

Fast and Furious (at Publishers): The Motivations behind Crowdsourced Research Sharing

Carolyn Caffrey and Gabriel J. Gardner

Crowdsourced research sharing takes place across social media platforms including Twitter hashtags such as #icanhazpdf, Reddit Scholar, and Facebook. This study surveys users of these peer-to-peer exchanges on demographic information, frequency of use, and their motivations in both providing and obtaining scholarly information on these platforms. Respondents also provided their perspectives on the database terms of service and/or copyright violations in these exchanges. Findings indicate that the motivations of this community are utilitarian or ideological in nature, similar to other peer-to-peer file sharing online. Implications for library services including instruction, outreach, and interlibrary loan are discussed.



Online communities exist on almost every topic imaginable, so it should be no surprise that there are also dedicated spaces across social media platforms devoted to the exchange of scholarly publications. These exchanges, known as crowdsourcing or peer-to-peer sharing (P2P), can be found on platforms such as Twitter, Reddit, and Facebook and often violate commercial database terms of service (ToS) and/or copyright. For example, Nature Publishing Group's Terms of Service Agreement includes, in part, "you must not ... distribute, transmit, syndicate, sell or offer to sell or otherwise make available any content, files, feeds or data from a Site, whether [publicly] available or not, except as specifically permitted by that Site," which would cover many if not all of the activities in these exchanges.¹ Like online illegal filesharing of music and movies, these communities are a labor of love. It takes time and passion to monitor these networks, find and retrieve the requested publication, remove identifying information, and upload the document for the person requesting it. In what ways are these scholarly networks similar to those for exchanging recreational content? Who is using P2P services for scholarly information? How often do they choose these services? Why are they using these methods compared to traditional library-centric means (such as interlibrary loan)? And what do these users think about the likely ToS and copyright violations? These communities represent only a tiny portion of overall traffic to scholarly resources, but their anonymous nature, strong sense

Carolyn Caffrey is Information Literacy Coordinator at California State University, Dominguez Hills; e-mail: cgardner@csudh.edu. Gabriel J. Gardner is Reference & Instructional Librarian at California State University, Long Beach; e-mail: gabriel.gardner@csulb.edu. ©2017 Carolyn Caffrey and Gabriel J. Gardner, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/4.0/>) CC BY-NC.

of community, and devoted users encourage further exploration as a unique look at the landscape of scholarly communication.

The peer-to-peer sharing of scholarly research examined here is coordinated by the use of social media platforms and online forums. The primary platforms used to organize these communities are Twitter, Reddit, and Facebook, and the primary websites used to host the shared files are AvaxHome, LibGen, and Sci-Hub. Commercial online cloud storage accounts (such as Dropbox or Google Drive) and e-mail are used to share content in more informal ways than those explored here. Users interested in obtaining an article or book chapter will disseminate their request over one or more of the platforms detailed below. Other users of those platforms try to get the requested content via their library accounts or employer-provided access; the actual files being exchanged are often hosted on other websites or e-mailed to the requesting users. While some of these communities are relatively new, P2P file sharing has a long Internet history and has been well explained anthropologically by modeling communities as nonmonetary gift economies described as “a system of social solidarity based on a structured set of gift exchange and social relationships among consumers.”² Data collected in our survey shed light on how participants view their gifts as well as the solidarity that many scholars, who support these communities with their time and access, feel. Additionally, we provide recommendations for libraries and librarians about marketing of interlibrary loan services, support for open-access publications, and information literacy instruction based on the survey results.

Literature Review

Previous Research on Scholarly Peer-to-Peer Sharing

Anecdotally, much scholarly information sharing happens using e-mail.³ Recent work by Carol Tenopir, in collaboration with Elsevier and others, estimates that articles are shared approximately eleven times for every one download; in other words, for every twelve scholars reading an article, only one will have downloaded it from the publisher; e-mail is the predominant method of sharing.⁴ The authors frequently observed that commercial online cloud storage services such as Dropbox or Ge.tt are also used, a finding confirmed by Tenopir.⁵ These decentralized options pale in volume, however, to three centralized efforts: AvaxHome, LibGen, and Sci-Hub. The Library Genesis platform (LibGen) is the only one that has received detailed analysis and critique.⁶ As of October 2014, LibGen contained more than 25 million documents totaling more than 42 terabytes of data; it is estimated to contain 36 percent of all articles that have a DOI assigned. Just three publishers, Elsevier, Springer, and Wiley, account for 83 percent of all of the content freely available on LibGen.⁷ Predictably, Elsevier has been monitoring the growth of LibGen with interest; in June 2015 they filed suit against the proprietors (one named defendant and ninety-nine John Does) of LibGen and Sci-Hub for copyright infringement and violations of the Computer Fraud and Abuse Act.⁸ Cabanac reviewed LibGen file cache logs and noted that, while a majority of the content was uploaded in less than two weeks (which shows evidence of a “biblioleaks” scenario), there is strong evidence of ongoing crowdsourcing, as approximately 2,720 new articles are added to LibGen every day.⁹

In an earlier study of Twitter crowdsourcing, Gardner and Gardner analyzed 674 #icanhazPDF journal articles requests from April to August 2014.¹⁰ Life sciences and biomedicine subject journals comprised 62 percent of those requests, according to Web of Science classification. Besides subject categories, year of publication was also examined to see if publisher embargoes could be playing a role in a user’s inability to access a journal article through an institution he or she was affiliated with. While 34.5 percent of items were requested from 2014, the year of study, only 55 percent of all requests

were from the past five years; suggesting that publisher embargoes may be a reason, though not the only reason, for lack of access. Most users during that time period only requested one article through #icanhazPDF, suggesting it is just one way for users to access a particularly difficult-to-find publication and not the primary means of access.

Interlibrary loan, often neglected in the scholarly communication literature, is the library service that likely loses users to the aforementioned P2P communities; for sharers who are affiliated with a library, every social request represents an unmade (or perhaps unfulfilled) ILL request. Well-run ILL departments can fulfill many of the requests made using peer-to-peer methods, the exceptions likely being materials that are too new or under embargo.¹¹

Peer-to-Peer Sharing and Gift Economies

Multiple theories exist purporting to explain motivations for online file sharing. In his book *Free Culture*, Lawrence Lessig outlined four different types of file sharing: 1) recourse to sharing as a substitute for a purchase; 2) obtaining content shared by others as a way to sample, with intention or desire to purchase; 3) recourse to sharing to get copyrighted content that is difficult to obtain because transaction costs are too high; 4) obtaining content that is not copyrighted through sharing networks or content that the copyright holder wants to give away.¹² Empirical studies on the motivations of file sharers in peer-to-peer networks demonstrate that much file sharing corresponds to Lessig's framework.¹³ In a study of more than 500 college students, Xiao Wang and Steven McClung found that utilitarian motivations surrounding cost, convenience, and availability predicted intentions to download copyrighted digital files.¹⁴

