

Who's downloading pirated papers? Everyone

By [John Bohannon](#)

Apr. 28, 2016 , 2:00 PM

Just as spring arrived last month in Iran, Meysam Rahimi sat down at his university computer and immediately ran into a problem: how to get the scientific papers he needed. He had to write up a research proposal for his engineering Ph.D. at Amirkabir University of Technology in Tehran. His project straddles both operations management and behavioral economics, so Rahimi had a lot of ground to cover.

But every time he found the abstract of a relevant paper, he hit a paywall. Although Amirkabir is one of the top research universities in Iran, international sanctions and economic woes have left it with poor access to journals. To read a 2011 paper in *Applied Mathematics and Computation*, Rahimi would have to pay the publisher, Elsevier, \$28. A 2015 paper in *Operations Research*, published by the U.S.-based company INFORMS, would cost \$30.

Related content:

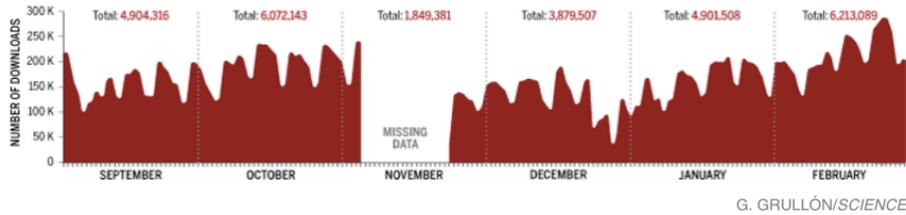
- [The frustrated science student behind Sci-Hub](#)
Alexandra Elbakyan founded Sci-Hub to thwart journal paywalls
- [My love-hate of Sci-Hub](#)
Editorial by Marcia McNutt, Editor-in-Chief, Science suite of journals
- [It's a Sci-Hub world data set](#)
Data set and details on Sci-Hub server

He looked at his list of abstracts and did the math. Purchasing the papers was going to cost \$1000 this week alone—about as much as his monthly living expenses—and he would probably need to read research papers at this rate for years to come. Rahimi was peeved. “Publishers give nothing to the authors, so why should they receive anything more than a small amount for managing the journal?”

Many academic publishers offer programs to help researchers in poor countries access papers, but only one, called Share Link, seemed relevant to the papers that Rahimi sought. It would require him to contact authors individually to get links to their work, and such links go dead 50 days after a paper’s publication. The choice seemed clear: Either quit the Ph.D. or illegally obtain copies of the papers. So like millions of other researchers, he turned to Sci-Hub, the world’s largest pirate website for scholarly literature. Rahimi felt no guilt. As he sees it, high-priced journals “may

be slowing down the growth of science severely.”

The journal publishers take a very different view. “I’m all for universal access, but not theft!” tweeted Elsevier’s director of universal access, Alicia Wise, on 14 March during a heated public debate over Sci-Hub. “There are lots of legal ways to get access.” Wise’s tweet included a link to a list of 20 of the company’s access initiatives, including Share Link.



But in increasing numbers, researchers around the world are turning to Sci-Hub, which hosts 50 million papers and counting. Over the 6 months leading up to March, Sci-Hub served up 28 million documents. More than 2.6 million download requests came from Iran, 3.4 million from India, and 4.4 million from China. The papers cover every scientific topic, from obscure physics experiments published decades ago to the latest breakthroughs in biotechnology. The publisher with the most requested Sci-Hub articles? It is Elsevier by a long shot—Sci-Hub provided half-a-million downloads of Elsevier papers in one recent week.

These statistics are based on extensive server log data supplied by [Alexandra Elbakyan, the neuroscientist who created Sci-Hub in 2011](#) as a 22-year-old graduate student in Kazakhstan. I asked her for the data because, in spite of the flurry of polarized opinion pieces, blog posts, and tweets about Sci-Hub and what effect it has on research and academic publishing, some of the most basic questions remain unanswered: Who are Sci-Hub’s users, where are they, and what are they reading?

For someone denounced as a criminal by powerful corporations and scholarly societies, Elbakyan was surprisingly forthcoming and transparent. After establishing contact through an encrypted chat system, she worked with me over the course of several weeks to create a data set for public release: every download event over the 6-month period starting 1 September 2015, including the digital object identifier (DOI) for every paper. To protect the privacy of Sci-Hub users, we agreed that she would first aggregate users’ geographic locations to the nearest city using data from Google Maps; no identifying internet protocol (IP) addresses were given to me. (The [data set and details](#) on how it was analyzed are freely accessible)

It’s a Sci-Hub world

Server log data for the website Sci-Hub from September 2015 through February paint a revealing portrait of its users and their diverse interests. Sci-Hub had 28 million download requests, from all regions of the world and covering most scientific disciplines.



