Science AMA series: This is Daniel Himmelstein, PhD, and Casey Greene, PhD. We found that the Sci-Hub website has created a pirate repository of nearly all scholarly articles, which will push publishing towards more open models. Ask Us Anything!

eLife$_AMA$[1]$andr/ScienceAMAs$[1]

[1]Affiliation not available

April 17, 2023

Abstract

See the eLife flyer and this post for pictures! Daniel Himmelstein (@dhimmel on Reddit, Steem, and Twitter) – Hi Reddit! I'm a data scientist in Casey Greene's lab at the University of Pennsylvania. Before this, I got my PhD in Biological & Medical Informatics at the University of California, San Francisco. One reason I took the job at Penn (watch me accept the job on YouTube) was because I wanted to continue advancing open science – the idea that science will progress most quickly if research is immediately open without barriers to reuse and collaboration. Sci-Hub is a website that brands itself as the first pirate website in the world to provide mass and public access to tens of millions of research papers. It is a controversial form of open science, because it infringes upon the copyright of publishers. However, it's interesting because we think it will push scholarly publishing towards more open business models. Therefore, when Sci-Hub tweeted the list of every article in its database in March 2017, we began analyzing it openly on GitHub. Fast-forward almost a year and, after the publication of three preprint articles, we published our findings in the journal eLife with the title Sci-Hub provides access to nearly all scholarly literature. We also created a Stats Browser to help anyone explore the data. Casey Greene (@greenescientist on Reddit, Steem, and Twitter) – My research lab is in the Department of Systems Pharmacology and Translational Therapeutics at the University of Pennsylvania. Our primary focus is on developing machine learning methods to better understand human health and disease. I also run the Childhood Cancer Data Lab for Alex's Lemonade Stand Foundation, which is focused on integrating large-scale data to accelerate the pace of discovery. In addition to our research, I have an interest in the process of scientific communication, including our work studying Sci-Hub, our efforts to write a review paper entirely in the open via GitHub, and our biOverlay effort to launch an overlay for the life sciences. We're here to answer questions about our eLife paper, or our work more broadly. We'll start answering questions at 2pm EDT. AMA!

the WINNOWER

REDDIT

Science AMA series: This is Daniel Himmelstein, PhD, and Casey Greene, PhD. We found that the Sci-Hub website has created a pirate repository of nearly all scholarly articles, which will push publishing towards more open models. Ask Us Anything!

ELIFE_AMA R/SCIENCE

See the eLife flyer and this post for pictures!

Daniel Himmelstein (**@dhimmel** on Reddit, Steem, and Twitter) – Hi Reddit! I'm a data scientist in Casey Greene's lab at the University of Pennsylvania. Before this, I got my PhD in Biological & Medical Informatics at the University of California, San Francisco. One reason I took the job at Penn (watch me accept the job on YouTube) was because I wanted to continue advancing open science – the idea that science will progress most quickly if research is immediately open without barriers to reuse and collaboration.

Sci-Hub is a website that brands itself as the first pirate website in the world to provide mass and public access to tens of millions of research papers. It is a controversial form of open science, because it infringes upon the copyright of publishers. However, it's interesting because we think it will push scholarly publishing towards more open business models. Therefore, when Sci-Hub tweeted the list of every article in its database in March 2017, we began analyzing it openly on GitHub. Fast-forward almost a year and, after the publication of three preprint articles, we **published** our findings in the journal *eLife* with the title ***Sci-Hub provides access to nearly all scholarly literature***. We also created a Stats Browser to help anyone explore the data.

Casey Greene (**@greenescientist** on Reddit, Steem, and Twitter) – My research lab is in the Department of Systems Pharmacology and Translational Therapeutics at the University of Pennsylvania. Our primary focus is on developing machine learning methods to better understand human health and disease. I also run the Childhood Cancer Data Lab for Alex's Lemonade Stand Foundation, which is focused on integrating large-scale data to accelerate the pace of discovery. In addition to our research, I have an interest in the process of scientific communication, including our work studying Sci-Hub, our efforts to write a review paper entirely in the open via GitHub, and our bioRxiv effort to launch an overlay for the life sciences.