Motivations for uploading content appear to be less utilitarian and more moralized. Some sharers view their effort ideologically and consider it a form of civil disobedience against an unjust intellectual property regime.¹⁵ The activist Aaron Swartz upped the rhetorical ante by declaring the (free) sharing of knowledge "a moral imperative."¹⁶ Others have more prosaic motivations. The sharing of recreational files online has been described as a "gift economy."¹⁷ Gift systems are an ancient human social structure that the digital world is giving new life to; they are defined as having 1) social distinctions; 2) reciprocity norms; and 3) rituals and symbolism.¹⁸ Reciprocity norms, also described as a "sharing ethic," are important; without them, crowdsourced sharing systems could not survive. If the purpose of these communities is only for individuals to get "free" content, each individual has little incentive to contribute.¹⁹ Social distinctions, the sense that sharing content is different from obtaining it via a marketplace, as well as the feeling that those sharing content constitute a "community," are important to sustaining the sharing enterprise. The social distinctions of academic scholarship (particularly science research), the shared symbols that go along with these pursuits, as well as observed get-give transactions in the comments on r/Scholar all indicate that gift system analysis also applies to the sharing of scholarly research throughout the online communities surveyed for this article.

Methods of Crowdsourcing Scholarly Material

Twitter

The use of Twitter to facilitate sharing of scholarly information dates back to 2011, when the #icanhazPDF "hashtag" was coined.²⁰ Users looking for a source tweet a link or partial citation for a paywalled article and append the metadata label, or "hashtag," #icanhazPDF. This label renders the request discoverable through Twitter's linking and search functions. Other Twitter users, once aware of the #icanhazPDF request, use their institutional subscriptions or personal memberships to download the desired PDF and provide it to the requestor. Often the requesting tweet will include an e-mail address, to simplify the exchange. Common practice is to delete any tweets associated with

the request once it has been fulfilled; this avoids duplication of effort and lessens the chance that a copyright holder will notice the transaction.²¹

Reddit Scholar

Reddit Scholar (hereafter r/Scholar) is an online subforum of the popular online bulletin board Reddit. r/Scholar was created in 2009 for the purpose of “requesting and sharing specific articles available in various databases.”²² It has grown steadily since then, reaching 10,000 subscribers in 2012 and 20,000 in 2014.²³ To request an article using r/Scholar, registered Reddit users post to the forum; these posts include the title of the work desired, whether it is an article or a book chapter, or an entire book, and a link to the paywalled source. The forum’s moderators and dedicated users use their institutional subscriptions or personal memberships to download the desired content and provide it to the requestor. This is often done by uploading a file to LibGen or other cloud storage service and posting the nonpaywalled link in reply to the requesting post. Of the various sharing communities, r/Scholar is the only one that urges members to use interlibrary loan (ILL) if requests are not “urgent.” A notice that appears on every page of the forum reads: “ILL avoids potential copyright issues and lets libraries know which subscriptions are useful.” Pseudonymous usernames are common on Reddit; this limits the chance that any transaction could be easily connected with a person’s real identity. The mores of r/Scholar are such that past requests are preserved, though Reddit’s website limits the number of old posts that can be viewed.

Facebook

Facebook is another platform that scholars use to disseminate research and share articles. There are four groups on Facebook devoted to this purpose; some are “communities” in Facebook’s parlance; others are “groups.” Some are open, and some are closed to outside users. Many have a name like “Research Articles, Books and Literature.” The largest community is a closed group with more than 77,000 members.²⁴ Sharing etiquette differs slightly among the various groups, but the basic mechanics are the same. Requesting users post links to the articles they would like to read on the group’s Timeline. The requests are fulfilled either by other group members using Facebook’s messaging system to send the files, typically PDFs, to the requestor or by posting a link to a nonpaywalled copy of the article in the comments section of the requesting post. Some groups allow files to be uploaded and shared with the entire group.

AvaxHome, LibGen, and Sci-Hub

The loci of nonpaywalled copies of articles are the LibGen and Sci-Hub websites; AvaxHome is a popular hosting option for DRM-free (digital rights management) e-books.²⁵ WHOIS queries of the domain names reveal scant information about the individuals responsible for these websites and where they might be hosted.²⁶ According to Alexa Internet rankings, AvaxHome is the 2,840th most popular website on the Internet; LibGen is the 8,518th; Sci-Hub is the 12,137th.²⁷ Web traffic to these sites comes from a variety of countries, with Iran, China, and the United States figuring prominently in LibGen and Sci-Hub use. Previous research has found the use of the #icanhazPDF hashtag on Twitter to be overwhelmingly an Anglophone phenomenon.²⁸ But, as Alexa traffic data reveals, aggregating and sharing nonpaywalled research is a popular international effort. These services are noncommercial entities that face challenges to long-term stability, the most obvious one being that essentially all of the content they host is created by publishers who hold copyright on said content. The fact that the creators or administrators take pains to obscure any information that might be learned about them through WHOIS queries indicates that they do not

want to be connected by name to these services in case of any legal repercussions for hosting copyrighted content.

Methods

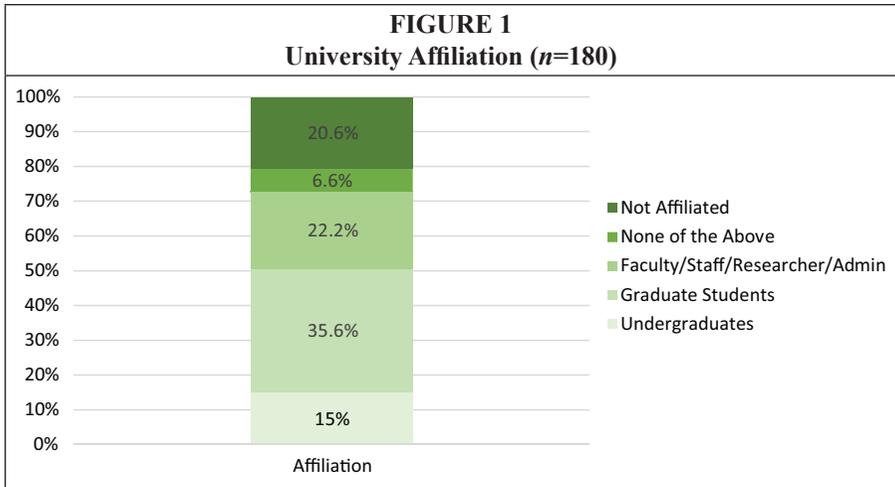
A survey was developed using feedback from our Association of College and Research Libraries (ACRL) conference paper presentation.²⁹ Questions attendees asked during the session helped guide the questions the authors explored through this survey. The survey included categorical, quantitative, and open-ended questions. Categorical questions addressed demographic information such as university affiliation (undergraduate, graduate student, and the like), subject area studied, country of residence, and age based on a range of years and information on which methods the respondent used to request or provide scholarly sources. Quantitative questions focused on frequency of peer-to-peer sharing methods. The open-ended questions asked for respondents to detail their motivations for obtaining and providing scholarly information using peer-to-peer sharing services, their concerns and perceptions on the legality of doing so, and what factors determined whether they used interlibrary loan or a peer-to-peer service.

