving users who fulfill requests

via #icanhazPDF have artificially inflated download statistics, while libraries whose (potential) users obtain articles over Twitter have artificially deflated ILL statistics. The magnitude of these errors is unknown and an area for further research. Therefore, Jill Emery, reacting to Liu and Bond, urged librarians involved with collection development and technical services to treat #icanhazPDF as an impetus to improve our services, specifically in the area of document delivery.[23] Users bypassing interlibrary loan, particularly students and professors who have institutional access, reveal their preferences for a different method of fulfillment that is simpler and often faster.[24] This study seeks to analyze our competitor and take a closer look at the prevalence of #icanhazPDF requests and analyze them in order to understand user demand and improve library services.

Methods

The deletion of the original requesting tweet after fulfillment makes gathering information difficult. To solve this issue of data collection, the authors explored several tweet archiving services and settled upon using Tweet Archivist (https://www.tweetarchivist.com/). Tweet archivist is relatively affordable, and most importantly, it captures tweets in real-time automatically by hashtag, solving the deletion conundrum. The authors activated Tweet Archivist and captured 1,238 publically available tweets using #icanhazPDF from the end of April through the beginning of August 2014 and did not include other less common variations of the hashtag (such as #icanhasPDF). Private tweets are only accessible to friends of that user, so while there are also likely private tweets using #icanhazPDF during this time frame, the authors did not have access to that data. Twitter Archivist captures the full text of the tweet, user name, geographic location, the number of followers the user has, time stamp, and language of the tweet.

Of the 1,238 tweets collected, 824 made a request for material either through partial citation information or a link to the original source. The remaining tweets ran the gamut from discussions over the ethics of #icanhazpdf to suggested rules for #icanhazpdf request structure. Of those 824 requests—74 were retweets by the original user or other accounts and were thus excluded from the pool for further analysis. The authors tracked down the full citation information for each item requested, and recorded it in a shared spreadsheet. While they made every attempt using the limited information available, 14 requests were unable to be fully captured. In most cases these requests were links with no other information, and the links were parsed through university proxy servers that the authors could not access. For Tweets in which the authors were able to determine the correct citation, they recorded the title of the material, journal title if applicable, publication date, publisher, and content format (journal article, book chapter, etc.).

While some users did not supply location information, 378 of the 475 users who had requested items had entered in an identifiable geographic location such as a city, state, province, or country associated with their Twitter profile. Many more users included location information that was facetious such as "the internet" or "everywhere" and these results were discarded from country analysis.

As Priem and Costello concluded that Twitter use was not dominated by any one discipline, the authors were interested in seeing if similar patterns exist among #icanhazPDF requests. Web of Science subject categories were chosen as a level of analysis because they are reviewed regularly by experts and are non-hierarchical.[25] Web of Science's classification of journals with multiple subject categories provides a clearer picture of what disciplines are represented in #icanhazPDF requests by mirroring the often interdisciplinary nature of scholarship. The authors searched for the journal titles in the Web of Science publication index, and if available collected the "Research Domain" information for the title. Since some journals were not indexed in Web of Science, the authors also searched Ulrichsweb.com and recorded the Ulrich's subject classifications. With only 4 months' worth of #icanhazPDF requests it was also necessary to group the disciplines into larger categories to get

a clear picture of the disciplines represented. The authors then categorized the collected subject categories into the four larger categories of: Life Sciences & Biomedicine, Physical Sciences, Technology, Arts & Humanities, and Social Sciences using the Research Areas chart in the Web of Science help pages.[26] Books, chapters, conference papers, and other miscellaneous items were not further analyzed by subject in either Ulrich's or Web of Science.

Results & Discussion

Who is Using icanhazPDF?