We're here to answer questions about our *eLife* paper, or our work more broadly. We'll start answering questions at 2pm EDT. AMA!

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

DATE RECEIVED:
March 22, 2018

DOI:
10.15200/winn.152163.36667

ARCHIVED:
March 21, 2018

CITATION:
eLife_AMA , r/Science , Science AMA series: This is Daniel Himmelstein, PhD, and Casey Greene, PhD. We found that the Sci-Hub website has created a pirate repository of nearly all scholarly articles, which will push publishing

I hate traditional publishers for being money hoarder as any other reasonable person. And I've even used sci-hub just for convenience when I was at home and I needed to read papers instead of waiting to be at the office.

Yet again, I really cannot accept the idea of journals going "Open Access". the idea that authors should pay to publish is simply unacceptable, and it would make research from "poorer" teams just impossible. Who on earth has 3000€ for a single paper??????

Also, open access is not going to fix the problem of outrageous publisher profits. It's just shifting the form according to which my lab ends up paying the same money, if not more, to Elsevier and Springer. How can we fix this?

lucaxx85

By study coauthor Thomas Munro

In response to lucaxx85's specific points:

towards more open models. Ask Us Anything!, *The Winnower* 5:e152163.36667 , 2018 , DOI: 10.15200/winn.152163.36667

I really cannot accept the idea of journals going "Open Access". the idea that authors should pay to publish is simply unacceptable, and it would make research from "poorer" teams just impossible.

This argument rests on several false assumptions:

1. All science is published in journals;
2. All open access (OA) journals charge publishing fees;
3. Authors pay all these fees themselves;
4. Paywalled journals do not charge publishing fees.

In fact,

1. Authors with little funding can publish preprints free of charge, as we did for this paper. To give a celebrated example, Perelman's proof of the Poincaré conjecture was only published on arXiv, not in a journal, but was universally acclaimed as a breakthrough.

2. Poor authors can publish free in the vast majority of OA journals, more than two thirds of whichdo not charge publishing fees such as APCs. These journals usually depend on institutional subsidies instead.

3. Most OA charges are paid by funding bodies or institutions, so there is no direct cost to the authors; see p. 9 of this article.

4. Many paywalled journals charge author-side fees. The most common is the print-era throwback of charging for color figures. While this can in principle be avoided by using monochrome figures, in practice color figures can be found in almost every article in prestigious journals, and the charges can amount to thousands of dollars per article, while the median APC in OA journals is zero, and the mean is less than a thousand dollars.

By far the highest publishing fees are charged by paywalled journals: reprint charges for medical journals. In some cases, drug companies pay millions of dollars to make an article freely available to doctors as reprints; these fees, and subscriptions for doctors paid by drug companies, make up a large part of the revenues for leading medical journals. By contrast, the highest APC of any OA journal is $5,200.

A good source on these and other myths is Peter Suber's classic book "Open Access".

Who on earth has 3000€ for a single paper?????? Also, open access is not going to fix the problem of outrageous publisher profits. It's just shifting the form according to which my lab ends up paying the same money, if not more, to Elsevier and Springer.

As noted above, the median APC in OA journals is zero, and the mean is less than $1,000. Meanwhile, the mean cost to society of a paywalled article is thousands of dollars, as Daniel noted. The maximum costs are also vastly higher for paywalled journals: as that article notes, "Philip Campbell, editor-in-chief of Nature, estimates his journal's internal costs at ... $30,000–40,000 per paper", even before their extremely high profit margin is added.

lucaxx85's questions themselves illustrate how paywalls raise costs, by allowing authors to externalize these ruinous costs to society: a vast public subsidy - tens of billions of dollars a year - of the concealment of publicly-funded research from the public. We argue that Sci-Hub is hastening the end of this grotesque situation.

I hate traditional publishers for being money hoarder as any other reasonable person. And I've even used sci-hub just for convenience when I was at home and I needed to read papers instead of waiting to be at the office.