Elbakyan also answered nearly every question I had about her operation of the website, interaction with users, and even her personal life. Among the few things she would not disclose is her current location, because she is at risk of financial ruin, extradition, and imprisonment because of a lawsuit launched by Elsevier last year.

The Sci-Hub data provide the first detailed view of what is becoming the world's de facto open-access research library. Among the revelations that may surprise both fans and foes alike: Sci-Hub users are not limited to the developing world. Some critics of Sci-Hub have complained that many users can access the same papers through their libraries but turn to Sci-Hub instead—for convenience rather than necessity. The data provide some support for that claim. The United States is the fifth largest downloader after Russia, and a quarter of the Sci-Hub requests for papers came from the 34 members of the Organization for Economic Cooperation and Development, the wealthiest nations with, supposedly, the best journal access. In fact, some of the most intense use of Sci-Hub appears to be happening on the campuses of U.S. and European universities.

In October last year, a New York judge ruled in favor of Elsevier, decreeing that Sci-Hub infringes on the publisher's legal rights as the copyright holder of its journal content, and ordered that the website desist. The injunction has had little effect, as the server data reveal. Although the sci-hub.org web domain was seized in November 2015, the servers that power Sci-Hub are based in Russia, beyond the influence of the U.S. legal system. Barely skipping a beat, the site popped back up on a different domain.

It's hard to discern how threatened by Sci-Hub Elsevier and other major publishers truly feel, in part because legal download totals aren't typically made public. An Elsevier report in 2010, however, estimated more than 1 billion downloads for all publishers for the year, suggesting Sci-Hub may be siphoning off under 5% of normal traffic. Still, many are concerned that Sci-Hub will prove as disruptive to the academic publishing business as the pirate site Napster was for the music industry (see [editorial by Marcia McNutt](#) on her love-hate of Sci-Hub). "I don't endorse illegal tactics," says Peter Suber, director of the Office for Scholarly Communications at Harvard University and one of the leading experts on open-access publishing. However, "a lawsuit isn't going to stop it, nor is there any obvious technical means. Everyone should be thinking about the fact that this is here to stay."

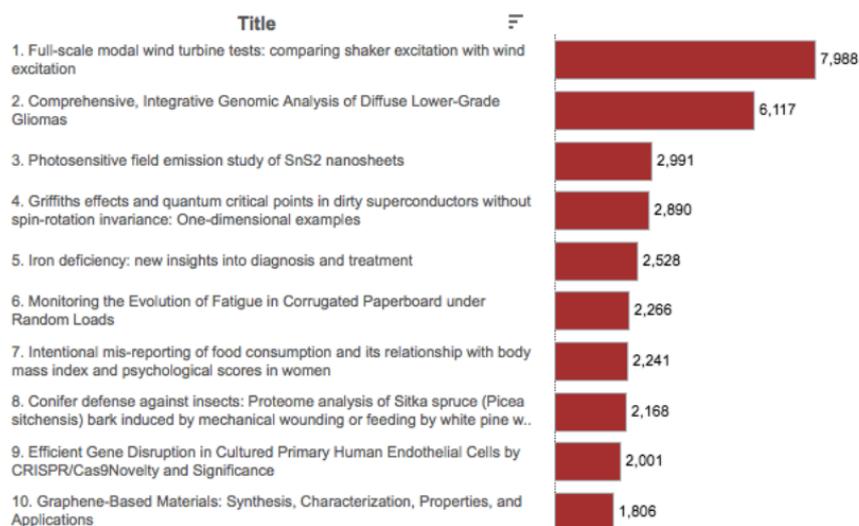
It is easy to understand why journal publishers might see Sci-Hub as a threat. It is as simple to use as Google's search engine, and as long as you know the DOI or title of a paper, it is more reliable for finding the full text. Chances are, you'll find what you're looking for. Along with book chapters, monographs, and conference proceedings, Sci-Hub has amassed copies of the majority of scholarly articles ever published. It continues to grow: When someone requests a paper not already on Sci-Hub, it pirates a copy and adds it to the repository.

