
Kluwer Copyright Blog

The science of piracy, the piracy of science. Who are the science pirates and where do they come from: Part 1

Balázs Bodó (Institute for Information Law (IViR)) · Wednesday, March 6th, 2019

Introduction

In 2016, Science published a [short report](#) on the usage of SciHub, a piratical scholarly journal article distribution service. Set up by Alexandra Elbakyan, a kazakhstani scientist, SciHub allows users to bypass journal publishers' paywalls, so everyone can have access to journal articles for free. The report, based on a [dataset provided by Elbakyan](#), offered a stunning insight into the underground circulation of scholarly knowledge. The colourful maps made it clear that it is not just developing countries that seem to struggle with access issues: high income European and North American countries are also eager pirates of scholarly articles.

The topic of this post is a closely related phenomenon: the underground circulation of scholarly books. Using a dataset provided to us by one of the administrators of a prominent shadow library, we mapped both the supply of and the demand for academic monographs, textbooks and other learning materials via piratical shadow libraries. Our primary findings suggest that scholarly book piracy is a ubiquitous global phenomenon, with no apparent end in sight. If that is indeed true, what might the consequences be for the status quo in scholarly publishing?

Shadow libraries

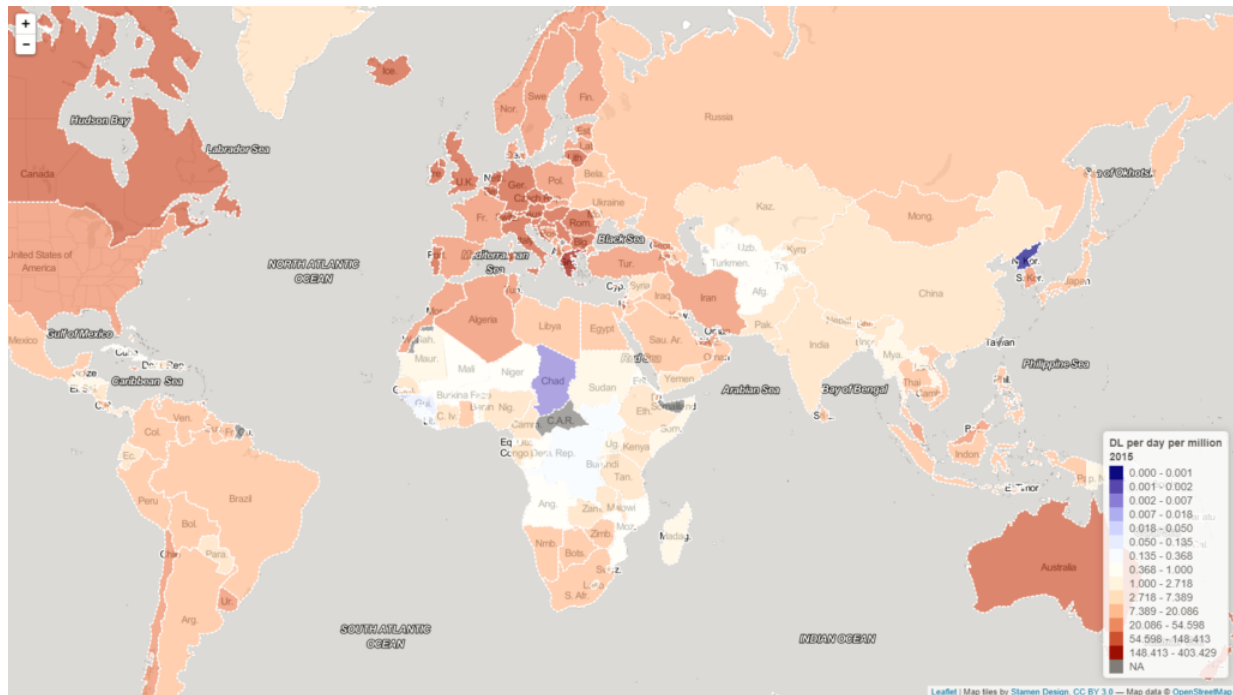


Figure 1: The number of downloads of scholarly books per day per capita (million) via a shadow library in 2015

