Pirates in the library –
an inquiry into the guerilla open access movement

Balazs Bodo (bodo@uva.nl), Institute for Information Law, University of Amsterdam

Abstract

2016 is the year when piracy finally became an unavoidable topic in the domain of scholarly communications. The public exposure of Sci-Hub, a copyright infringing site that provides free access to paywalled journal databases, electrified the decade old debates about the role of scholars, (commercial) publishers, libraries, and copyright in creating an environment, where results of scholarly inquiry are equally accessible for all.

This article tries to give an insight into the Guerilla Open Access (GOA) movement, which is responsible for the creation and maintenance of massive, copyright infringing, freely accessible online shadow libraries of scholarly works: journal articles, monographs, textbooks. It reconstructs the developments in the western and global academia and scholarly publishing which led to the birth of the movement, and identifies some of the factors its ongoing existence depends on.

The article discusses several aspects of the GOA movement: the alliance of scholars in the global centers and at the global peripheries, the alliance of public and clandestine operations, and its relationship with, and its differences from the Open Access (OA) approach, which aims to facilitate the accessibility of scholarly communications through legal means.

The goal of this article is to contribute to the discussions of the future of scholarly communications through the description of a phenomenon which poses the single greatest challenge to the scholarly publishing status quo in recent history.

KEYWORDS: piracy, Guerilla, Open Access, Sci-Hub, Library Genesis, shadow libraries, Aaron Swartz, Alexandra Elbakyan

Introduction

Sci-Hub is a copyright infringing service that provides unauthorized backdoor access to proprietary scholarly journal databases. Its primary aim is to help researchers, who do not have institutional access to digital journal databases, and who are unwilling or unable to pay the per article access fees. The site is one of the multiple shadow libraries that provide copyright infringing access to scholarly communication (Bodó, 2015b). It gathered widespread attention after a young Kazakh scientist, Alexandra Elbakyan, who was identified as the owner of the website, mounted a fierce public opposition to a lawsuit filed by Elsevier against the site and herself in a New York court (Elsevier Inc. et al v. Sci-Hub et al Case No. 1:15-cv-04282-RWS).

Rather than lying low, or focusing on a legal defense, Elbakyan responded to the court case with a highly visible public campaign against the publisher who sued her (Elsevier), and the business practices currently dominating the scholarly publishing market (Elbakyan, n.d.). Her arguments revolve around a few key points: scientific knowledge should be freely accessible to everyone; the business model of journal publishers, which is based on selling subscriptions to proprietary databases is not only unethical, but it is also highly damaging to science and society; and while her actions may be illegal, she is engaged in a just fight against greedy corporate powers and those legal frameworks that enable such abuses. While these arguments are not novel, Elbakyan is unique in the sense, that she is one of the very few individuals who, for reasons to be discussed later in this article, decided to defend their private, copyright infringing actions in the public, and turn a court case on copyright into a political struggle around access to knowledge, culture, and information.

Most Open Access Guerillas try to avoid exposure, and prefer to operate under the radar to avoid prosecution. Most critiques of the Intellectual Property status quo are keen to be seen in full compliance with the law. Elbakyan, like Aaron Swartz, and a number of others, offer their critique of copyright and scholarly publishing in the form of blatantly illegal services. This paper explores this Guerilla Open Access movement and offers an analysis of the challenges and the dilemmas they face.