Yet again, I really cannot accept the idea of journals going "Open Access". the idea that authors should pay to publish is simply unacceptable, and it would make research from "poorer" teams just impossible. Who on earth has 3000€ for a single paper??????

Also, open access is not going to fix the problem of outrageous publisher profits. It's just shifting the form according to which my lab ends up paying the same money, if not more, to Elsevier and Springer. How can we fix this?

lucaxx85

By study coauthor Daniel Himmelstein

Great point. Open access is not a panacea for problems with publishing and the outrageous cost of scholarly communication.

However, first we should note that not all peer-reviewed open access journals charge APCs (article processing charges that authors must pay to publish). In fact, only about 30% do. Nonetheless, in your field, you the OA journals may charge APCs. You should check to see if your university or library has allotted funds to cover open access APCs. Many have. In fact, many universities are going a step further and negotiating bulk deals with journals so all their members can publish OA for free. However, these deals are still in their infancy. Some funders are also willing to pay OA fees (without taking this amount out of the total grant).

But even so, publication should be cheaper. I think the reliance on journal prestige as a factor to assess scholarly achievement is a big reason there has been little pressure to make OA publishing cost-effective. Hopefully, soon scholars will start evaluating work based on its actual content rather than journal. When this happens, then you could preprint for free and never have to even engage with a journal.

A current project I'm working on is called Manubot. We're trying to build the most advanced publication system, and it's open source and free to use. Hopefully new tools like this, combined with preprints and more sophisticated article-level metrics, will bring competition and price elasticity to scholarly publishing.

How many dmca take down notices have you received? Do you'll have a legal fund that we can contribute towards?

xifu_mi

By study coauthor Casey Greene

We are not affiliated with Sci-Hub. We did not download the content that Sci-Hub contains. Instead, we analyzed the list of things that they contain. We compared that list against the CrossRef database of existing literature to determine what proportion of articles they contained.

We do not have a legal fund that you can contribute towards. However, if you liked this you may also find the Research Parasite awards interesting. We do have a way to donate to the Parasite Award via this Penn giving page. To my knowledge, we have raised ~$27k towards our goal and the match has been used up. I'm not sure why the page hasn't been updated, but I just put in a request to get it fixed.

How many dmca take down notices have you received? Do you'll have a legal fund that we can contribute towards?

xifu_mi

By study coauthor Stephen Reid McLaughlin

Sci-Hub and LibGen both accept BitCoin donations, but so far they haven't needed a legal fund per se ... because they haven't mounted a legal defense. They basically ignore all legal proceedings and takedown requests.

Last year a federal court in New York awarded the publisher Elsevier $15 million in a lawsuit against Sci-Hub and LibGen, and a court in Virginia awarded $5 million to the American Chemical Society in a suit against Sci-Hub. Neither is likely to see any of that money anytime soon.

Have you found a relationship between the number of citations of a particular study and the likelihood that it appears on SciHub?

adenovato

By study coauthor Daniel Himmelstein

Great question. We didn't look on a per-article basis, but did investigate on a per-journal basis. See Figure 9A. On the x-axis is CiteScore, "which measures the average number of citations that articles published in 2012–2014 received during 2015". The y-axis shows the mean coverage for all journals within a given CiteScore range. We comment:

Highly cited journals tended to have higher coverage in Sci-Hub (Figure 9A). The 1,734 least cited journals (lowest decile) had 40.9% coverage on average, whereas the 1,733 most cited journals (top decile) averaged 90.0% coverage.

In Figure 9B, we show that articles in highly-cited journals were downloaded more frequently in the Sci-Hub log data released for a sixth month period starting in late 2015.