There were 475 unique users requesting items through #icanhazPDF during the data collection period. 80% (378) of users had a Twitter profile with an identifiable location. This sample is nearly 4 times larger than Liu's and confirms her results. The top two countries with requests were the United States and Great Britain, with other countries contributing marginally. Taking Canada and Australia into account demonstrates that #icanhazPDF is overwhelmingly an Anglophone phenomenon as indicated in Table 1.
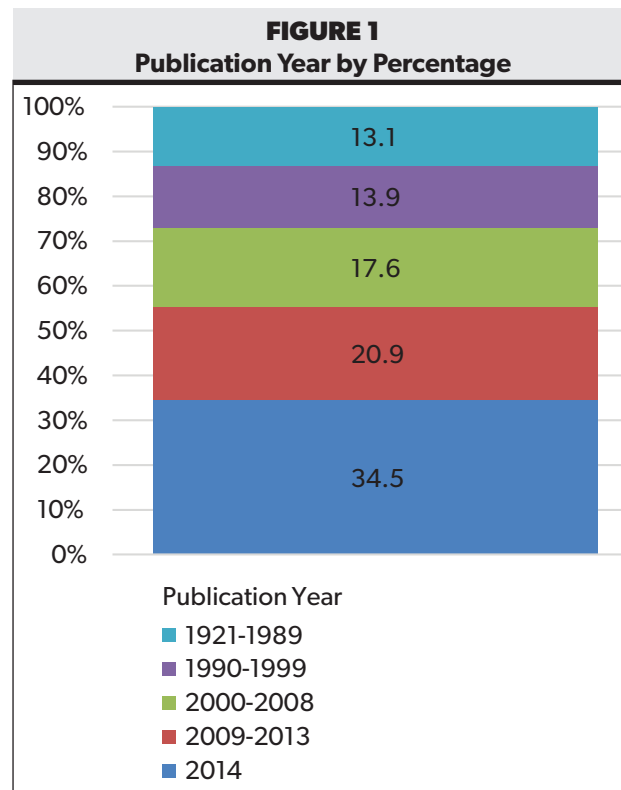
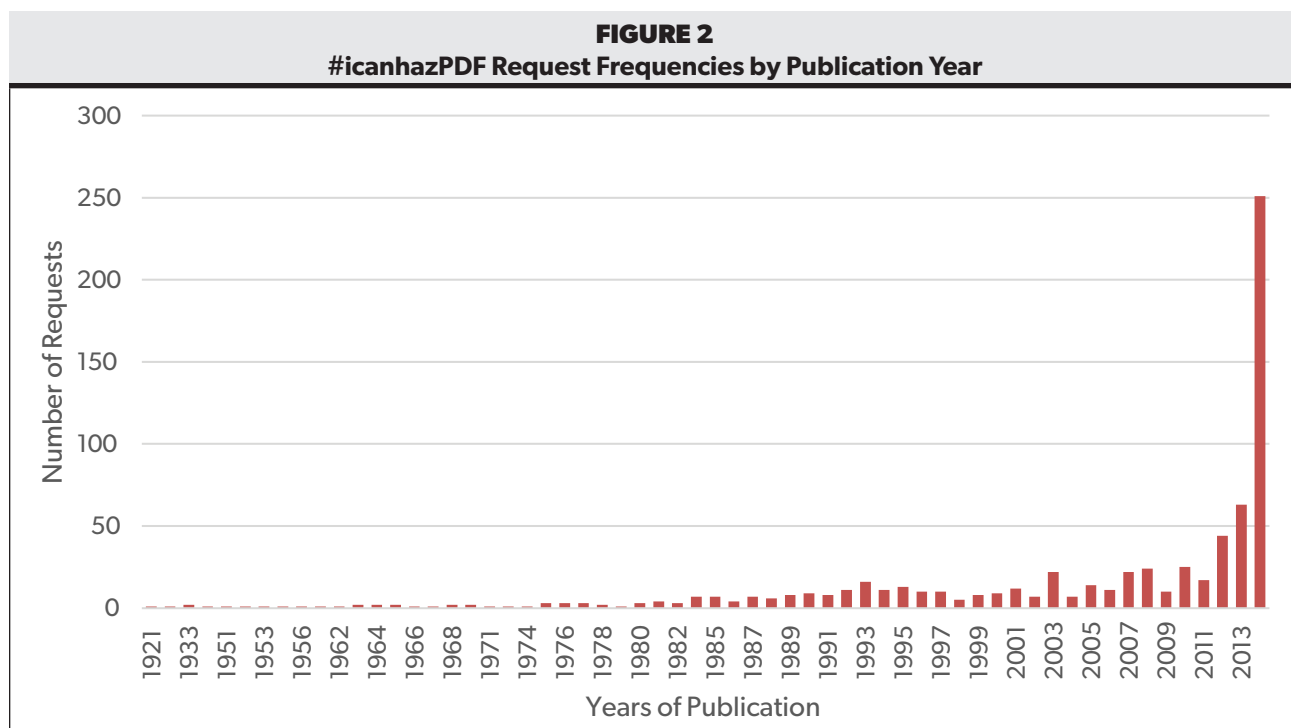| TABLE 1 #icanhazPDF Requests by Countries	
Countries	**Number of Requests**
United States	128
Great Britain	110
Germany	20
Canada	17
Australia	13
France	13
Netherlands	10
Brazil	9
Sweden	6
Chile	5

While many #icanhazPDF requests are filled by devoted followers of the hashtag, these requests reach much farther than one might expect. The mean follower count was 1,207, which does not account for when these tweets are retweeted to ever expanding networks.