One of the major developments during the 2010s in scholarly publishing was the emergence of so-called shadow libraries – illegal archives of scholarly journal articles, textbooks, monographs, and other forms of academic work. Services like [Gigapedia](#), [LibGen](#), [SciHub](#), [Aaaaarg](#), [monoskop](#), [Hansi](#), and [chomikuj](#) provide free, unrestricted, often copyright infringing access to more than a hundred million pay-walled scholarly articles and millions of books for anyone interested.

While the piracy of science is nothing new, online science piracy is a unique product of the conditions of global higher education systems' development in the last half century. The online science pirate libraries constitute only a relatively small part of much larger, informal human and institutional networks through which scholarly works circulate. A recently published study: “[Shadow Libraries – Access to Knowledge in Global Higher Education](#)” gives an in-depth insight into the different informal practices which facilitate the often-illicit circulation of learning materials in emerging economies. The study lays out the contexts which contributed to the emergence of scholarly piracy: the rapid expansion of tertiary education in the world and the subsequent growth in the demand for books; the post-WWII concentration of the western academic publishing market and the access barriers set up by publishing monopolies; and the innovative solutions with which scholars try to bridge the gap between the supply of and the demand for scholarly works.

Supply and demand in global academic publishing

In the decade between 1995 and 2005, the number of people with post-secondary education grew from 283 million to 725 million. As Joe Karaganis, the editor of the aforementioned study, remarks, much of that growth originated in the developing world: India, Brazil, South Africa, and Central and Eastern European Post-Soviet countries. Unlike in the previous growth period, when the post-WWII Western countries increased their higher education sectors and their research capacities through huge public investment, this time the education boom took place in middle- and low-income countries, and was much less reliant on public funds. Consequently, in the last few

decades, hundreds of millions of people have tried to gain access to a global knowledge commons, while the infrastructural conditions for such participation, in the form of well stocked, easily accessible libraries, were less than ideal.

The rapid, global growth in the demand for scholarly works, and the tightening financial conditions of higher education coincided with a rapid concentration and commercialization of Western scholarly publishing. The post-WWII boom in western tertiary education radically expanded the size of the scholarly publishing market, which led to a series of mergers and acquisitions, where commercial publishers started to consolidate smaller, non-profit scientific publishers into a handful of strong, vertically integrated oligopolies. The same companies introduced innovations, such as different Science Citation Indices, which led to the development of a few highly sought after, therefore very valuable, journals in many disciplines. These journals, in turn, became powerful points of control in scholarly publishing. The publishers who control such key resources are able to charge excessive access fees despite the fact that every other input for these journals (the articles themselves, peer review, etc.) are provided by the academic community for free.

These twin developments of rapidly rising costs and rapidly rising demand coincided with the widespread availability of increasingly cheap reproduction technologies: first the photocopy machine, and later digital technologies. On these technological platforms a plethora of practices evolved that tried to bridge the gap between supply and demand. The Open Access initiatives in the 1990s created the standards of green and gold open access, and of self-archiving. The rapid proliferation of copy shops on and around the campuses provided copies at the marginal cost of physical reproduction. Parallel markets for scholarly works emerged through the re-importation of learning materials and textbooks from low-income countries, where they were sold at a discount, into high income countries. And, finally, online shadow libraries emerged that digitized, archived and distributed books online.

The emergence of online shadow libraries

Personal computers and digital networks provide an ideal environment for the creation, accumulation and circulation of texts in digital form. One of the earliest online projects is in fact a digital library from 1971, the Gutenberg Project, which set out to digitize public domain cultural heritage. Copyright issues were able to limit bottom-up digitization and sharing of copyrighted works for a while, however when copyright wars turned the legal struggle around intellectual property protection into a war on moral values, some in academia started to see IP infringement as a morally justifiable act of resistance against unjust publishing monopolies. In 2008, Aaron Swartz, a US activist, published his *Guerrilla Open Access Manifesto*, in which he called for the liberation of pay-walled scholarly content to show solidarity with scholars with no access. Around the same time, various text archives, which were digitized and compiled by Russian research institutions in the early 2000s and which up till then had circulated on DVDs and ftp servers, started to consolidate and appear as the online shadow library *Library Genesis* (LibGen).