Since Sci-Hub attempts to downloads articles when they're requested (if they're not already in its database), we'd expect higher coverage for highly-cited articles. We do see this. Sci-Hub has very high coverage of articles, when weighting by actual citations (rather than just random articles):

We identified 7,312,607 outgoing citations from articles published since 2015. 6,657,410 of the recent citations (91.0%) referenced an article that was in Sci-Hub. However, if only considering the 6,264,257 citations to articles in toll access journals, Sci-Hub covered 96.2% of recent citations. On the other hand, for the 866,115 citations to articles in open access journals, Sci-Hub covered only 62.3%.

was there any resistance (from publishers, or others) to getting this work published? have you gotten any pushback or criticism from publishing companies since publishing the paper?

kittttttens

By study coauthor Stephen Reid McLaughlin

One of my coauthors on another paper, Gabriel Gardner, got put on blast by the president of the Association of American Publishers for speaking positively about Sci-Hub. His institution went to bat for him, which is encouraging: https://www.insidehighered.com/news/2016/08/08/letter-publishers-group-adds-debate-over-sci-hub-and-librarians-who-study-it

Here's the paper we later wrote:
http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/2017/ShadowLibrariesandYou.pdf

was there any resistance (from publishers, or others) to getting this work published? have you gotten any pushback or criticism from publishing companies since publishing the paper?

kittttttens

By study coauthor Casey Greene

I was invited to speak at CrossRef Live this year, which a number of publisher representatives also attended. My interpretation of the Q&A at that meeting plus additional conversations is that major for-profit publishers appear somewhat consigned to a world where the value-add needs to be more than a paywall. This is not a universally held view, and I got the sense that there is much more resistance among society publishers. You can see the recorded Q&A here if you'd like (or the full talk by rolling back to the beginning).

I did not perceive any pushback on the publication of this work, but we did submit it toeLife which is an open access journal.

was there any resistance (from publishers, or others) to getting this work published? have you gotten any pushback or criticism from publishing companies since publishing the paper?

kittttttens

By study coauthor Daniel Himmelstein

I haven't personally received any push-back or criticism (that I'm aware of). In fact, I think a lot of publishers have found our analysis informative. A couple have even contacted me regarding questions about our Sci-Hub Stats Browser, which provides coverage information for each publisher.

In general, if someone emails me a question about the project, I ask them to post their question as a GitHub issue. Here's an interesting conversation we had on GitHub with Stuart Taylor, who's Publishing Director at The Royal Society. Stuart was questioning whether we overestimated the effect Sci-Hub will have on subscription publishing. However, the conversation was civil and productive.

At the end of the day, I think publishers will start switching to more open models. I think our study starts to make clear the necessity of this move. So smart publishers are looking at our study and using it to inform their strategic vision. Pushback (i.e. shooting the messenger) is the stupid strategy and risks igniting a Streisand effect.

What do you think about Sci-Bay?

https://sci-bay.org/

michaelhoffman

By study coauthor Daniel Himmelstein

Very timely question. I just heard of *Sci-Bay Scholar* yesterday and have not had a chance to really try it out. It's a new service whose domain was registered on 2018-03-15. I have heard, but not verified, that this site is "hosted in Singapore on generic cloud infrastructure". The description on the site reads "MORE THAN a combination of Google Scholar AND Sci-Hub! Google it. Download it. All in one site."

So in essence, it seems to be Google Scholar with links to Sci-Hub for articles. Note that Sci-Hub doesn't really provide much of a search engine. Users are expected to know the article before coming to Sci-Hub. So Sci-Bay Scholar integrates the Google Scholar search engine with quick links to Sci-Hub. Perhaps this will be convenient for many users? It also could also be a way to abstract the current domain names of Sci-Hub away from casual users (lots of non-technical users seem to have difficulty finding Sci-Hub domains when the domain they used gets suspended).

One note would be that Sci-Bay Scholar depends on two proprietary services (Google Scholar and Sci-Hub), neither of which have open APIs (as far as I know). Therefore, it'll be interesting to see whether it can persist.