A note on ethics and methods

I approach the topic in multiple capacities. First, I am obviously a researcher, which requires me to keep an emotional and analytical distance from the topic. But in my capacity as a scholar, I cannot pretend to be an innocent bystander to the struggles around scholarly publishing and Open Access. My decisions on where I publish, and how I make my work available force me to “take sides” in those debates, whether I like it or not. During my career I also spent substantial amounts of time on both sides of the access divide. As a scholar working in Eastern Europe, I also had to think hard about how to conduct competitive research and provide competitive degrees to my students without having access to the same texts as my Western European colleagues. This background explains why, at one point I stepped beyond the narrowly defined role of a researcher, and for many years I have been participating in the debates around the accessibility, production, distribution of culture as an activist, propagating the use of Creative Commons licenses, and other legal and technological tools of Open Access.
My association with shadow libraries should also be known. In fact, I am a signatory to an online call for solidarity for the two subjects of this paper, Sci-Hub and LibGen (Barok et al., 2015). This close proximity to the subjects of this inquiry has both its advantages and its risks. Shadow librarians are a geographically, culturally and linguistically dispersed, reclusive community, engaged in copyright infringing activities. Facing immense personal legal, financial risks, shadow librarians are highly protective of themselves, their communities, their individual identities. The trust I seem to enjoy in some of these communities enabled me to understand the values, motivations, acts of these communities in ways not usually available for outsiders, even if this extra depth comes with its own set of ethical and methodological challenges (Davis, 1991; Tenen & Foxman, 2014). On the other hand, being too close, may have a number of adverse effects on the quality of my scholarship. The biggest challenge for me, however was not how to remove myself from the description, but how to identify the blind spots in the “thick description” (Geertz, 1994) I attempted to provide here. Given the limited length of this contribution, probably there will be more than one aspects missing from this account. I tried to ensure that they are not the result of any systematic bias in my approach.

This article covers the piracy of both journal articles, such as the activities and politics of Sci-Hub, and the piracy of scholarly monographs, textbooks, readers, and other printed materials, as done by Library Genesis (LG) (Bodó, 2015b). The economics, organization and history of scholarly journal publishing are quite different from how academic presses operate, and these differences would warrant a much more nuanced, if not completely separate analysis of the two. There are, however, a few reasons I felt comfortable with treating the two domains as mostly interchangeable for the purposes of this analysis. First, despite being two distinct fields, they both serve the same audiences, rely on the same author pool, and sell to the same set of buyers: academic institutions and their libraries. This means that they are intrinsically linked: “the journal crisis, concentrated in the sciences, has precipitated a monograph crisis, concentrated in the humanities.” (Suber, 2013, p. 33). Second, scientists struggle with serious access issues in both domains. Third, while LG, the shadow library that serves the underground book market is organizationally and infrastructurally separate from Sci-Hub, they are co-defendants in the Elsevier lawsuit. And finally, while their politics may show substantial differences, these two types of piratical services are part of the same struggle, and offer similar solutions to similar problems. As such, they constitute two distinct, but closely related elements of a wider Guerilla Open access movement, which uses piracy as a political tool to address to systemic failures of scholarly publishing.

Pirates in the (research) library

One would have thought that due to the ease of copying texts, the printed word would be the first to be affected by digital piracy. Instead, more than 16 years after Napster, there is little evidence of

---

1 Use the word pirate and piracy to refer to the willful infringement copyright. Though I am aware of the debate around the appropriateness to use this word to describe people who copy and/or share information (Stallman, 2002), I am not convinced that this term’s meaning is closed and without any ambiguity. Legitimate political parties use the term in their names (Burkart, 2014), and it is sold as a lovable opponent to relentless and aggressive capitalist expansion in family entertainment media products (Verbinski, 2003, 2006). With the conscious use of this word I point to this ambiguity of meaning, and the openness of interpretations.
widespread, well organized e-book piracy networks (Flood, 2015; Meister, 2013; Pogue, 2013; Reimers, 2014), apart from the scholarly shadow libraries. While the few piratical libraries that offer literary works, bestsellers and the likes are usually small, disorganized, fragmented services, of low technical quality, at the very edge of the publishing markets, there are multiple well-organized underground text collections which collect and make accessible scholarly publications: scientific monographs, textbooks, journal articles by the millions. (Bodó, 2015a, 2015b; Cabanac, 2015) The rather late appearance, and highly specific nature of e-book piracy raises three closely related questions: why scholarly publishing is affected and fiction is not? Why did it take so long for pirate to appear? And, being so late, why did they appear at all?

The limits of this article do not allow me to fully explore the first question, but answering the latter two is essential, if we want to understand the full extent of the current scholarly piracy crisis.

The commercialization of western scholarly publishing

“In the last 50 years, publishers have managed to transform scholarly journals—traditionally, a secondary, unpromising publishing venture at best—into big business. What is the real basis behind this astounding capability? What is the source of their power? How can it be subverted?”