Most users (76%) only used #icanhazPDF once during the data collection dates. However, there were some repeat users. Not counting retweets, the most items requested by any one user was 12. This suggests that for most users #icanhazPDF is not their primary method of access to scholarly material, but is instead used for those hard-to-access publications in a one-off request.

When Were the Requested Articles Published?

The authors hypothesized that #icanhazPDF requests might be originating because of publisher embargos for the most recent items at a user's institution. While the most requested publication date for items was the current year, 2014, there were more historical materials than anticipated. Items from the current year, 2014, only made up 34.5% of all requests with publication dates. Of further surprise, the past five years (2009-2014), only brings the percentage up to 56.4% of requests (Figure 1).

FIGURE 1
Publication Year by Percentage

Publication Year
- 1921-1989
- 1990-1999
- 2000-2008
- 2009-2013
- 2014

FIGURE 2
#icanhazPDF Request Frequencies by Publication Year

The earliest published item requested was from 1921, and item requests were scattered across the 20th century (Figure 2).

More research which takes into consideration other forms of "guerrilla open access", such as article sharing over email or discussion boards, is required before explicitly endorsing the embargo hypothesis. However, the data collected for this study suggest that embargos are not the only reason users turn to #icanhazPDF.

What is Being Requested?

The majority of requests, 89.86% (674) were for individual journal articles. The remaining 73 requests were for other materials, which included conference papers, book chapters, entire books, business reports, music scores, and ISO standards. The journal article requests came from 493 unique titles. While most journals only had one #icanhazPDF request, there were some outliers. *Nature* had the most requests, with 16 unique article titles; *Science* followed with 14, and *Proceedings of the Royal Society B: Biological Sciences* had 7. Large science-focused publishers were well represented in the requests. Table 2 identifies the 10 most represented publishers:

A majority of the articles requested, 87.4% (589), were indexed in Web of Science. There were 188 different research domains represented and 34% (78) of them occurred only once. Likewise, the majority of requests came for journals that fell within ISI's category for Life Sciences & Biomedicine journals (Figure 3). Conversely, the Arts & Humanities made up a

TABLE 2
#icanhazPDF Requests by Publisher

Top Publishers of Articles Requested	Number of Requests
John Wiley & Sons	91
Elsevier	83
Nature Publishing Group	61
Taylor & Francis	52
Springer	26
American Association for the Advancement of Science	19
Oxford University Press	17
Lippincott Williams & Wilkins	15
Royal Society Publishing	13
Sage Publications	12
Bentham Science	11

FIGURE 3
Journal Article Requests by Subject Category
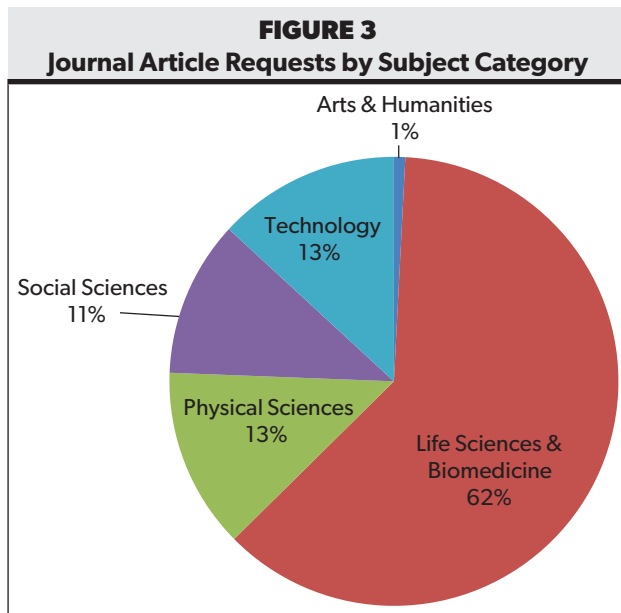
small sliver of requests. This may be in part because of #icanhazPDF being a journal article focused hashtag, which may not be as readily applicable to humanities researchers, which predominantly uses monographs. Many of the journals not indexed in Web of Science were listed in Ulrich's. Of the 674 unique requests for journal articles, 91.5% (617), were indexed in Ulrich's. However, analysis of Ulrich's Classifications yielded a similar request pattern to that identified by ISI category analysis.