Elbakyan declined to say exactly how she obtains the papers, but she did confirm that it involves online credentials: the user IDs and passwords of people or institutions with legitimate access to journal content. She says that many academics have donated them voluntarily. Publishers have alleged that Sci-Hub relies on phishing emails to trick researchers, for example by having them log in at fake journal websites. "I cannot confirm the exact source of the credentials," Elbakyan told me, "but can confirm that I did not send any phishing emails myself."

So by design, Sci-Hub's content is driven by what scholars seek. The January paper in *The Astronomical Journal* describing a possible new planet on the outskirts of our solar system? The 2015 *Nature* paper describing oxygen on comet 67P/Churyumov-Gerasimenko? The paper in which a team genetically engineered HIV resistance into human embryos with the CRISPR method, published a month ago in the *Journal of Assisted Reproduction and Genetics*? Sci-Hub has them all.

Sci-Hub's top 10 most downloaded papers

Of the 28 million documents requested from September 2015 through February, these were the most popular.



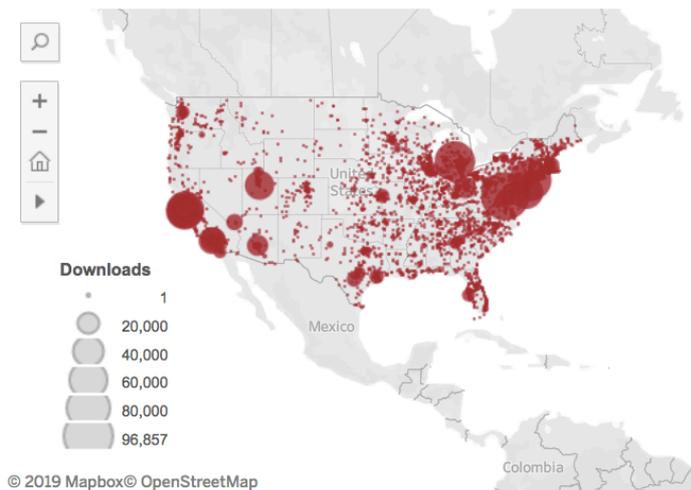
It has news articles from scientific journals—including many of mine in *Science*—as well as copies of open-access papers, perhaps because of confusion on the part of users or because they are simply using Sci-Hub as their all-in-one portal for papers. More than 4000 different papers from PLOS's various open-access journals, for example, can be downloaded from Sci-Hub.

The flow of Sci-Hub activity over time reflects the working lives of researchers, growing over the course of each day and then ebbing—but never stopping—as night falls. (There is an 18-day gap in the data starting 4 November 2015 when the domain sci-hub.org went down and the server logs were improperly configured.) By the end of February, the flow of Sci-Hub papers had risen to its highest level yet: more than 200,000 download requests per day.

How many Sci-Hub users are there? The download requests came from 3 million unique IP addresses, which provides a lower bound. But the true number is much higher because thousands of people on a university campus can share the same IP address. Sci-Hub downloaders live on every continent except Antarctica. Of the 24,000 city locations to which they cluster, the busiest is Tehran, with 1.27 million requests. Much of that is from Iranians using programs to automatically download huge swaths of Sci-Hub's papers to make a local mirror of the site. Elbakyan says. Rahimi, the engineering student in Tehran, confirms this. "There are several Persian sites similar to Sci-Hub," he says. "So you should consider Iranian illegal [paper] downloads to be five to six times higher" than what Sci-Hub alone reveals.

The geography of Sci-Hub usage generally looks like a map of scientific productivity, but with some of the richer and poorer science-focused nations flipped. The smaller countries have stories of their own. Someone in Nuuk, Greenland, is reading a paper about how best to provide cancer treatment to indigenous populations. Research goes on in Libya, even as a civil war rages there. Someone in Benghazi is investigating a method for transmitting data between computers across an air gap. Far to the south in the oil-rich desert, someone near the town of Sabha is delving into fluid dynamics. Mapping IP addresses to real-world locations can paint a false picture if people hide behind web proxies or anonymous routing services. But according to Elbakyan, fewer than 3% of Sci-Hub users are using those.

In the United States and Europe, Sci-Hub users concentrate where academic researchers are working. Over the 6-month period, 74,000 download requests came from IP addresses in New York City, home to multiple universities and scientific institutions. There were 19,000 download requests from Columbus, a city with less than a tenth of New York's population, and 68,000 from East Lansing, Michigan, which has less than a hundredth. These are the homes of Ohio State University and Michigan State University (MSU), respectively.