(Guedon, 2001, pp. 1–2)

The development of scholarly communication can roughly be divided into two phases. In the first phase print was the dominant medium, and various geographically, linguistically separated local markets developed in relative separation from each other. The beginning of the second phase is marked by digital technologies, globalized communications networks, and an increased level of globalization in science and higher education in general. Pirates played different roles in these two phases. In the following section I attempt to reconstruct how piracy was shaped by the specific conditions of the markets in each phase.

The domain of scholarly publishing is a unique market in the sense that both the demand and the supply of scholarly publications is rather insensitive to changes in price. Since authors publish for non-monetary rewards, they are willing to supply the input for the publishing industry for free. On the other hand, particular articles constitute an often non-substitutable input for research and education. This makes demand for these works also very insensitive to what they cost, especially, that few readers actually pay directly for the works they consume, as in most cases it is their institutions and/or libraries who have to pay the bill.

Up until the second half of the 20th century, the majority of western scholarly communication was published by scientific/professional societies, university presses and educational publishers. These institutions sold subscriptions to the journals they published, but profit was traditionally of low priority, and scholarly publishing remained mostly the internal affair of the scientific community. In the second half of the 20th century, however, a number of developments led to the dramatic increase of commercial influence in the subscription based scholarly publishing market, transforming a rather unprofitable,
decentralized marketplace into a highly profitable, highly concentrated global enterprise, mostly run by commercial publishers. These changes include, among others,

- **the post-WWII boom in Western higher education and research**, which increased the pool of authors, the number of articles, the size of audiences (Tenopir & King, 1997), and therefore the size of the scholarly publishing market.

- **The development of ISI and other Science Citation Indices**, which identified a relatively narrow set of “core journals” from each discipline as being the most relevant publications for that field. (Guedon, 2001) As a result of these seemingly objective citation and impact metrics, core journals become highly desired publication venues both by authors (whose academic career relies on publishing in high prestige journals) and by readers (who rely on the preselection and quality assurance functions of these venues). This two sided demand for core journals, in turn, greatly increases their value as potential business assets.

- **Digitization**, which radically reduced the costs associated with the traditional tasks of the publisher without actually putting a downward pressure on prices.

- **A series of mergers and acquisitions**, which created a handful of strong, vertically integrated oligopolies in scholarly publishing (Larivière, Haustein, & Mongeon, 2015).

- **The inability of libraries to effectively resist price increases**, and redefine their role in the era of digital publishing. Rather than being the information brokers that they were, digital licensing deals forced them into the role of powerless “knowledge pumps” who were tasked to enforce access and licensing terms they did not necessarily agree with in the first place.

- **The general passivity of scientists**, who were not directly bearing the costs, but were directly enjoying the benefits of this publication system. The scientific journal was as much a tool to claim primacy and ownership over new ideas, as a vehicle to disseminate knowledge (Guedon, 2001). Most scientists were comfortable with working with and for the western publishers, and tolerate the publishers’ ever expanding ownership claims over the dissemination vehicles, as long as they could effectively use these vehicles to establish their ownership claims over the ideas.

These developments enabled commercial publishers not only to capture a significant chunk of the scholarly communication market, but to set the prices for their journal subscriptions based on their value, rather than what the articles actually cost for them (O’Donoghue, 2016). Consequently this market was increasingly characterized by a growing commercial influence, fast rising subscription costs and a general lack of access beyond the boundaries of rich western campuses.

The relentlessly rising subscription costs, and the promises of the digital revolution prompted the first serious discussions about the alternatives to and within the prevailing academic publishing practices in the early 1990’s (Okerson & O’Donnell, 1995). These discussions’ primary focus was the problem of rapid costs inflation. The discussions finally settled on three major fields of action, thought to be best suited to address this issue: Open Archiving set out to encourage individual and institutional self-archiving practices; Open Access explored publishing models in which the cost of publishing are not covered through access restrictions; while libraries started to set up professional organizations to improve their negotiating position vis-à-vis publishers.
Open Archives, Open Access, the loyal opposition to the status quo

In 1994, Steven Harnad, a Canadian cognitive scientist at the University of Southampton sent a “Subversive proposal” to an online mailing list, in which suggested to set up open ftp servers to store and disseminate scholarly articles in pre-print versions (Okerson & O'Donnell, 1995). At that time only a few of such pre-print archives were in operation. One of the very first ones, arxiv.org started only few years earlier, in 1991. The online discussion around Harnad’s proposal slowly matured into a wider, institutional repository based, self-archiving movement, which, in 2002 was strong enough to give birth to the Budapest Open Access Initiative (BOAI). The BOAI, and subsequent similar declarations² set out to provide the standards, the (meta data sharing) tools, and facilitated the self-archiving efforts of academic institutions by pooling knowledge, best practices, and expertise.