After the survey was developed in Qualtrics, it was reviewed by several colleagues for usability and to ensure the use of nonjudgmental language. Subsequently, the survey and its dissemination methods were approved by the university's Institutional Review Board (IRB). The survey instrument is included in the appendix; it was open to respondents for four weeks. Using personal Twitter accounts, the authors distributed the link publicly using the #icanhazpdf Twitter hashtag. By using the hashtag, the tweets were then retweeted to a larger network through bot accounts that have been programmed to automatically retweet requests to a network of #icanhazpdf users. For r/Scholar, the authors reached out to the moderators of the Reddit subforum for permission to post the survey link. The forum moderators agreed that the survey could be reposted once a week so that it would be visible to infrequent users and also "stickied" it (pinned as the top post) for one week. On Reddit, the survey link could also be voted and commented on by forum users. "Up-votes" (and "down-votes") are a type of content moderation used to keep items of interest to this community visible. Across three separate postings to the r/Scholar forum, the survey received fifty-nine "up-votes," which denote positive interest from the forum subscribers and increased the post's location (and thus its visibility) in the queue.

One interesting element of conducting Internet research is that your respondents can contact you instantly; several respondents e-mailed the authors asking for check boxes as opposed to a multiselect column on the survey. The authors obliged to make the survey more mobile device-friendly. Each question was optional to follow IRB compliance; few respondents answered all questions. Additionally, not all questions were visible to all users—for example, users who answered that they had never used r/Scholar before did not see a question asking about their motivation for using r/Scholar, as it was not applicable. With this in mind, to be clear about the results, we are summarizing the number of respondents; (*n*) is noted for each question in the results section. Closed-choice answers were analyzed for descriptive statistics in SPSS. Questions with open-text responses were reviewed by both authors to identify emergent themes. After identifying initial themes, the authors coded the responses and revised themes as necessary. Finally, both authors coded the responses independently using the emergent themes and reconciled any differences.

Results

A total of 252 respondents consented and completed at least one question in the survey. As seven participants consented to the survey but then did not complete any questions, we removed those unanswered surveys from our analysis.



Demographic & Discipline Questions

Of the 178 users who answered this question, 36.5 percent fell between 18 and 25 years old, 41 percent between 26 and 34, 20.2 percent between 35 and 54, and only 2 percent for 55 and over.

Most people across all demographic categories responded that they resided in the United States (57.3%), followed by the United Kingdom (11.7%), Canada (7%), France (5.3%), and Germany (2.9%) (n=171). Other countries were also selected to lesser degrees. Cross-tabbed analysis between country of residence and university affiliation, as well as country of residence with frequency of article requested/provided did not reveal any patterns of significance.

To determine if any one discipline was overrepresented in crowdsourced sharing, we asked respondents to choose from more than 100 distinct Web of Science categories that they typically requested or were provided scholarly materials from. Respondents could select more than one category: for example, if a neuroscientist often requested articles from psychology and biomedical sciences, she could select both categories. Web of Science subject categories were chosen as a level of analysis because they are reviewed by experts and are exhaustive without being overwhelming.³⁰ Web of Science’s classification of journals with multiple subject categories also provides a clearer picture of what disciplines are represented by mirroring the often interdisciplinary nature of scholarship. Of the 147 people who answered this question, 62 percent selected more than one subject. With so many different subjects, however, there were no clear patterns. Therefore, the authors used the broader Web of Science Research Domains to compute a new variable encompassing the individual subjects into the larger

categories of: Arts & Humanities, Social Science, Physical Sciences, Biomedical and Life Sciences, and Technology.

TABLE 1
Users by Research Domain (n=147)

Research Domain	Number of Users
Arts & Humanities	25
Technology	48
Social Sciences	65
Physical Sciences	65
Biomedical and Life Sciences	91

Categorical Questions

Users were asked, “Have you ever used any of the following services to *obtain* scholarly materials (such as journal articles)?” and to check all services that they had used (n=216). Respondents also had the option to select “I have not ob-

tained scholarly materials using any of these services" (n=13) or enter another option in an open-text field. LibGen or a LibGen mirror was mentioned by twenty-one people in the open text field. Other open-ended responses included ways to obtain materials outside a defined system such as asking a friend or colleague.

The next question asked about services used to provide materials (n=208). It included the same options with the exception of interlibrary loan and Open Access Button, as those systems do not crowdsource the providing side of a transaction. Respondents also had the option to select "I have not provided scholarly materials using any of these services" (n=61) or "Other." "Other" (n=77) was not elaborated on in this instance and presumably would also include LibGen. In retrospect, not enumerating LibGen, Sci-Hub, and their mirrors as responses to the question was an oversight on our part.

Service	Number of Users
BitTorrent	18
#icanhazpdf	45
r/Scholar	71
Other	77

Open-Ended Questions

The reasons users included for obtaining scholarly materials from peer-to-peer services were coded into the following themes: Speed, User Experience, Ideology, Access, and Cost (n=148). Text responses could include multiple themes, each of which was coded. Speed was used when respondents referred to how quickly they could have a request filled. User Experience was used when respondents said things like "easy" or "convenient" — some users explicitly mentioned cumbersome and lengthy ILL request forms in contrast to peer-to-peer methods. The Ideology theme included responses that focused on the belief that scholarly information should be free as a public good. Access was used for responses that described these methods as their only believed access point, referencing inability to use ILL, lack of institutional subscription, or lack of institutional affiliation. Cost included responses that talked about certain publications being cost prohibitive as well as users who felt any cost was too high. Responses that contained reasons that did not fit within these categories or were otherwise too difficult to tell what the respondent intended were included in a catch-all "Other" category. Responses in the Other category also included two passive observers of r/Scholar who read papers that others post but are not actively seeking out a particular article to meet an immediate information need.

Service	Number of Users
#icanhazpdf	47
BitTorrent	57
Open Access Button	67
Interlibrary loan	119
r/Scholar	127
Other	102

When asked how often users provide or obtain materials using these methods (n=149), many described infrequent usage. The majority of respondents (51.7%) indicated they use these services less than once a month, 27.5 percent indicated 1–3 times a month, 8.1 percent said once a week, 8 percent indicated 2–3 times a week, and only 4.7 percent indicated 4 or more times a week. r/Scholar and other methods (e-mail, LibGen, and other such methods) were used more frequently than #icanhazpdf and BitTorrent.