The numbers for Ashburn, Virginia, the top U.S. city with nearly 100,000 Sci-Hub requests, are harder to interpret. The George Washington University (GWU) in Washington, D.C., has its science and technology campus there, but Ashburn is also home to Janelia Research Campus, the elite Howard Hughes Medical Institute outpost, as well as the servers of the Wikimedia Foundation, the headquarters of the online encyclopedia Wikipedia. Spokespeople for the latter two say their employees are unlikely to account for the traffic. The GWU press office responded defensively, sending me to an online statement that the university recently issued about the impact of journal subscription rate hikes on its library budget. “Scholarly resources are not luxury goods,” it says. “But they are priced as though they were.”

Several GWU students confessed to being Sci-Hub fans. When she moved from Argentina to the United States in 2014 to start her engineering Ph.D., Natalia Clementi says her access to some key journals within the field actually worsened because GWU didn’t have subscriptions to them. Researchers in Argentina may have trouble obtaining some specialty journals, she notes, but “most of them have no problem accessing big journals because the government pays the subscription at all the public universities around the country.”

Even for journals to which the university has access, Sci-Hub is becoming the go-to resource, says Gil Forsyth, another GWU engineering Ph.D. student. “If I do a search on Google Scholar and there’s no immediate PDF link, I have to click through to ‘Check Access through GWU’ and then it’s hit or miss,” he says. “If I put [the paper’s title or DOI] into Sci-Hub, it will just work.” He says that Elsevier publishes the journals that he has had the most trouble accessing.

The GWU library system “offers a document delivery system specifically for math, physics, chemistry, and engineering faculty,” I was told by Maralee Csellar, the university’s director of media relations. “Graduate students who want to access an article from the Elsevier system should work with their department chair, professor of the class, or their faculty thesis adviser for assistance.”

The intense Sci-Hub activity in East Lansing reveals yet another motivation for using the site. Most of the downloads seem to be the work of a few or even just one person running a “scraping” program over the December 2015 holidays, downloading papers at superhuman speeds. I asked Elbakyan whether those download requests came from MSU’s IP addresses, and she confirmed that they did. The papers are all from chemistry journals, most of them published by the American Chemical Society. So the apparent goal is to build a massive private repository of chemical literature. But why?

A lawsuit isn’t going to stop [Sci-Hub], nor is there any obvious technical means. Everyone should be thinking about the fact that this is here to stay.

“ A lawsuit isn’t going to stop [Sci-Hub], nor is there any obvious technical means. Everyone should be thinking about the fact that this is here to stay. ”

Peter Suber, Harvard University

Bill Hart-Davidson, MSU’s associate dean for graduate education, suggests that the likely answer is “text-mining,” the use of computer programs to analyze large collections of documents to generate data. When I called Hart-Davidson, I suggested that the East Lansing Sci-Hub scraper might be someone from his own research team. But he

laughed and said that he had no idea who it was. But he understands why the scraper goes to Sci-Hub even though MSU subscribes to the downloaded journals. For his own research on the linguistic structure of scientific discourse, Hart-Davidson obtained more than 100 years of biology papers the hard way—legally with the help of the publishers. “It took an entire year just to get permission,” says Thomas Padilla, the MSU librarian who did the negotiating. And once the hard drive full of papers arrived, it came with strict rules of use. At the end of each day of running computer programs on it from an offline computer, Padilla had to walk the resulting data across campus on a thumb drive for analysis with Hart-Davidson.

Yet Sci-Hub has drawbacks for text-mining research, Hart-Davidson says. The pirated papers are in unstructured PDF format, which is hard for programs to parse. But the bigger issue, he says, is that the data source is illegal. “How are you going to publish your work?” Then again, having a massive private repository of papers does allow a researcher to rapidly test hypotheses before bothering with libraries at all. And it’s all just a click away.