The same initiatives also paved the way towards Open Access journals. These journals, unlike Open Archives of various kinds, offer the very same functions (such as peer review, editing and typesetting) as traditional, subscription based journals, but do not charge their readers for access. Instead, they hope to fund their services by charging article processing fees from authors, or rely on institutional funding to cover their costs. (Suber & Darnton, 2016; Suber, 2013)

Libraries and their professional organizations formed the third major locus of activities. Their involvement in open (self) archiving and the operation of institutional document repositories was a natural fit with their historical responsibilities. Some of them also ventured into publishing: a number of research libraries launched the SPARC initiative in 1998 to establish low-cost competitors to journals published by commercial entities (Lustria & Case, 2005; Savenije, 2004). Libraries were also the ones most directly affected by the inflation of subscription costs. Like Cornell in 2003 (Faculty Library Advisory Board, 2003), or Harvard almost a decade later (Faculty Advisory Council, 2012), at one point or another every academic and research library had to address the financial sustainability of its journal subscriptions. One solution to that issue was, of course, a strong commitment to OA principles, and the encouragement of faculty to publish in OA journals. Another, more immediate solution was the formation of various consortia and professional associations, which formed buyers’ clubs to achieve stronger negotiating positions and better licensing terms from publishers.

Over the years the effectiveness of these strategies started to become more apparent. The rise of the subscription costs went on uninterrupted, but library consortia were seemingly getting better value for their money from publishers, who started to package their journals into “big deals”. For many libraries these deals also locked in their budgets, and reduced their flexibility in purchase decisions. On the other hand, the Open Access movement proved to be a substantial success both in generating organic growth³, and in convincing the two biggest research funders, the USA and the EU to finally mandate the

² Such as the 2003 Bethesda Statement on Open Access Publishing (http://legacy.earlham.edu/~peters/fos/bethesda.htm), and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (https://openaccess.mpg.de/Berlin-Declaration) in the same year.

³ The Directory of Open Access Books reported more than 150 publishers having published more than 4700 peer-reviewed open access books in mid-2016. The Directory of Open Access Journal reported more than 11 thousand open access journals in the first quarter of 2016. The Bielefeld Academic Search Engine is the second largest academic search engine after Google Scholar. It indexes more than 90 million documents. BASE estimates that about 60% of the articles are OA (Morrison, 2016). For a concise summery of the growth of OA see (Moody, 2016).
open accessibility of all publicly funded research (Holdren, 2013; The Dutch Ministry of Education Culture and Science, 2016).

Despite the successes, none of these actions revolutionized scholarly publishing. Peter Suber’s list of what Open Access is, and is not (Suber, 2013, pp. 20–27) makes it very clear, that the radical transformation of the status quo was not amongst the goal of OA. OA was not designed to “reform, violate, or abolish copyright”, it was not “an attempt to punish or undermine conventional publishers”, and finally, by stating that OA “isn’t about universal access”, it was made clear, that OA as a movement wished to refrain from the wider struggles around digital freedoms, online censorship, etc. ¹ Not facing any major and imminent financial, or social pressures, OA decided to play the role of the “loyal opposition” of commercial publishing, opting for a slow, gradual, and friendly change. Its course of action was to develop alternatives that, on the long run would complement, rather than replace the fundamentally commercial, paywall based systems of academic publishing.

These strategies seemed to work reasonably well within their narrowly defined western, English language contexts up until the 2000’s. They also proved to be incapable of addressing the challenges that came with the rapid transformation of science and higher education around the globe in the last few decades. With the rapid rise of middle classes in the developing world, this domain experienced a rapid and dramatic growth in multiple locations around the globe. At one point the internet united these local scientific communities in a single global domain, merging, on the way, the global periphery with the comfortably isolated Western institutional center.

