James Heathers Follow
Scientist. Writer. Scallywag.
May 2 · 13 min read

Why Sci-Hub Will Win

Sci-Hub is a content aggregator for published academic work.

It's a portal where anyone can access the vast majority of academic publishing for free. It's 5 years old but only recently received worldwide attention.

It's also balls-out illegal, because the copyright for the articles it shares overwhelmingly belongs to large academic publishing companies.

According to Elsevier, the bloated Henry VIII at the top of the publishing heap, the site is "an international network of piracy and copyright infringement by circumventing legal and authorized means of access".

Well, yes. Elsevier may be a roc's nest of obese Poland Chinas, truffling around in a huge mudpit of public money, but facts is facts—piracy and copyright infringement is a perfectly reasonable description. In New York, last November, a court agreed.

As this is the internet, the ruling made no difference whatsoever, and the site lives. Probably buoyed by the publicity, actually. Vintage Streisand.

As of a few seconds ago.

Now they're a little better known, the game of Whack-A-Mole has started—that is, when one website is DNS blocked, a redirect will appear—www.sci-hub.cc, www.sci-hub.bz, and so on. There are probably more by now, either more Sci-Hub addresses at different extensions or shell sites.

A great deal has been written about Sci-Hub. It's illegal. It's irresponsible. It's ruining the Open Access movement. The creator of the site is a pioneer of academic civil disobedience. The creator is a criminal. It's Robin Hood-esque.

(People say this one a lot. Robbing plutocrats to help poor people doesn't have a lot of titular historical figures.)

I won't link to these opinions, because they don't matter. My opinion doesn't matter either. And neither does yours.

Regardless of what anyone thinks, Sci-Hub is going to win.

Here's why.

(I'm going to use the term 'piracy' here to describe the unauthorised access of copyrighted academic material in the same tradition of music and video piracy.)

Piracy already happens. A lot.

*Academics, researchers, scientists, etc. can generally access everything they need to read already … **eventually**.*

The key word is EVENTUALLY.

Barriers to accessing published academic work are really porous, and jourlans are actually very open to everyone who cares to look: people who work at small institutions, in foreign countries, in poor countries, people with bad library access, *everyone*. A publication needs to be really robustly obscure for it to be totally inaccessible to someone with no money and few contacts.

How do people get this access? Through an ad-hoc series of methods.

> **They ask each other.** This is the most common method of obtaining copyrighted content. Professor X writes an article on neurogenesis, you want to read it. You email Professor X. She sends you a copy. I have *never* found a scientist unwilling to share their own work. (Note: not answering your email because you are doddering old fool doesn't make you necessarily unwilling.)
>
> **They organise it formally on Twitter.** There is a hashtag for this, #icanhazPDF. You tweet the title of an article, and hashtag it with that —and some kindly soul sends you a copy privately. Also, probably the most significant time a cat in dire need of a cheeseburger has affected academic discourse.
>
> You can ferret out what you need on **Google**. Many, many institutions maintain publicly accessible uploaded copies of journal articles—these are generally for teaching, for lab access, or convenience, and generally not supposed to be public. But the Google spiders know no rest, and will archive them with the 'formal' search results if you are looking for the article.
>
> They find papers on **lab websites** (some laboratories thoughtfully archive all their published work on their own servers) or saved as teaching materials (some universities publicly accessible copies of different papers they wish to share which are relevant to teaching, journal clubs, etc.) Naturally, the all-seeing eye of Google usually finds these too.
>
> They form small communities of people who have **shared institutional**

access on various corners of the web (and in newsgroups). You can find them if you look hard enough.

The same as 5. but on **the Darknet**, where copyright is a forgotten joke.

They look for a **pre-print**. Some authors (quietly or otherwise) release a 'pre-review' copy of their work in various repositories or locations.

They **physically go to a university with access**. Many universities have their access available through public terminals or WiFi on campus. When you are physically located there, the gates are usually wide open.