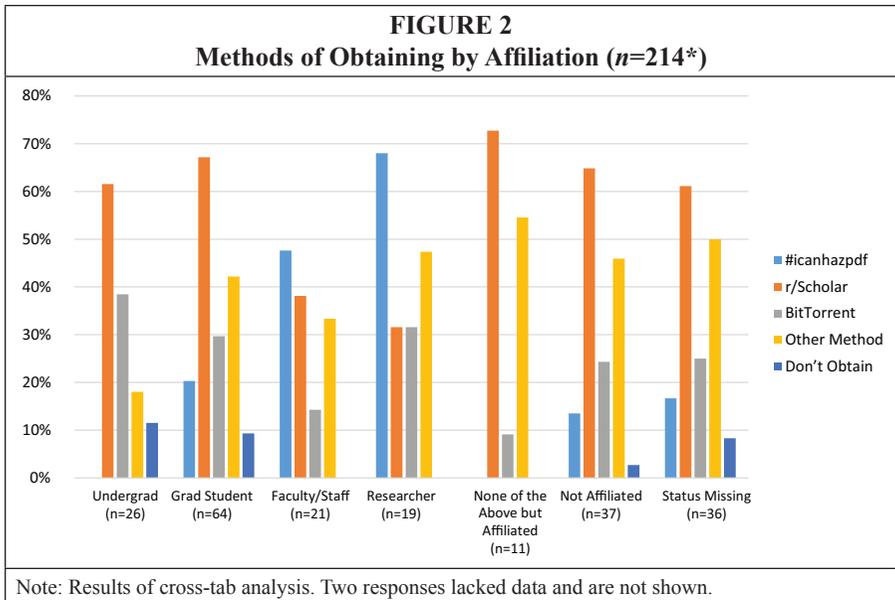
These selected quotes represent the range of responses:

- “Fist [sic], the simplicity of the request. It’s one DOI link to Twitter compared to a page long form to interlibrary loans. Second, speed of fulfillment: a few minutes compared to days or weeks. Libraries have to do better.” (*Categorized as Speed and User Experience*)
- “Being from a third world country that recently lost access to HINARI (which gave us a good source of access), it is really a struggle to access research material, especially when paying \$30+ to see if an article is worth reading, or when even universities can’t afford to pay access to good subscription services, and have to suffice with other services which seem to be outdated and with limited access (e.g. embargoes, subpar books, etc.)” (*Categorized for Access and Cost*)
- “To avoid paying the outlandish amounts for information that should be free.” (*Categorized as Cost and Ideology*)

The reasons users included for providing scholarly materials to others were coded

Themes	Number of Responses
Access	76
Speed	63
User Experience	48
Cost	32
Other	12
Ideology	11

into the themes: Reciprocity, Community, Solidarity, Ideology/Civil Disobedience, and Other (n=104). Reciprocity was used when respondents described feeling like they needed to have a fair trade or “giving back” to the peer-to-peer sharing system. Community and Solidarity were often found together; Community was used when respondents explicitly mentioned a community (either in the online community or their professional colleagues at large). Solidarity was used when the respondent mentioned understanding what it was like to be without access or in a similar position to the



requester. Ideology/Civil Disobedience included responses that focused on the belief that scholarly information should be free and included references to public good, tax-funded research, and often the very phrase “civil disobedience.” Responses that contained reasons that did not fit within these categories or were otherwise too difficult to tell what the respondent intended were again included in a catch-all “Other” category. Reasons provided that fell in the “Other” category included boredom and ease of providing articles.

The selected quotes represent the responses received:

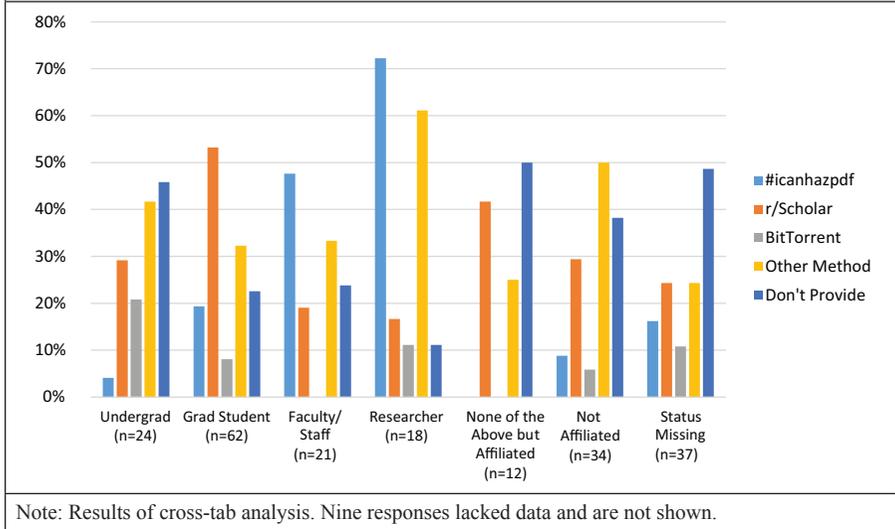
- “I generally do 2:1 (I give 2 articles the same day I ask for one). I presuppose if I want to receive eventually, I also have to give to maintain community alive.” (*Categorized for Reciprocity and Community*)
- “My own research would not have been possible without help from these caring, sharing communities. Thus it is natural and right that I should give help back to others in need of access to research literature. The legacy academic publishers (e.g. Elsevier, Wiley, Taylor & Francis, NPG, Springer...) make billions from actively restricting access to research—this kind of civil disobedience is an inevitable and logical response to such ruthless monopolists.” (*Categorized for Reciprocity and Ideology/Civil Disobedience*)
- “If you’re gonna take, you should give, too, right? Also, I’m a librarian at an academic institution. In the past, as part of my reference shift, I would go through r/scholar requests and check them against my library’s full-text access, filling as many as I could within those time periods. I viewed this as continuing my role to help provide access and furthering the research of others.” (*Categorized for Reciprocity and Ideology/Civil Disobedience*)
- “Other people are stuck at universities with crappy libraries, too.” (*Categorized for Solidarity*)

Themes	Number of Responses
Reciprocity	58
Ideology	32
Community	25
Solidarity	23
Other	9

researchers participating in the peer-review process “for free,” and other beliefs that scholarly information should be available at no charge to the public. Responses coded under Not a Violation included comments that peer-to-peer services are akin to emailing a colleague so they should not be considered a violation of ToS or copyright and other arguments that these methods do not breach existing contracts. Responses included under Don’t Care included respondents who had considered the violation and determined they did not care enough to not send an article as well as those who had not yet considered the copyright implications of these services. Animus Toward Publishers was used when responses indicated academic publishing systems, specific publishers, or publishing practices they disagree with. Displeasure with Current Copyright Regime covered responses focused on copyright more generally. Finally, an Other category