Aaron Swartz and Alexandra Elbakyan gave a name and a face to the countless anonymous individuals who maintain the global shadow library ecosystem. Swartz represents the privileged Western scholars, the insiders, who have access to almost everything through their first class, well-endowed academic libraries. Some of these scholars recognised their privileged position, and decided to show solidarity with others by sharing their digital access opportunities. They are the ones who smuggle the knowledge out from behind the paywalls. On the other end, we find Elbakyan, who represents scholars at the very peripheries of privilege, wealth and access. Such

scholars are outsiders. They are the ones on the wrong side of the access paywalls. They tend to live in countries, like Russia, which have rich histories of highly efficient clandestine knowledge distribution networks to circumvent political oppression, economic hardships, evade enforcement, and build influential underground knowledge repositories under hostile circumstances. They put their historic experiences to use, and build and run the infrastructures necessary to archive and distribute the copyright infringing content.

Through the collaboration between scholars at the centre and at the periphery, powerful shadow libraries now facilitate a historically unprecedented transfer of knowledge across the globe.

The users of shadow libraries

Based on publicly available data on the catalogue and privately shared data on the usage of one of LibGen's mirrors we were able to shed some light on who uses shadow libraries and for what purpose. In 2012, the Library Genesis catalogue contained 836,479 records. Three years later, in 2015, the catalogue had almost doubled to 1,317,424 records, and by the time of writing in 2018, Library Genesis hosted more than 2,237,940 documents, almost all scholarly publications [with a particular focus on the western scholarly canon](#). In addition, there is an extensive collection of literary works, comics, and of course the 100 million journal articles archived through the SciHub.

There are multiple websites which make this catalogue available. Although the exact number fluctuates as a result of various legal, technical, organizational and financial issues, at least one service has been consistently online for the last eight years. The anonymous administrators of this single mirror provided us with usage logs for two months in 2012 and for four months in 2014/2015. In these three years the traffic on this single server tripled, from an average of 41,000 downloads per day in 2012 to an average of 120,000 downloads per day in 2015.

Where does this growth come from? The countries and regions that account for the bulk of the usage (US, India, China, Europe) show average growth, or as in the case of Russian speaking countries, outright decline, as Russian users seem to have migrated to other services. On the other hand, there has been a staggering growth in Latin America, which in 2012 was hardly using (this particular mirror of) Library Genesis at all, but by 2015 had become one of the most intensive users of the library. These rapid, and violent changes in the use of shadow libraries are the signs of a systemic, unserved demand for knowledge, which is constantly seeking the venues which can satisfy its need for access.

Region	Daily downloads per million inhabitants (2012)	Daily downloads per million inhabitants (2015)	Growth (%)
Europe & Central Asia	28.81	42.87	149%
Middle East & North Africa	8.41	24.31	289%
South Asia	1.22	6.38	523%
Latin America & Caribbean	0.06	11.68	19467%
Sub-Saharan Africa	0.28	3.98	1421%
East Asia & Pacific	2.13	7.4	347%
North America	10.47	42.76	408%

Table 1: Per capita downloads in 2012 and 2015 per region

On the other hand, it is also apparent that the biggest per capita users are the high income North American and European countries.