They **ask their friend at Large University** to download it for them. This comes last, because it's the only one that inconveniences someone you know.

I'm sure you get the picture by now—basically, you rely on an informal network of a) people or places that have journal access and b) random places research can be publicly uploaded. This shabby ad-hoc network is pretty good the majority of the time for the majority of published work.

Basically, it's much easier to get ANYTHING with a digital copy, *no matter how rigidly it's paywalled or hidden*, than anything in print. Even ten years ago as a beginning Masters student, people would bitch with surprising vehemence about having to actually *physically go* to the library (it was about two city blocks distance from our office).

What's the real problem?

Lag.

Often, you don't need to read one crucial paper which will illuminate the problem that you're working on, you need to read twelve, or twenty, or a hundred. More often you need to find a paper, which cites a paper, which cites the paper you really need. That is, you don't need to read whole articles as much as have them direct you. You need a lot of access to a lot of journals to chase down the right foundational references, or to do exploratory reading through the spiraling branches of mutual references that exist between works.

Let's take an example and look at a paper of mine from a few years ago:

> ***Heathers (2013)—Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research***.

This is a fairly short paper: two straightforward experiments, a simple correction method, some reporting of parameters. It covers seven and a half pages of text, and it has *sixty-two* references.

If I had to access them individually without an institutional login, I cannot imagine how many emails, hashtags, random searches and furrowed brows would be involved. That is to say, I'm sure I COULD access almost all these references eventually in the absence of my University credentials—but it would take about as long as the study did. Not the same amount of effort, of course, but the same amount of time.

Scientists don't want to be librarians, forming our own curated collections of work we need to read. We want the information to be pre-curated, please, and available quickly. Which brings us to the worrying bit, if you're a publisher.

Piracy is a service problem...

The above is the first half of a quote by Gabe Newell (the internet's favourite deity). We'll get to the second half in just a minute. The quote continues:

> *For example, if a pirate offers a product anywhere in the world, 24 x 7, purchasable from the convenience of your personal computer, and the legal provider says the product is region-locked, will come to your country 3 months after the U.S. release, and can only be purchased at a brick and mortar store, then the pirate's service is more valuable.*

This was from an interview *five years ago*, but people in academic publishing seem to only now be grappling with this: when you throw away licencing, copyright, and digital rights issues, it's very easy to build a *superior* service.

Forget about morality for a second, and let's just look at the service here. Say I'm hypothetically in search of a journal article, whose name I know…

Sci-Hub

Google article, first page of results is 100% likely to contain PubMed or journal page.

Click relevant site.

Copy PubMed ID or DOI (digital object identifier; the ‘serial number’ of published online documents) to Sci-Hub main page.

Push enter.

Read the article.

Yes, that’s it.

My Institutional Access (mileage may vary)

Load library search page.

Click ‘journals’ in the OPAC.

Enter the name of the journal where the article resides.

Select a journal from the list presented.

Select a form of access to the journal (often this is provided from different databases, and you need to select the right once. For instance a “Legacy” collection may only access from 1977 to 2001, and a current collection may access from 1991 to present).

Insert your institutional login / password.

Wait while the hamsters in the proxy server shake off their sawdust, adjust their tiny, adorable trousers and start turning the wheels.

Insert the title name in the search bar of the journal, push enter.

Hope it works—these have a tendency to either a) reject queries for being too long b) reject queries for not being long enough (i.e. not recognising text you pasted into the search bar), c) throwing an error because you put in a ‘special’ character, such as a semi-colon, colon, question mark, hyphen etc.

Click the article if the search works. If not, browse through the journal tree (Year, Volume, Page Number) until you find the right research. Click. Science time.

Caveat: mobile login is similar, *only worse*. At least one screen in this chain will have no mobile website capability, and will go strange on you.

And I should conclude this by saying my institutional access is *GOOD*, and even my notoriously terrible attention span can suffer through this process intact most of the time.

And, let's put this in context of checking my references: **multiply the above by sixty two**. That's a lot of minutes.