Thanks for the AMA! Could you provide a short summary of your paper for those only browsing this thread?

useful_person

Good idea! Here's the abstract of our study:

The website Sci-Hub enables users to download PDF versions of scholarly articles, including many articles that are paywalled at their journal's site. Sci-Hub has grown rapidly since its creation in 2011, but the extent of its coverage has been unclear. Here we report that, as of March 2017, Sci-Hub's database contains 68.9% of the 81.6 million scholarly articles registered with Crossref and 85.1% of articles published in toll access journals. We find that coverage varies by discipline and publisher, and that Sci-Hub preferentially covers popular, paywalled content. For toll access articles, we find that Sci-Hub provides greater coverage than the University of Pennsylvania, a major research university in the United States. Green open access to toll access articles via licit services, on the other hand, remains quite limited. Our interactive browser at https://greenelab.github.io/scihub allows users to explore these findings in more detail. For the first time, nearly all scholarly literature is available gratis to anyone with an Internet connection, suggesting the toll access business model may become unsustainable.

The complete author list of the study (we've invited all the authors to join the AMA at 2 PM EDT) is:

Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene

Wait, what? You found it? Who pirated the articles? Is this Aaron Swartz's secret stash of JSTOR articles? If they're pirated from legitimate paywall science journals, how wil you succeed where he did not?

shiningPate

By study coauthor Stephen Reid McLaughlin

The cat is essentially out of the bag. Instead of downloading everything as fast as possible, Alexandra Elbakyan collected articles in a slow trickle, distributed geographically and over time. Even if publishers block her from downloading new articles tomorrow, she already has the vast majority of everything published to date, in JSTOR or any other commercial database.

As old domain names get taken down, she adds new ones: https://sci-hub.tw https://sci-hub.hk https://sci-hub.la

And the whole collection is mirrored separately: https://sci.libgen.pw

All the PDFs (a few dozen terabytes) are even available via BitTorrent. It's a pretty remarkable project.

Wait, what? You found it? Who pirated the articles? Is this Aaron Swartz's secret stash of JSTOR articles? If they're pirated from legitimate paywall science journals, how wil you succeed where he did not?

shiningPate

By study coauthor Daniel Himmelstein

First, I should clarify we're not affiliated with Sci-Hub in any way. We just analyzed publicly-available data and reported what we found.

Is this Aaron Swartz's secret stash of JSTOR articles?

I am not an expert on the Swartz case, but I don't believe he shared the JSTOR articles he downloaded on MIT's network. Note that the thirteen felonies he was indicted for did not include copyright infringement.

Who pirated the articles?

Sci-Hub was created by Alexandra Elbakyan, who I believe currently resides in Russia. A recentblog post of hers explains the history of Sci-Hub's repository:

Later in 2013 LibGen experienced problems with its hard drives, around 40,000 collected papers were completely lost. There was only one copy! I started a crowdfunding campaign on Sci-Hub to buy additional drives, and soon had my own copy of the database collected by LibGen, around 21 million papers. Around one million of these papers was uploaded from Sci-Hub, the other, as I was told, came from databases that were downloaded on the Internet/Darknet.

The list of Sci-Hub articles we analyzed from March 2017 contained ~63 million DOIs (digital object identifiers, i.e. article IDs). My understanding is that Elbakyan / Sci-Hub has downloaded the articles to grow the repository from 21 million to 63 million (and still growing). Sci-Hub seems to use a mix of credentials for institutions that subscribe to journals as well as directly infiltrating publisher systems to retrieve articles.

Hi Friends! I absolutely agree that access to papers is hugely important. Scholarly writing itself can be

so convoluted that it presents another barrier. How do you keep your scientific language accessible? - Holly

ucsc_treehouse

By study coauthor Daniel Himmelstein

Great point. Not only are articles super long usually, they tend to be written using a difficult-to-understand style. Sometimes difficult prose are necessary to precisely communicate an idea, but usually they are not.

In my experience, it's much easier to communicate in mediums that are more like conversations. For example, discussion over software development is accomplished primarily in a forum style, such as GitHub Issues. Software development rarely needs publications to communicate ideas. I'm excited for platforms that support more modular, interactive, small-scale scholarly communication. I'm not really sure traditional publications are necessary for many types of research.