The question, “What are your views regarding the potential violation of copyright or a database’s terms of service that may occur by using r/Scholar, Torrents, or #icanhazPDF?” was answered by 160 participants, and the authors used emergent theme coding to categorize responses into the following categories: Information Should Be Free, Not a Violation, Don’t Care, Animus Toward Publishers, and Displeasure with Current Copyright Regime. Information Should Be Free included responses that mentioned publicly funded research,

FIGURE 3
Methods of Providing by Affiliation (n=208*)



was used for responses that did not fall into these categories, which included several responses that said their views did not matter because these sources provide access to information regardless. The responses to this question included profanities and the selected quotes below have not been censored to ensure they accurately capture tone:

- “Meh. we’re hardly breaking a multi-billion dollar industry. They’ll be fine.” (Categorized for *Animus Toward Publishers*)
- “Fuck the publishers if they have ‘transferred’ the copyright from the authors—I entirely disregard ‘publisher-owned’ copyright. I expect 99.99% of most academics would have no problem with other academics copying their research so that they have a copy to read. Essentially, I just don’t care. ... Ever since Napster, people of my generation have been used to routinely ‘infringing’ copyright in a digital context. It’s an everyday activity—not just in academic realms.” (Categorized for *Displeasure with Current Copyright Regime, Don’t Care, and Animus Toward Publishers*)
- “Ugh, copyright. I get it but at the same time, I just need to find these articles for my research. I’ll cite them!” (Categorized for *Don’t Care*)
- “If the companies didn’t suck (especially since I’m generating the content and acting as a reviewer for others’ content, yet still don’t get access) I might not feel bad. ... I mean, seriously, my institute’s library doesn’t even have a subscription to some of major journals in my field we publish in regularly. I’ve wound up using Reddit Scholar as a faster way of getting /my own/ articles.” (Categorized for *Animus Toward Publishers*)
- “I would GLADLY publish my own data for free if there was an easy way to, but my PI and a majority of others maybe don’t feel the same way. Yet, they accept these papers I get through r/scholar with no regrets. Data should be free, and putting copyright on data to lock it away is disgusting.” (Categorized for *Information Should Be Free*)

Finally, the last open-ended question asked users who had earlier selected ILL as a method they used: “What determines whether you obtain materials through interlibrary loan services compared to another method?” The reasons users included were

Themes	Number of Responses
Don't Care	67
Information Should Be Free	48
Animus Towards Publishers	30
Not a Violation	24
Displeasure with Copyright	23
Other	14

electronic (it is beyond the scope of this study to determine if this perception is the actual library practice). Five respondents cited the cost of their ILL services as prohibitive. Access was used when the respondent said he or she did not have access to ILL; this includes those without an institutional affiliation. The User Experience category contained responses that mentioned ease of use, convenience, poor scans, and reliability of service. Responses that contained reasons that did not fit within these categories or were otherwise too difficult to tell what the respondent intended were again included in an Other category.

Selected quotes:

- “#icanhazpdf is much quicker (no form to fill / no question asked)” (*Categorized for User Experience*)
- “Since finding these services I have ceased using interlibrary loan. Waiting several days for interlibrary loan, then having to photocopy an article is simply far more inconvenient. Getting the PDF to print through these services results in precisely the same end state and typically takes minutes to hours rather than days.” (*Categorized for Speed*)
- “If I cannot find someone with access through /r/scholar, science communities, or friends at other research universities then I would utilize ILL.” (*Categorized for ILL as Last Resort*)

Discussion

It is unclear how generalizable these results may be, because there is not a known number of #icanhazpdf or r/Scholar users. In a previous study, Gardner and Gardner found that, on average, 158 unique individuals used #icanhazpdf publicly every month during the period analyzed.³¹ A later study by Swab and Romme found an increase in users during their three-month period of analysis.³² As word of these communities spreads among scholars, more of them appear to be drawn to them. The survey response rate seems to suggest we captured a substantial percentage of

categorized into the themes: Speed, ILL as a Last Resort Only, ILL Is Print Only, Cost, Access, and User Experience. Speed included anecdotes that ILL services often takes days or weeks, whereas respondents compared peer-to-peer services as only taking hours. The “last resort” theme was used for respondents who described only using ILL as a last resort when all other methods (including peer-to-peer) fail. Other respondents cite their ILL services as being primarily for print sources, and those articles may be received in print when the respondent would prefer elec-

Themes	Number of Responses
Speed	40
User Experience	19
ILL is print only	18
Access	15
Last Resort	11
Cost	5
Other	5

#icanhazpdf users. Yet, r/Scholar has more than 20,000 subscribers suggesting that the survey data presented here only scratches the surface of that community. Another issue affecting generalizability is language. Transactions on Twitter, Reddit, and Facebook are typically in English; our survey was available in English only. Yet Iran and China, two countries not represented among survey respondents, are the main sources of traffic to LibGen and Sci-Hub; very little can be said about the users of those sites based on our results.³³ Regardless, we believe that the survey responses provide insight into the motivations of engaged participants in these P2P communities that are important for libraries and the larger field of scholarly communications to consider. Additional research should be conducted to determine the extent of these user populations.

Demographic questions were at the end of the survey and seemed to have a lower response rate as a result. Additionally, while the consent form promises anonymity, users still may have felt uncomfortable disclosing demographic information in combination with potentially illicit behaviors. The data on users' subject specialties matched previous content analysis study of articles requested, with life sciences and biomedicine predominating and the arts and humanities comprising a minuscule subset.³⁴ One hypothesis for the discrepancy between disciplines is that Arts and Humanities fields rely less on journal articles and therefore are not as well represented in #icanhazpdf and r/Scholar, which are geared toward journal articles. Our own field, Library and Information Science, is not exempt from these crowdsourced systems; users did select that option, and they are represented in the Technology research domain. The authors were surprised by how many users had an academic affiliation, meaning that these services are not isolated to only those without access to library systems.