So, yes, there's a 'cute' story here that has a human interest angle: one single solitary woman from Kazakhstan, with whatever assistance she can cobble together in her own time, has built a model which circumvents the closely-guarded highly contentious copyright of billion dollar academic companies wholesale.

But circumventing copyright is old news, academe is already collectively, quietly, and privately circumventing the hell out of copyright. *We don't make individual purchases.* When we need to read research we either a) use what we have been bought by a library, or b) find it ourselves.

What's important is that **Sci-Hub is a better system than yours**.

With its MS Paint frontpage and its funny load times and its presumably Kazakh server.

It's not significant that it *has access* to your material. It's significant that it *has simpler, more straightforward, ad-free, faster, more streamlined access*.

There was a hilarious post in Scholarly Kitchen a month or so ago which stayed with me, because it was so wonderfully bone-headed:

> *Sci-Hub has been part of the background radiation for many years, but about a month ago it jumped front and center on my radar screen. The occasion was that a young scientist told me that even though he has access to almost all the research papers he needs from his university library, he routinely uses Sci-Hub instead because of its more congenial user experience.* **I wasn't sure if he was**

> **pulling my leg or not (I think he wasn't), but his comment or wisecrack raises the interesting possibility that some of Sci-Hub's enormous traffic is generated by researchers who are indeed authorized to get access to the material they seek, but not at Sci-Hub.** *On reflection this is not surprising.*

On reflection?

What happened to these sage, sensible thinkers when other industries were having discussions about what piracy *meant*?

Home Taping Is A Crime

You Wouldn't Steal A Car

Ubisoft Goes Free-to-Play

etc. etc.

The fact this only just occurred to you? *That people will do what's easy and what they can get away with?* Don't even become a psychologist, I'm not sure your sage insights into the human condition will be able to cut it.

If we're this far into the digital age and your immediate response is to bet against human nature when the business environment changes, **you will lose so hard**.

What happens when they kill a torrent site or tracker? The site gets re-established with a different URL or cloned. (Note: as above, this has already happened.)

What happens when they kill Napster and Limewire? We get peer-to-peer sharing.

What actually makes a difference to piracy? Paid services which solve service problems. Spotify. Netflix. Hulu.

"But, we have a company like that! It's called Deepdyve!"

Deepdyve is a rental model for scientific papers. Access is much, much cheaper than it is to buy the content. And, like I said, it's a *rental* model. You

rent access to the paper, which is locked. In my field, rentals are normally $6.

And that's great.

IF your journal is available in Deepdyve (most of the big ones aren't).

IF you don't need 62 papers to write your own.

IF you're paid properly, so dropping '*just the price of a cup of coffee*' sixty-two times to write your own paper isn't a problem. Actually, I'm not sure what kind of Fair Trade civet-cat ultra-rare cup of Joe you're buying for $6, so it's *one hundred and twenty four cups* of coffee.

That's way too much coffee.

This embedded content is from a site that does not comply with the Do Not Track (DNT) setting now enabled on your browser.

Please note, if you click through and view it anyway, you may be tracked by the website hosting the embed.

Learn More about Medium's DNT policy

SHOW EMBED

Deepdyve is a Band-Aid over a sucking chest wound.

Now, let's address the rest of our previous quote:

Piracy is a service problem *not a pricing problem*...

Or at least, that's what Gabe Newell said about software.

That is, when you geo-lock software and fill it full of DRM and won't support problems with it and stuff consumers around, they'll pirate to *reduce hassle.* It isn't really a matter of saving money, it's the fact that the convenience of digital life was being ruined so copyright holders could maintain control.

But our problem here goes one step further, because in academe, it's now BOTH. Piracy is a service problem AND a pricing problem.

As covered above, journal databases are hard to access but they are also often a crying, bleeding shambles. It's frankly amazing that there's billion dollar companies maintaining them. They have 'on demand' updates that don't update. They frequently refuse to work at all. They have search bars which don't work. They have horrible user-unfriendly pages. They live behind complicated paywalls, in silos sorted by publishers.