What do you think the future of scientific publishing is?

What is your vision for an ideal publishing world, or at least some qualities you hope it will possess?

future_wombat

By study coauthor Daniel Himmelstein

Quoting from a recent development proposal we wrote for the Manubot, I want publishing to be: "transparent & reproducible, immediate & permissionless, versioned & automated, collaborative & open, linked & provenanced, decentralized & hackable, interactive & annotated, and free of charge."

Note that most of these goals are not possible with the current journal / static article framework. Hence, I hope that journals start playing a smaller and smaller role in scholarly communication.

What do you think about journals that charges a fee to the author in order to make the article open access? Is it a sustainable option?

pengrobinson

By study coauthor Casey Greene

The who-pays question in scientific publishing is a good one. I like to first think about what value the scientific publishing system provides.

Science is a method that we use to figure out how the world works. The process of disseminating results after peer review, which we currently call publishing, provides an opportunity for community comments to alter the trajectory of the research before it reaches an archived state (which, currently, is publication in a journal).

In the current system, costs are incurred at the level of infrastructure building (software, etc), maintenance (bandwidth, storage, etc), organization and interest assessment (professional editors, etc), peer review (though unpaid, this imposes a cost on the researchers and the institutions that employ them), article structure and maintainability (copyediting, typesetting, reference checking, etc). With for-profit publishers, there is also a cost to the research enterprise in the form of publisher profits. These funds are removed from the research ecosystem and returned to investors or owners.

For those costs, this system provides a distributed trust network that assesses the importance and correctness of individual contributions.

There are some things that worry me about for-pay publishing in this context. First, rejected manuscripts are lost profits. There’s no way around it. Some publishers (Nature, Elsevier with Cell Press) appear to be creating a broad swathe of journals that allow publications to “filter down” to their level of perceived importance. A manuscript might be submitted to *Nature* or *Nature Genetics* where it

is deemed to be of too little import, and it may filter down to *Nature Communications* or *Scientific Reports*. This approach may lead to a system where the journals that do not consider perceived importance also lose their ability to reject manuscripts that should be rejected because rejections at that stage are profits that are entirely lost to the publisher. The extent to which these pressures result in situations like a recent conflict at Scientific Reports is unclear but are things to keep an eye on.

However, there are also things that worry me about toll-access publishing. First, the work ends up locked behind a paywall and is generally inaccessible to the people who supported it. This also means that access may be provided only to wealthy researchers (or those from wealthy institutions or countries). In our study, we found that Penn - a very well funded research institution - has access to fewer papers than Sci-Hub, so it's clear that even our collection is not complete.

I think that the broad communications platform that is the internet provides new opportunities for scholarly communications. We no longer need to mail around hard copies of manuscripts to anonymous peer reviewers. I am hopeful that the proliferation of preprint servers which make literature available at no charge, combined with clear value-add from publishers that perform rigorous peer review and/or other services that improve the quality or rigor of the work, will allow incentives to become realigned in scholarly publishing.

We've started some experiments in this area. For example, we recently started bioOverlay to provide rigorous reviews of publicly shared preprinted work and highlight work that we found to be exceptionally interesting.

I hope that the publishing landscape in 5 years looks drastically different than it does now. In summary, I don't see the for-pay OA option in its current form as a long-term sustainable model. But it may be needed on an interim basis to get where we are going.

How many dmca take down notices have you received? Do you'll have a legal fund that we can contribute towards?

xifu_mi

Casey Greene here again - looks like this got asked twice. Answers are in the one above:

https://www.reddit.com/r/science/comments/861hzi/science_ama_series_this_is_daniel_himmelstein_phd/dw1s4rv/

Have you been tracking the instances of findings that were consistent vs disproven of the articles?

And do you feel this broad open sourcing will have a better confirmation of papers if a wider variety of people can read them, ie the backyard mathematician who likes to do it for fun reviewing papers vs the professor at a university?

powerlesshero111

By study coauthor Daniel Himmelstein

That's a difficult analysis to automatically extract findings of papers and whether other articles confirm or disprove them. We also didn't download the actual PDFs for articles in Sci-Hub (just the list of what articles were in Sci-Hub).