The majority of respondents (92%) claim to be obtaining articles for utilitarian reasons described by Wang and McClung; motivations of speed, access, and usability are what drive them to Twitter, r/Scholar, or Facebook.³⁵ They did not voice explicit support for open access or against the current copyright regime—crowdsourcing is just their preferred alternative to using interlibrary loan. Among those respondents who provide articles in these networks, however, moral motivations and ideology figure more prominently. P2P sharing has been deemed an act of civil disobedience by influential scientists for years, so this is somewhat expected.³⁶ Of the 104 respondents to the open-ended question about motivations for providing, thirty-two had answers coded for Ideology. Yet the top reason people provide articles to others is the reciprocity norm; these results correspond to the study from Cenite et al., which found reciprocity norms to be “the crux of P2P file sharing,” with only a small subset of providers viewing themselves as providing content that was “*meant to be shared.*”³⁷ The small subset is probably larger in our sample due to the nature of the information shared; as many commenters pointed out, research in the sciences often receives government funding. Some vocal respondents involved in crowdsourcing view themselves within the tradition of “guerilla open-access” pioneered by Aaron Swartz, who achieved notoriety by attempting to make articles from JSTOR and from the federal government’s PACER (Public Access to Court Electronic Records) database freely available online.³⁸ Only three people selected that they provide materials but do not obtain them. Some users form an altruistic core in these communities, helping to sustain them despite the presence of free-riding. On the other hand, sixty-one respondents (24%) obtain materials without contributing back to the community. Our methodology precludes any definitive statements about the amount of “free-riding” in these communities, but it appears to be unusual behavior. Reciprocity norms and a sense of solidarity will likely sustain P2P sharing communities for the foreseeable future.

When it comes the ways in which respondents view their actions in terms of potential copyright and ToS violations, it is important to note that several respondents skipped

every single question in the survey except this one, indicating it was likely an issue they felt passionate about and thus may include more extreme views. Additionally, in methods that are a one-to-one peer exchange (such as #icanhazpdf), some users argue this can fall under the protections of “fair use” and is no different from e-mailing an article to a colleague. We authors are not lawyers, but we understand that license terms generally override copyright law where the two differ; in our own experiences, license terms often restrict distribution via e-mail for a course rather than linking through the database.³⁹ In Elsevier’s terms and conditions, for example, this type of sharing is expressly prohibited: “You may print or download Content from the Site for your own personal, non-commercial, informational or scholarly use, provided that you keep intact all copyright and other proprietary notices. You may not copy, display, distribute, modify, publish, reproduce, store, transmit, post, translate or create other derivative works from, or sell, rent or license all or any part of the Content, products or services obtained from the Site in any medium to anyone, except as otherwise expressly permitted under these Terms and Conditions, relevant license or subscription agreement or authorization by us.”⁴⁰

While Lessig suggested that peer-to-peer sharing of entertainment content might be a temporary problem, as copyright holders adjusted distribution for the digital landscape and improved their payment options, the persistence of—indeed, increasing popularity of—the sharing communities discussed above make it clear that his prediction did not hold for this nonentertainment content.⁴¹ The major scholarly publishers have offered precisely the type of simple online payment options (normal subscriptions as well as pay-per-use) that he suggested might blunt the use of illegal downloading for recreational purposes. The sharing of scholarly research might at first seem like a clear-cut example of Lessig’s Type A: sharing as purchase substitution; to the publishers who hold copyright over the articles, it must seem that way. It is obviously not a case of Type C: sharing to access copyrighted content that is not for sale online. Yet our results indicate that many participants view these activities as if they are Type D sharing: “content that is not copyrighted or that the copyright owner wants to give away.”⁴² Scientific authors are typically thrilled when their work is consulted or cited; nor do they publish journal articles to make a profit: their research exists to be used and shared. The current publishing system and copyright regime in which authors sign over the copyright to the publisher is overlooked or thought to be illegitimate from some respondents. In some circumstances, the scholars who participate in these communities and use the LibGen and Sci-Hub repositories are partially responsible for the very conditions they lament. Many publishers allow authors to post versions of their work online in institutional repositories or similar outlets. The RoMEO database provides authors with information about self-archiving (preprint or postprint versions) policies allowed by journals in their field.⁴³ Also, scholars can choose to publish in gratis or libre open access publications.⁴⁴ Many scholars are under a great deal of pressure to publish their work in specific journals, journals that may not have liberal self-archiving policies.⁴⁵ However, to download an article from LibGen or Sci-Hub for use in a future paywalled article, for which the author took no steps to make open access, is surely hypocritical.

Poor usability is also hindering our patrons from gaining access to materials. Librarians need to apply user experience thinking to all our online systems. At our respective libraries we have to click multiple times just to discover if an item is owned. Besides complicated discovery methods, software or holdings errors are possible. As Ben Rowe would say, some of the error messages received do not appear to be written by humans or for humans.⁴⁶ Librarians need to view these crowdsourced communities as alternatives that fill a gap that we have yet to meet as opposed to purely underground and

shadowy communities. Barriers to use that users mentioned in their responses included: long forms, the perception that too much information was required, clicking multiple times, and difficulty with proxy servers. Low-quality scans, scans as images that are not full-text searchable, and inaccurate scans also influenced respondents' opinion of ILL and came up as reasons why the service is not their first stop. It is important that librarians not chastise peer-to-peer sharers or dismiss them as "not knowing" about ILL. For example, the r/Scholar forum explicitly mentions and encourages ILL, and a majority of survey respondents who answered how they obtain materials listed ILL as one method they use. To better serve our users compared with crowdsourced access to research, library ILL departments must improve the usability of their services and reduce fulfillment times — awareness and publicity are a smaller piece of the problem than one might think for these communities. Advertising quick turnaround times, since speed was the most frequent reason given for not using ILL, seem especially important.

Conclusion

Peer-to-peer article sharing is a multicausal scholarly communication dilemma with roots that go beyond document delivery to the legal bedrock that is our current copyright and intellectual property system. Recognizing the complexities at issue here, there are two areas, beyond improving the general user experience in our online environments, that we recommend libraries and librarians focus on. First, continued work for open access advocacy must continue. This includes talking with these authors, primarily our graduate students and faculty, about publishing in open access publications, negotiating their rights as authors, exploring article processing charge subventions for gold Open Access publications, and institutional open access mandates.

Second, librarians and faculty should continue to collaborate on educating students, at all levels, on dispositions and knowledge practices raised in the "Information Has Value" section of the ACRL *Framework for Information Literacy for Higher Education*. Students will ideally engage with the many meanings of value, including as a commodity, to help them understand the socioeconomic and political interests involved in its production and dissemination.⁴⁷ Possible learning outcomes raised in this frame are often relegated to brief conversations at the end of a one-shot instruction session about content being paid for on the library's website or as around citations or clicking on the libraries web page since we "pay for it" or, as Patti Ryan and Lisa Sloniowski put it, "we often operate as extensions of the database vendors whose products we rent and encourage our students to consume."⁴⁸ The authors are implicated in these pedagogical shortcuts too, but we believe that engaging with these issues critically before students graduate will better prepare them to seek out open access information as an unaffiliated user and to understand and influence the scholarly communication infrastructure as creators. Increased criticality in librarianship and information literacy pedagogy more specifically as demonstrated in the recent publication of titles related to critical theory can provide guidance in having nuanced conversations about the cost of knowledge. Yasmin Sokkar Harker provides lesson plan suggestions for discussing legal information with students accustomed to cost-efficiency narratives to include critiques on "who profits from the legal information industry, who is disadvantaged by it and discuss alternatives to commercial legal information."⁴⁹

Finally, future research into these sharing communities could focus on the total user population as well as quantifying traffic volume. Quantification and more detailed user profiles would allow libraries to accurately determine the impact that peer-to-peer scholarly sharing has on their interlibrary loan volumes and services. Additional research may also examine how the typical profile of a P2P user compares to the average scholar's information-seeking and -sharing behaviors in particular dis-

ciplines. Given our survey's sampling method, we cannot be certain; but, *if* our sample is representative of the sharing population, the typical user is not a scientist toiling away in the developing world locked out of the scholarly community due to "the cost of knowledge."⁵⁰ Rather, she is a social or hard science researcher who has academic library privileges but prefers crowdsourced methods of obtaining access for any of the reasons enumerated above.