Massification and globalization of science

It is interesting to note that the Western scholarly publishing crisis never really triggered a major piratical response. As I discuss in the next section, a plethora of informal practices of contestable legalities existed, but these informal practices rarely triggered serious legal conflicts⁵. To understand what led to the sudden, and rather dramatic explosion⁶ of scholarly piracy after 2010, one needs to look beyond the boundaries of western campuses, and look for he causes around the sudden expansion of higher education in the world.

The number of people enrolled in tertiary education grew almost sevenfold, from ~30 M in 1970 to nearly 200M in 2013 (Altbach, Reisberg, & Rumbley, 2009). This growth created an immense demand for articles and books, textbooks and other learning materials. In the West, the established infrastructures of scholarly communications were there to supply this demand. For the developing world, this growing demand was met with severely inadequate supply. Local production was lacking because local markets

¹ Note, that these limitations are significantly more restrictive than the what the wider creative commons movement has thought to be still appropriate: Lessig, the figurehead of the movement did try to challenge the fundamentals of US copyright (Eldred v. Ashcroft 537 U.S. 186, 2003), and he did argue against the intellectual monopolies of big media companies (Lessig, 2004).

⁵ This is not to ignore the ongoing legal struggles in the US around the fair use of copyrighted materials in educational settings, but the contested copying activities of universities and libraries were rarely willful infringements of clearly defined rules. See for example (Bartow, 1998) for details.

⁶ See later sections and (Bodó, 2015a; Cabanac, 2015) for more exact data on the growth of these services.
needed time to mature, while western publishers were also slow to provide appropriately priced texts. These structural problems created huge hurdles to the local accessibility of knowledge (Chon, 2010), and triggered a number of discourses to address the situation (Gaëlle Krikorian & Kapczynski, 2010). As the most apparent sign of the crisis, local, photocopy based, often piratical cottage industries stepped in to serve the local scholarly demand around campuses (Bollag, 2004; Chon, 2010; GARWE, 2014; Kaser, 1969; Lin-Liu, 2004; Lloyd, 2004; Overland, 2004; Rens, Prabhala, & Kawooya, 2006, pp. 29–30).

For a long time both the access problems, and their ad-hoc solutions remained local. But in the 1990’s this has started to change. Internet penetration grew. The forces of economic and cultural globalization, increased physical mobility, and various digital services opened up many of the local, non-English speaking education and research communities, and exposed them to the already partly globalized marketplaces and networks of education, research, employment, organizations and ideas. Suddenly millions of scholars appeared and wanted to participate in the global scholarly discourse.

While it went without question that the scholarly canon of this globalized domain is that of the West, written in, or translated to English, published by the major commercial publishers, the mechanisms to provide access to that canon were missing. The sudden surge in demand could not find its legal supply. The English language scholarly canon was paywalled. OA was not yet ready, and was not a good enough substitute. The local print based cottage industries were rapidly becoming obsolete in terms of the quality and quantity of materials they were able to carry. The new generation of students and researchers who were born into these conditions had to realize that they face insurmountable competitive disadvantages if they try to stick to the rules.

Meanwhile, in the US, and other western countries, the research and education sector had to cope with the economic crises, and tightening education budgets. Institutional budget cuts coincided with rising individual education costs and ballooning student debt. These developments slowly upset the relatively stable status quo on the western scholarly publishing markets as well, adding a large number of western students and scholars to the rapidly growing number of shadow library users.

Informal text sharing practices in the scholarly community

As we have just seen, piracy as a problem only became manifest at the peripheries of the globalizing world of scholarly publishing. This, however, does not mean that there weren’t any access problems in the center. While western scholars could rely on a rich and dense access infrastructure, a plethora of practices existed and are still in use to deal with the occasional inaccessibility problems. What is the quickest, most convenient way to get an article not carried by the library? Should I write to the author, ask around in the friend network, or ask on Twitter? How to disseminate the readings for the students: should I produced a printed reader, or is it enough if I distribute pdfs online? Should it be a closed ftp server, or I just leave a pen-drive with all the scanned articles in the classroom, and pretend not to notice what students do with it? When compiling that reader, should I scan, or copy the original? Should I save and organize the files since they are there already? Should I share them with colleagues? Should there be closed departmental e-text repository, which stores all these illegally scanned documents, which we then keep in secret? Many of these questions should be familiar to those working in research and education.
A number of informal practices developed in response the questions above. Wikipedia, for example, has its own home-brew academic journal exchange to acquire source materials since 2007. Journal articles which were previously exchanged via emails, mailing lists, ftp servers and online fora are now circulated through ad-hoc favor-networks, such as dedicated reddit channels, or special Twitter hashtags (Cabanac, 2015). Article swapping started on Reddit in April, 2010. The first tweet with the #icanhazpdf hashtag used for requesting inaccessible articles dates back to January 2011. Services like MedicineGround or Gigapaper were used to both request articles, and share remote access credentials to library systems and university networks. Scholarly books are sometimes made accessible through cyberlockers. In other cases, specialized shadow libraries, such as monoskop, or aaaaarg.org developed to serve disciplinary niches (“Interview with Dusan Barok,” 2013).