I do think widespread open access of the scholarly record will be beneficial to the public as well as science. There are lot's of individuals out there with the aptitude to rigorously engage with scientific reasoning. Erecting paywalls around scientific knowledge is a good way to ensure they never get started. One example is for rare diseases where patients, without a science background, end up becoming leading experts on their disease. See for example, the story of Sonia and Eric.

Hey guys!

Richard Stallman has pointed out that although unauthorized copying (known more commonly by the

slur "piracy") is not immoral (it is good to share things with your neighbor), it is still detrimental to free software culture as it continues the spread and reliance upon proprietary software.

Do you think that Sci-Hub will have a similar impact in the Open Science culture? Will there be less imperative to support open science initiatives if everybody already has access to proprietary knowledge?

Thanks!

miserlou

By study coauthor Daniel Himmelstein

Great question and this is a concern I've had as well. It'd be unfortunate if scholars were less motivated to publish open access because they reason that Sci-Hub will provide access regardless. Of course, Sci-Hub could go away or become censored in some jurisdictions. And open licensing of articles is extremely important. For text and data mining of articles to really blossum, we need openly licensed corpuses of articles. For example, I recently wrote a blog post that looked at 1.9 million articles. I would have done more... but the lack of open licenses were the prohibiting factor.

On the other hand, I think Sci-Hub will shift publishers to open access business models because subscription models will no longer be economically viable. So while scholars may care less about OA, their choices will be much more likely to be OA. Overall I think the shift in publishing models that Sci-Hub is triggering along with preprints and funder policies will more than compensate for a reduced need to publish OA so people can read your study.

Your paper clearly shows that Sci-Hub is a very effective solution to the lack of access to scholarly literature, but I think we need more than that. In the long term, we need Open Access: legal access and distribution of information; and also open and transparent infrastructures for scholarly publishing. I would like to hear your opinion on long-term effects of Sci-Hub: do you think Sci-Hub contributes to these goals? How does Sci-Hub affect the Open Access movement?

havre-kakor

By study coauthor Daniel Himmelstein

Absolutely, I think Sci-Hub contributes to the goal of having all publicly-funded research published under open licenses. In fact, this was my main motivation for doing the study, which I've stated before:

I think the larger picture of this study is that this is the beginning of the end for subscription scholarly publishing. I think it is at this point inevitable that the subscription model is going to fail and more open models will be necessitated. One motivation for doing the study is that I want to bring that eventuality into reality more quickly.

We go over why we think Sci-Hub will disrupt publishing towards more open models in thediscussion extensively, so I'll leave you with my favorite paragraph:

In the worst case for toll access publishers, growing Sci-Hub usage will become both the cause and the effect of dwindling subscriptions. Librarians rely on usage metrics and user feedback to evaluate subscriptions (Roth, 1990). Sci-Hub could decrease the use of library subscriptions as many users find it more convenient than authorized access (Travis, 2016). Furthermore, librarians may receive fewer complaints after canceling subscriptions, as users become more aware of alternatives. Green open access also provides an access route outside of institutional subscription. The posting of preprints and postprints has been growing rapidly (Piwowar et al., 2018; Kaiser, 2017), with new search tools to help locate them (Singh Chawla, 2017c). The trend of increasing green availability is poised to continue as funders mandate postprints (Van Noorden, 2014) and preprints help researchers sidestep the slow pace of scholarly publishing (Powell, 2016). In essence, scholarly publishers may have already lost the access battle. Publishers will be forced to adapt quickly to open access publishing models. In the words of Alexandra Elbakyan (Elbakyan, 2016b): “The effect of long-term operation of Sci-Hub will be that publishers change their publishing models to support Open Access, because closed access will make no sense anymore.”

I am also at the University of Pennsylvania, can we get an afternoon beer sometime soon?

plantsex

By study coauthor Daniel Himmelstein

Sure! DM me on Reddit.