Appendix: Survey Questions

Survey was distributed via Qualtrics software.

1. Have you ever used any of the following services to *obtain* scholarly materials (such as journal articles)? Check all that apply.
 - #icanhazPDF
 - r/Scholar
 - BitTorrent
 - Interlibrary Loan
 - Open Access Button
 - Other
 - I have not obtained scholarly materials using any of these services.
2. Have you ever used any of the following services to *provide* scholarly materials to others (such as journal articles)? Check all that apply.
 - #icanhazPDF
 - r/Scholar
 - BitTorrent
 - Other
 - I have not provided scholarly materials using any of these services.
3. [Shown only to respondents who answered a prior question with a relevant answer.] Describe the other method(s):
4. [Shown only to respondents who answered a prior question with a relevant answer.] How did you discover #icanhazPDF?
 - Colleague recommendation
 - Social media post (Facebook, Twitter, and such)
 - At a conference
 - Scholarly literature
 - Other
5. [Shown only to respondents who answered a prior question with a relevant answer.] How did you discover r/Scholar?
 - Colleague recommendation
 - Social media post (Facebook, Twitter, and such)
 - Other Reddit community
 - At a conference
 - Scholarly literature
 - Other
6. [Shown only to respondents who answered a prior question with a relevant answer.] Why do you use r/Scholar, #icanhazPDF, or BitTorrent to *obtain* materials?
7. [Shown only to respondents who answered a prior question with a relevant answer.] Why do you use r/Scholar, #icanhazPDF, or BitTorrent to *provide* materials?
8. [Shown only to respondents who answered a prior question with a relevant answer.] How often do you *obtain or provide* materials through #icanhazPDF, r/Scholar, or BitTorrent?
 - Never
 - Less than once a month
 - 1–3 times a month
 - Once a week
 - 2–3 times a week
 - 4–5 times a week
 - More than 5 times a week
9. [Shown only to respondents who answered a prior question with a relevant answer.] What subject areas of scholarly materials do you typically *obtain or provide*?

- Select all that apply.
- (Listing of Web of Science categories)
10. [Shown only to respondents who answered a prior question with a relevant answer.] What determines whether you obtain materials through interlibrary loan services compared to another method?
 11. What are your views regarding the potential violation of copyright or a database's terms of service that may occur by using r/Scholar, Torrents, or #icanhazPDF?
 12. How are you affiliated with a college or university? If not affiliated please select, "Not affiliated"
 - Undergraduate student
 - Graduate student
 - Teaching faculty/staff
 - Researcher
 - Administrator
 - None of the above
 - Not affiliated
 13. How old are you?
 - Under 18
 - 18–25
 - 26–34
 - 35–54
 - 55–64
 - 65 or over
 14. In which country do you reside? (Listing of countries)

Notes

1. "Nature Terms," *Nature*, last modified June 23, 2015, available online at www.nature.com/info/tandc.html [accessed 10 July 2015].
2. Markus Giesler, "Consumer Gift Systems," *Journal of Consumer Research* 33, no. 2 (2006): 283–90.
3. Laura L. Haines et al., "Information-Seeking Behavior of Basic Science Researchers: Implications for Library Services," *Journal of the Medical Library Association: JMLA* 98, no. 1 (2010): 73–81, doi:10.3163/1536-5050.98.1.019.
4. Suzie Allard, "The Secret Life of Articles: From Download Metrics to Downstream Impact" (presentation, How Researchers Share Articles: Impact on Library Resources and Services, Library Connect Webinar, December 3, 2015).
5. Carol Tenopir et al., "To Boldly Go Beyond Downloads: How Are Journal Articles Shared and Used?" (presentation, The Charleston Conference, Charleston, N.C., November 5–8, 2014).
6. Guillaume Cabanac, "Bibliogifts in LibGen? A Study of a Text-Sharing Platform Driven by Biblioleaks and Crowdsourcing," *Journal of the Association for Information Science and Technology*, early view (2015): 1–11, doi:10.1002/asi.23445.
7. *Ibid.*, 3.
8. Gary Price, "Elsevier Files Lawsuit against Site Providing Access to Pirated Articles," *Library Journal INFOdocket* (blog), June 9, 2015, available online at www.infodocket.com/2015/06/09/elsevier-cracks-down-on-pirated-scientific-articles/ [accessed 9 June 2015]; *Compl. Elsevier Inc, Elsevier B.V., and Elsevier Ltd. v www.scihub.org, www.libgen.org, Alexandra Elbakyan, and John Does 1–99, No. 15-cv-4282 (RWS) (S.D.N.Y. June 3, 2015)*.
9. Adam G. Dunn, Enrico Coiera, and Kenneth D Mandl, "Is Biblioleaks Inevitable?" *Journal of Medical Internet Research* 16, no. 4 (2014): e112, doi:10.2196/jmir.3331; Cabanac, "Bibliogifts in LibGen?" 5.
10. Carolyn Caffrey Gardner and Gabriel J. Gardner, "Bypassing Interlibrary Loan via Twitter: An Exploration of #icanhazpdf Requests," in *Creating Sustainable Community: The Proceedings of the ACRL 2015* (ACRL 2015, Portland, Ore.: Association of College and Research Libraries, 2015), 95–101, available online at www.ala.org/acrl/sites/ala.org/acrl/files/content/conferences/confsandpreconfs/2015/Gardner.pdf [accessed 10 July 2015].

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

12. Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock down Culture and Control Creativity* (New York: Penguin Press, 2004), 68–69.

13. Mark Cenite et al., "More Than Just Free Content: Motivations of Peer-to-Peer File Sharers," *Journal of Communication Inquiry* 33, no. 3 (2009): 206–21, doi:10.1177/0196859909333697; Jan U. Becker and Michel Clement, "Dynamics of Illegal Participation in Peer-to-Peer Networks: Why Do People Illegally Share Media Files?" *Journal of Media Economics* 19, no. 1 (2006): 7–32, doi:10.1207/s15327736me1901_2.