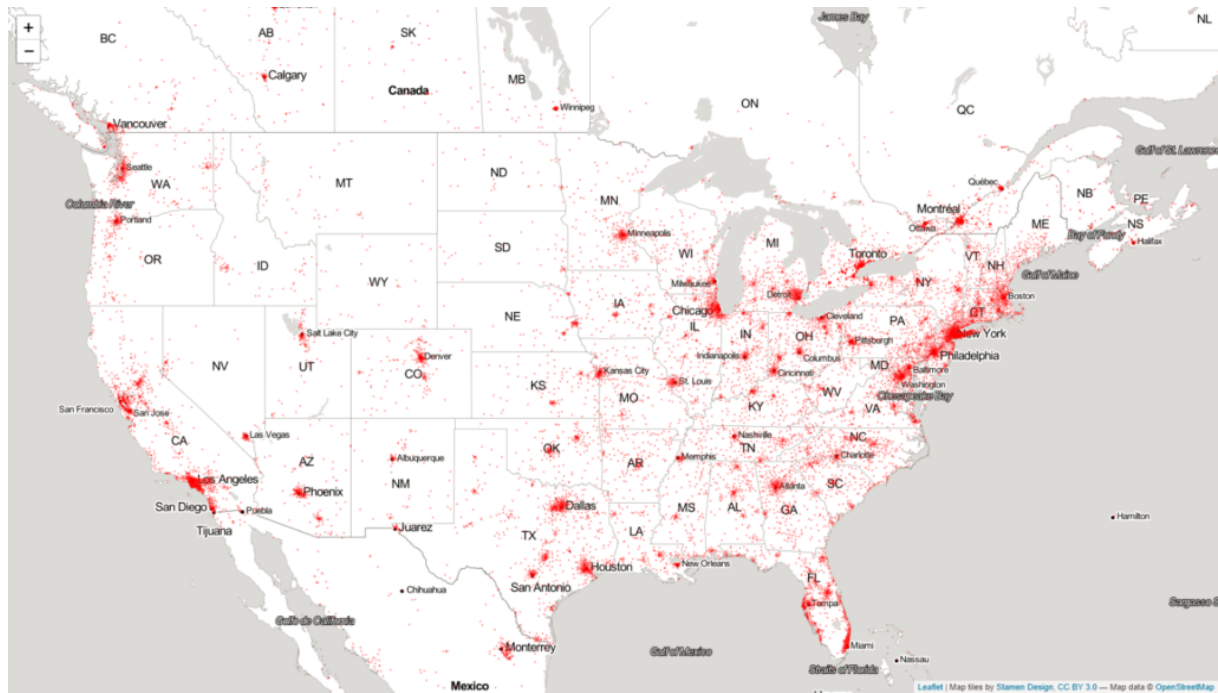


Figure 2: North America

ican download locations

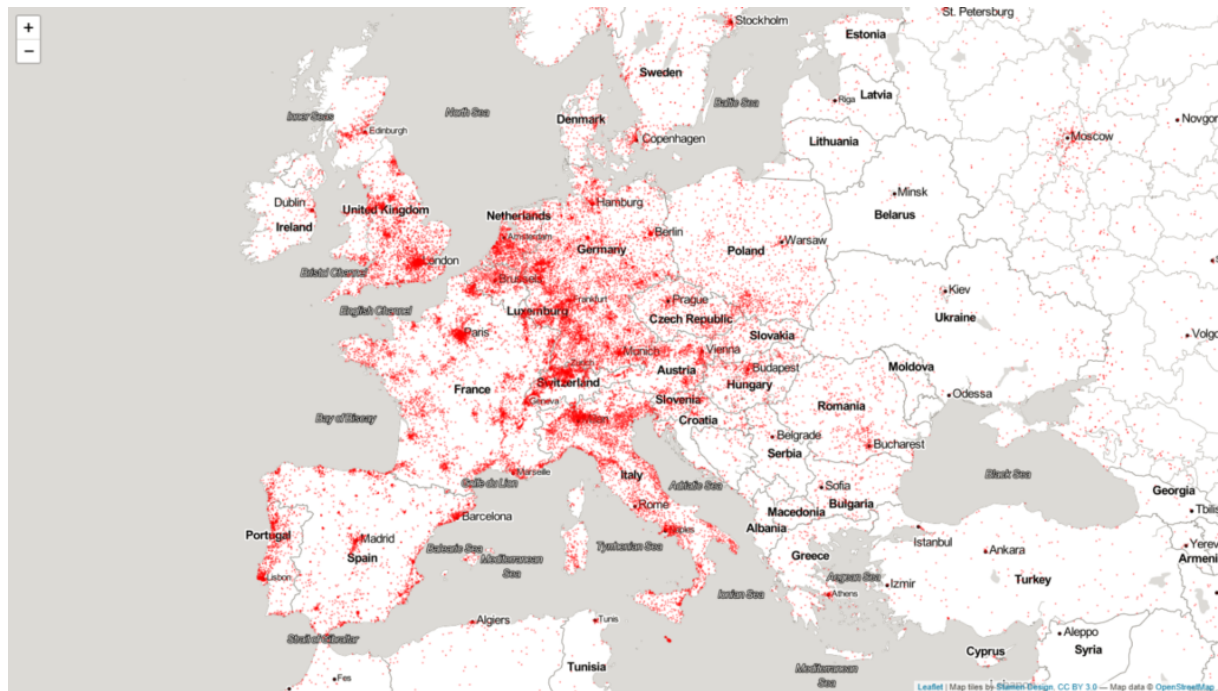


Figure 3: Europe

ownload locations

In fact, just a handful of countries – the United States (11.66%), India (8.58%), Germany (5.23%), the UK (4.10%), Iran (3.68%), China (3.67%), Italy (3.30%), Canada (2.36%), Indonesia (2.29%), Spain (2.28%), Turkey (2.24%), and Brazil (2.11%) account for more than half of all the downloads. Animated download maps are available for the [US](#), [SE Asia](#), [South America](#) and [Europe](#).

The fact that a large proportion of science pirates come from rich Western countries, with well-

financed higher education systems, dense library networks, and well-organized markets should give us reason to pause. Where are these downloads coming from? Lazy academics who find their institutional access too cumbersome, and turn to shadow libraries for their ease of use? Is there a substantial demand for science from beyond the walls of academic institutions? Would libraries be able to serve this demand, if they had the chance to lend such books electronically? Would individuals buy such books, even if they were on sale in electronic formats, for a reasonable price? What kind of effect does this type of piracy have on the future of scholarly publishing?