While more research is needed, at initial glance the emphasis on life science publications says less about the price of academic journals in the discipline and more about who is using #icanhazPDF. According to the *Library Journal*'s Annual Periodical Price Survey 2014, the top three categories in terms of price are Chemistry, Physics, and Astronomy, which are all in the ISI Physical Sciences classification.[27] In the case of Chemistry, the average price per title is more than 180% more expensive than the average Biology journal. This may suggest which academic communities #icanhazPDF is gaining traction with as opposed to which subjects have prohibitive prices for institutions or individuals.

Conclusion

More research is required to understand the motivations and preferences of researchers, students, and

journalists who obtain articles using #icanhazPDF and other illicit crowdsourced methods. Additional analysis may also address how journal impact factors and rankings correspond to the items that are requested in particular disciplines.

Librarians who have spoken out about #icanhazPDF have largely decried its existence and reminded patrons of interlibrary loan services. By exploring what is being requested, the authors hope to point out that this is not an isolated phenomenon of a few users and is continuing to grow. Librarians would benefit by being proactive in both preventing the need for the paywalls and assisting users in becoming aware of how to search for free resources, especially when institutional access ends. Interlibrary loan services experience the same barriers to information access as users, including cost-prohibitive pay per use agreements.[28] Interlibrary loan librarians remind us that one focus on scholarly communication initiatives at universities should be to develop cross-library agreements for interlibrary loan prices so that we are not reliant on publishers for content access.[29] If these initiatives and agreements fail or receive insufficient funding, it is likely that librarians will continue to shoulder the burden of providing legal and ethical access to pay-walled content. Preparing students to be life-long information seekers, including accessing scholarly material after graduation, is an important role for library educators. To better prepare graduating students in particular for what employers are looking for, librarians need to encourage persistence and flexibility in searching as well as equipping our users with the knowledge to find open-access and freely available materials upon graduation.[30] #icanhazPDF is a symptom of a broken scholarly publishing system and of the complexity of many libraries' interlibrary loan interfaces. The current scholarly publishing system is so broken that some researchers are forced to make requests like "Still looking for a pdf of my own paper! Please help."[31] Most libraries are unlikely to be able to fulfill requests at the speed of a crowd of Twitter users. Until the system is fixed, or ILL systems are streamlined and given more resources, "guerrilla open access" efforts like #icanhazPDF will persist.

Notes

1. Jason Priem and Kaitlin Light Costello, "How and Why Scholars Cite on Twitter," *Proceedings of the American Society for Information Science and Technology* 47, no. 1 (2010): 1–4. doi:10.1002/meet.14504701201.

2. Pablo Moriano, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. "Dissemination of Scholarly Literature in Social Media," *Altmetric.org,* May 23, 2014. http://figshare.com/articles/Dissemination_of_scholarly_literature_in_social_media/1035127.

3. Priem, 2.

4. Emily Darling, David Shiffman, Isabelle M. Côté, and Joshua A. Drew. "The Role of Twitter in the Life Cycle of a Scientific Publication," *PeerJ PrePrints*, 2013. https://peerj.com/preprints/16/.

5. Priem, 2.

6. Xin Shuai, Alberto Pepe, and Johan Bollen. "How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations," *PLoS ONE* 7, no. 11 (November 1, 2012): e47523. doi:10.1371/journal.pone.0047523.

7. Andrew Shuping, "The Modern Interlibrary Loan Office Channeling Holmes, MacGyver, and Neo," *College & Research Libraries News* 75, no. 7 (July 1, 2014): 385–86.

8. Collette Mak, "Resource Sharing among ARL Libraries in the US: 35 Years of Growth," *Interlending & Document Supply* 39, no. 1 (February 22, 2011): 26–31. doi:10.1108/02641611111112110.