14. Xiao Wang and Steven R. McClung, "Toward a Detailed Understanding of Illegal Digital Downloading Intentions: An Extended Theory of Planned Behavior Approach," *New Media & Society* 13, no. 4 (2011): 663–77, doi:10.1177/1461444810378225.

15. David Kroll, "#icanhazpdf: Civil Disobedience?" *Terra Sigillata* (blog), December 22, 2011, available online at <http://cenblog.org/terra-sigillata/2011/12/22/icanhazpdf-civil-disobedience/> [accessed 30 January 2015].

16. Aaron Swartz, "Guerrilla Open Access Manifesto" (Eremo, Italy, 2008), available online at <http://archive.org/details/GuerrillaOpenAccessManifesto> [accessed 26 October 2014].

17. Cenite et al., "More Than Just Free Content," 209–10; Markus Giesler and Mali Pohlmann, "The Anthropology of File Sharing: Consuming Napster as a Gift," *Advances in Consumer Research* 30, no. 1 (2003): 273–79; Giesler, "Consumer Gift Systems," 283–90; McKenzie Wark, "Copyright, Copyleft, Copygift," *Open: Cahier on Art and the Public Domain* 12 (2007): 22–29, available online at www.onlineopen.org/copyright-copyleft-copygift [accessed 8 May 2015].

18. Giesler, "Consumer Gift Systems," 284.

19. Becker and Clement, "Dynamics of Illegal Participation in Peer-to-Peer Networks," 12–16.

20. Andrea Kuszewski, Twitter post, January 20, 2011, 5:06 p.m., available online at <https://twitter.com/AndreaKuszewski/> [accessed 6 February 2015].

21. Bora Zivkovic, Twitter post, December 8, 2012, 7:49 p.m., available online at <https://twitter.com/BoraZ/> [accessed 6 February 2015].

22. "Scholar," *Reddit*, available online at <https://www.reddit.com/r/Scholar/> [accessed 5 June 2015].

23. "r/Scholar metrics," *Reddit Metrics*, available online at <http://redditmetrics.com/r/Scholar> [accessed 4 July 2015].

24. "Research Articles, Books and Literature," *Facebook*, available online at <https://www.facebook.com/freeliterature12> [accessed 4 July 2015].

25. LibGen is a multiheaded hydra with many mirrors and domains; at the time of this writing, <http://www.libgen.biz> is the simplest address. Sci-Hub also maintains mirrors at various domains; <http://www.sci-hub.org> is the original address. AvaxHome is located at <http://avxhome.se>.

26. For detail, see: <https://whois.icann.org/en/lookup?name=sci-hub.org>; <https://whois.icann.org/en/lookup?name=libgen.biz>; and <https://who.is/whois/avxhome.se>.

27. All numbers and percentages from Alexa as of 6/10/2015.

28. Jean Liu, "Interactions: The Numbers Behind #ICanHazPDF," *Altmetric* (blog), May 9, 2013, available online at www.altmetric.com/blog/interactions-the-numbers-behind-icanhazpdf/ [accessed 30 January 2015]; Gardner and Gardner, "Bypassing Interlibrary Loan Via Twitter," 95–101.

29. Gardner and Gardner, "Bypassing Interlibrary Loan Via Twitter," 95–101.

30. "Research Areas," *Web of Science Help*, last modified August 5, 2012, available online at http://images.webofknowledge.com/WOKRS57B4/help/WOS/hp_research_areas_easca.html [accessed 9 February 2015].

31. Gardner and Gardner, "Bypassing Interlibrary Loan via Twitter," 97.

32. Michelle Swab and Kristen Romme, "#icanhazpdf? User Requests for Medical Literature on Twitter" (poster presentation, Canadian Health Libraries Association Conference, Vancouver, BC, June 19–22, 2015).

33. See above and Note 31.

34. Gardner and Gardner, "Bypassing Interlibrary Loan via Twitter," 99.

35. Wang and McClung, "Toward a Detailed Understanding of Illegal Digital Downloading Intentions," 667–74.

36. Kroll, "#icanhazpdf: Civil Disobedience?"

37. Cenite et al., "More Than Just Free Content," 214. Emphasis in original.

38. Swartz, "Guerrilla Open Access Manifesto"; Wesley Yang, "The Life and Afterlife of Aaron Swartz," *New York* (February 18, 2013), available online at <http://nymag.com/news/features/aaron-swartz-2013-2/> [accessed 1 December 2015].

39. "Copyright Information for MIT Faculty," *Scholarly Publishing @ MIT Libraries*, available online at <https://libraries.mit.edu/scholarly/publishing/copyright-information-for-mit-faculty/> [accessed 7 December 2015].

40. "Elsevier Website Terms & Conditions," *Elsevier*, last modified August 26, 2010, available online at <https://www.elsevier.com/legal/elsevier-website-terms-and-conditions> [accessed 10 July 2015].
41. Lessig, *Free Culture*, 298.
42. Lessig, *Free Culture*, 68–69.
43. "About RoMEO," *SHERPA/RoMEO*, available online at www.sherpa.ac.uk/romeo/about.php [accessed 7 December 2015].
44. Peter Suber, "Gratis and Libre Open Access," *SPARC Open Access Newsletter* (Aug. 2008), available online at www.sparc.arl.org/resource/gratis-and-libre-open-access [accessed 2 December 2015].
45. Hendrik P. van Dalen and Kène Henkens, "Intended and Unintended Consequences of a Publish-Or-Perish Culture: A Worldwide Survey," *Journal of the American Society for Information Science and Technology* 63, no. 7 (2012): 1282–93.
46. Ben Rowe, "The 4 H's of Writing Error Messages," *UXmas* (blog), December 9, 2012, available online at <http://uxmas.com/2012/the-4-hs-of-writing-error-messages> [accessed 9 July 2015].
47. "Framework for Information Literacy for Higher Education," *Association of College & Research Libraries*, last modified February 2, 2015, available online at www.ala.org/acrl/standards/ilframework [accessed 8 July 2015].
48. Patti Ryan and Lisa Sloniowski, "The Public Academic Library: Friction in the Teflon Funnel," in *Information Literacy and Social Justice: Radical Professional Praxis*, ed. Lua Gregory, 280. Sacramento, CA: Library Juice Press, 2013.
49. Yasmin Sokkar Harker, "Critical Legal Information Literacy: Legal Information as a Social Construct," in *Information Literacy and Social Justice: Radical Professional Praxis*, ed. Lua Gregory, 210. Sacramento, CA: Library Juice Press, 2013.
50. Tyler Neylon, "The Cost of Knowledge," *The Cost of Knowledge*, available online at <http://thecostofknowledge.com/> [accessed 9 July 2014].