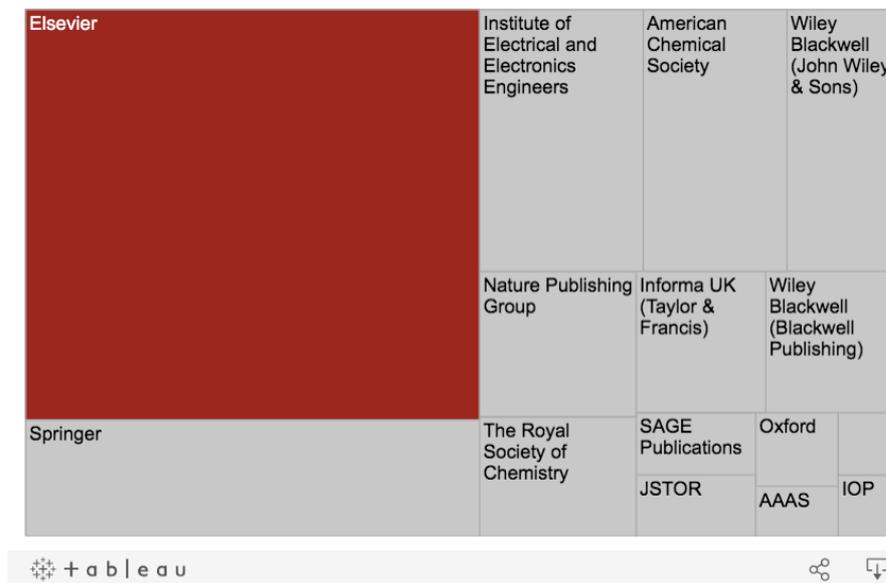
While Elsevier wages a legal battle against Elbakyan and Sci-Hub, many in the publishing industry see the fight as futile. “The numbers are just staggering,” one senior executive at a major publisher told me upon learning the Sci-Hub statistics. “It suggests an almost complete failure to provide a path of access for these researchers.” He works for a company that publishes some of the most heavily downloaded content on Sci-Hub and requested anonymity so he could speak candidly.

For researchers at institutions that cannot afford access to journals, he says, the publishers “need to make subscription or purchase more reasonable for them.” Richard Gedye, the director of outreach programs for STM, the International Association of Scientific, Technical and Medical Publishers, disputes this. Institutions in the developing world that take advantage of the publishing industry’s outreach programs “have the kind of breadth of access to peer-reviewed scientific research that is pretty much the equivalent of typical institutions in North America or Europe.”

And for all the researchers at Western universities who use Sci-Hub instead, the anonymous publisher lays the blame on librarians for not making their online systems easier to use and educating their researchers. “I don’t think the issue is access—it’s the perception that access is difficult,” he says.

Sci-Hub’s most downloaded publishers

This analysis of 28 million Sci-Hub download requests identified publishers via their Cross-Ref DOIs, which do not reflect the Springer-Nature merger.



“I don’t agree,” says Ivy Anderson, the director of collections for the California Digital Library in Oakland, which provides journal access to the 240,000 researchers of the University of California system. The authentication systems that university researchers must use to read subscription journals from off campus, and even sometimes on campus with personal computers, “are there to enforce publisher restrictions,” she says.

Will Sci-Hub push the industry toward an open-access model, where reader authentication is unnecessary? That’s not clear, Harvard’s Suber says. Although Sci-Hub helps a great many researchers, he notes, it may also carry a “strategic cost” for the open-access movement, because publishers may take advantage of “confusion” over the

legality of open-access scholarship in general and clamp down. “Lawful open access forces publishers to adapt,” he says, whereas “unlawful open access invites them to sue instead.”

Even if arrested, Elbakyan says Sci-Hub will not go dark. She has failsafes to keep it up and running, and user donations now cover the cost of Sci-Hub’s servers. She also notes that the entire collection of 50 million papers has been copied by others many times already. “[The papers] do not need to be downloaded again from universities.”

Indeed, the data suggest that the explosive growth of Sci-Hub is done. Elbakyan says that the proportion of download requests for papers not contained in the database is holding steady at 4.3%. If she runs out of credentials for pirating fresh content, that gap will grow again, however—and publishers and universities are constantly devising new authentication schemes that she and her supporters will need to outsmart. She even asked me to donate my own Science login and password—she was only half joking.

For Elbakyan herself, the future is even more uncertain. Elsevier is not only charging her with copyright infringement but with illegal hacking under the U.S. Computer Fraud and Abuse Act. “There is the possibility to be suddenly arrested for hacking,” Elbakyan admits. Others who ran afoul of this law have been extradited to the United States while traveling. And she is fully aware that another computer prodigy-turned-advocate, Aaron Swartz, was arrested on similar charges in 2011 after mass-downloading academic papers. Facing devastating financial penalties and jail time, Swartz hanged himself.

Like the rest of the scientific community, Elbakyan is watching the future of scholarly communication unfold fast. “I will see how all this turns out.”

The survey on attitudes toward Sci-Hub is now closed—[here are the final results](#).

***Correction, 12 May, 4:36 p.m.:** *The two GWU students noted in the story are pursuing engineering Ph.D.s, not physics.*