And they cost a fortune. Bundling journals (selling them in pre-curated collections, not always things you need access to) and continuous price increases make the access for universities *really* expensive. It's the pricing that makes resources scarce for a lot of people—your libraries can't buy the official access, it's just not feasible. Here's a great book review about what Big Deal bundling has some for academic library budgets… it's gory.

Maybe we can update that quote: "*academic* piracy is a service problem *and* a pricing problem."

So, yes, a *better, free* service is probably a threat.

Conclusion

Imagine we ran the following poll on scientists:

Large academic publishers are:

a) well-meaning companies providing a crucial & good value service,

b) a necessary evil,

c) rent-seeking monsters pissing away public money like draft beer, artificially extending a business model that should be dead, standing in the way of human progress.

Can you guess the answers? It'd be unbearably ugly.

All the silliness written by defenders of large academic publishers in the last little while especially has really highlighted just how monumentally out of touch they are. They fail to understand a) the core of the problem, and b) the extraordinary and deep well of resentment that they've built up.

A good thought experiment here is *The Garbage Strike Test.*

Sanitation engineers / garbagemen / ashcart men—depending on your point in history—have a long tradition of industrial action. It's pretty easy to imagine how that works.

"Give us XYZ."
"No."
"RIGHT! NO MORE GARBAGE COLLECTION!"

And they just stop.

What happens? Almost immediately, massive stinking middens of rancid trash build up. Streets became partially inaccessible. Rats run rampant. Cities marinate in their own furious stink. Rocks are thrown at strike-breakers and scabs. Mayors call meetings.

After a few days, both sides are glaring at each other across a laminated chipboard table, reaching a compromise which will retrieve their urban landscape from looking like a Bruegel made of rotting cabbage and Go-gurt cups.

Most of us have no idea the systems that run behind the scenes, cleaning, scrubbing, mending, replacing, and rewiring the world around us, systems on which we are utterly dependent.

And when one of these systems makes the conscious decision to chuck us overboard and just stop working *until our demands are MET IN FULL!* it

really helps if they are:

a) truly necessary

b) on the right side of public opinion

c) something whose absence horrifies people utterly

Garbage collectors often meet all three conditions. Do academic publishers?

Let's say all large publishers suddenly refused anyone any access to any of their copyrighted materials at 9am tomorrow morning—what would they be replaced with?

The answer is a system which differs in *almost every respect* from the status quo, and one which would start seamlessly and immediately.

People would start to retrieve journal archives and torrent them. Universities would set up basic systems for access. Someone would write a Chrome util that linked your PubMed page to a torrent of the article. People with PMIDs would leave comments on each article to access them. Researchers with huge personal archives research stored in Endnote, Zotero, Papers, ReadCube, etc. would find a way to export them.

Would we lose pieces of information? Yes. But that would mean there were no copies anywhere else at all… making them, for the most part at least, less important than everything else. An awful lot of research is never read, let alone cited.

My bold prediction is in about two days, the whole thing would be strongly framed as an opportunity, and various calls for assistance in sticking back together our entire library of knowledge would travel over the whole planet.

In a fortnight, we would have quasi-formal channels of storing, disseminating, reviewing and publishing information.

In three months, they would be established, and serious steps would be taken to make sure these channels were never corporatised or exploited ever again.

In six to twelve months, legislation mandating that publicly funded research

must be publicly accessible at all times would become a lot more popular. The politically-minded love piling on an issue when it's a sure thing.

The whole thing would be a glorious mad scramble and a messy, wonderful, fraught, difficult opportunity that would be seized with single-minded aggressiveness.

That is NOT a garbage strike. That's the opposite.

Basically, massive academic publishers need to change or die, because a system for instantly disseminating research that's built out of driftwood and thumbtacks has sprung up, and the big news is not that it's free, it's that *it's better*.

****

My Medium (it has longer writing)

My Facebook (it has shorter writing and snark)

My Podcast (it has science)