9. Colleen Kenefick, and Jennifer A. DeVito, "Google Expectations and Interlibrary Loan: Can We Ever Be Fast Enough?," *Journal of Interlibrary Loan, Document Delivery & Electronic Reserve* 23, no. 3 (July 1, 2013): 157–63. doi:10.1080/1072303X.2013.856365.

10. Selden Lamoureux, and James Stemper, "White Paper: Trends in Licensing," *Research Library Issues: A Quarterly Report from ARL, CNI, and SPARC* 275 (June 1, 2011). http://publications.arl.org/rli275/20.

11. Andrea Kuszewski, Twitter post, January 20, 2011, 7:06 p.m., http://twitter.com/AndreaKuszewski

12. Shivika Chandra and Pranab Chatterjee. "Digital Indiscretions: New Horizons in Medical Ethics," *The Australasian Medical Journal* 4, no. 8 (August 31, 2011): 453–56. doi:10.4066/AMJ.2011899.

13. Adam G. Dunn, Enrico Coiera, and Kenneth D Mandl. "Is Biblioleaks Inevitable?" *Journal of Medical Internet Research* 16, no. 4 (April 22, 2014). doi:10.2196/jmir.3331.

14. Ibid.

15. Aaron Swartz. *Guerilla Open Access Manifesto*. Accessed October 26, 2014. http://archive.org/details/GuerillaOpenAccessManifesto.

16. Kroll, David. "#icanhazpdf: Civil Disobedience?" *Terra Sigillata*, December 22, 2011. http://cenblog.org/terra-sigillata/2011/12/22/icanhazpdf-civil-disobedience/.

17. Bora Zivkovic, Twitter post, April 26, 2013, 12:08 p.m, https://twitter.com/BoraZ/

18. Liu, Jean. "Interactions: The Numbers Behind #ICanHazPDF," *Altmetric.com*, Accessed April 23, 2014. http://www.altmetric.com/blog/interactions-the-numbers-behind-icanhazpdf/.

19. Ibid.

20. Kathryn Greenhill and Constance Wiebrands, "No Library Required: The Free and Easy Backwaters of Online Content Sharing." In *VALA 2012 Proceedings*. Melbourne, Austrialia: Victorian Association for Library Automation, 2012. http://vala.org.au/vala2012-proceedings/vala2012-session-11-greenhill.

21. Ibid.

22. Alex Bond, "How #icanhazpdf Can Hurt Our Academic Libraries," *The Lab and Field*. Accessed April 23, 2014. https://labandfield.wordpress.com/2013/10/05/how-icanhazpdf-can-hurt-our-academic-libraries.

23. Jill Emery, "#ICanHazPDF vs. #ICanHazLibrary: Where Librarians Need to Rise to the Occasion," *ALCTS Collection Connection.* Accessed April 23, 2014. http://www.collection-connection.alcts.ala.org/?p=188.

24. Greenhill and Wiebrands.

25. "Database Platform: ISI Web of Knowledge: Academic Database Assessment Tool." Accessed January 30, 2015. http://adat.crl.edu/platforms/about/isi_web_of_knowledge.

26. "Research Areas: Web of Science Help." Accessed February 9, 2015. http://images.webofknowledge.com/WOKRS57B4/help/WOS/hp_research_areas_easca.html.

27. Stephen Bosc and Kittie Henderson, "Steps Down the Evolutionary Road: Periodicals Price Survey 2014," *Library Journal*, April 11, 2014. http://lj.libraryjournal.com/2014/04/publishing/steps-down-the-evolutionary-road-periodicals-price-survey-2014/.

28 Beth Posner, "The View from Interlibrary Loan Services: Catalyst for a Better Research Process," *College & Research Libraries News* 75, no. 7 (July 1, 2014): 378–81.

29. Ibid, 381.

30. Alison Head, *Learning Curve: How College Graduates Solve Information Problems Once They Join the Workplace*, Project Information Literacy Research Report, October 16, 2012. http://projectinfolit.org/images/pdfs/pil_fall2012_workplacestudy_fullreport_revised.pdf.

31. Ewan St John Smith, Twitter post, May 20, 2014, 07:17 a.m, https://twitter.com/psalmotoxin/.

This is the published PDF version of a contributed paper for the Association of College and Research Libraries (ACRL) 2015 conference in Portland, Oregon. This paper can be found at:

http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2015/Gardner.pdf