Part 2 of this post will analyse the nature of the works downloaded and discuss the implications of shadow libraries for the future of scholarly publishing.

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please [subscribe here](#).

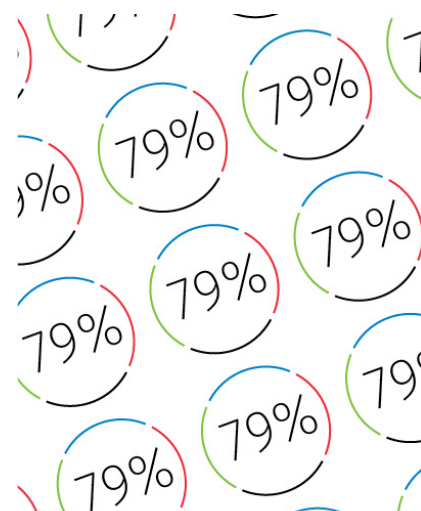
Kluwer IP Law

The **2022 Future Ready Lawyer survey** showed that 79% of lawyers think that the importance of legal technology will increase for next year. With Kluwer IP Law you can navigate the increasingly global practice of IP law with specialized, local and cross-border information and tools from every preferred location. Are you, as an IP professional, ready for the future?

Learn how **Kluwer IP Law** can support you.

79% of the lawyers think that the importance of legal technology will increase for next year.

Drive change with Kluwer IP Law.
The master resource for Intellectual Property rights and registration.



2022 SURVEY REPORT
The Wolters Kluwer Future Ready Lawyer
Leading change

This entry was posted on Wednesday, March 6th, 2019 at 3:50 pm and is filed under [European Union](#),

Making available (right of), USA

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.