These low-key, ad-hoc informal text sharing practices were able to address the majority of access issues, within the context of western academia. They weren’t perceived as a major threat for a number of reasons. Publishers could capture some value from the grey market of photocopies through pricing (Liebowitz, 1985), photocopying levies (in the EU), or through licensing deals (Bartow, 1998). Most of the online activity didn’t reach the volume, or the economic impact to trigger serious counter-action (Young, 2008). None of these practices were inherently political either. This explains why piracy appeared so late to the western scholarly publishing market: it was always there, but only as a low intensity activity, taking place in informal, favor-based networks. Piracy was elsewhere, at the peripheries, far from home.

This relatively comfortable situation was due to change soon. The scholarly back markets at the periphery were the first signs of the globalization of science and education, but as long as they were print-based and thus local, the problem remained isolated at the edges. But local print shops were soon to give way to the online competition. The transition of foreign scholarly black markets from print to digital set piracy free from its technological and geographical quarantines, and set off a chain of events that led to the birth of a global scholarly black market. As the demand for online works surged rapidly, some online piratical venues started to successfully consolidate the dispersed online resources into a few highly organized shadow libraries. These consolidated services were getting better and better in attracting audiences from diverse disciplinary and geographical backgrounds, in effect putting an end to the separation of piracy at the centrum and at the periphery, and bringing piracy closer to home.

These developments started in the early 2000’s, and by the end of that decade, they created a quickly emerging piracy crisis in the western heartlands of scholarly communications. This is how piracy, after being so late, finally appeared in this domain.

The radical opposition to the publishing status quo

By the first half of 2010’s all the ingredients of a highly radicalized and political scholarly pirate movement were in place. It had a suitable name: Guerilla Open Access, inherited from Aaron Swartz’s

---

8 Chomikuj.pl, for example, is a polish site which serves as the shared shadow library of the polish academic community.
2008 Manifesto. It had its iconic leaders, known by name, like Aaron Swartz or by their online pseudonyms, like bw, smiley or ringo-ring. It had its major shadow libraries that already amassed a substantial part of the western scholarly canon. Pirates’ first conflicts with the publishers\(^9\) also offered an insight into how high the stakes were (Kelty, 2012). In the public, the cause of open access seemed to enjoy the sympathy of western, privileged academic elites. In the underground, this sympathy was already transformed into a practical alliance with scholars at various geographic, institutional and scholarly peripheries. In the last section of this paper I reconstruct how the various elements amalgamated into a coherent movement.

The alliance of the center and the periphery

Of shadow librarians, there are many. In fact “with books ready to be shared, meticulously cataloged, everyone is a librarian.” (Mars, 2012) However, there are only a handful, whose name is known, and out of those few, only two became visible beyond their immediate communities.

Aaron Swartz and Alexandra Elbakyan. The names of these two individuals become inseparably linked to the Guerilla Open Access phenomenon, which provides piratical access to scholarly publications.

Aaron Swartz was an American computer prodigy, political activist and hacktivist. In 2008 he published a document called the “Guerilla Open Access Manifesto”, in which he wrote: “There is no justice in following unjust laws. It’s time to come into the light and, in the grand tradition of civil disobedience, declare our opposition to this private theft of public culture. We need to take information, wherever it is stored, make our copies and share them with the world. We need to take stuff that’s out of copyright and add it to the archive. We need to buy secret databases and put them on the web. We need to download scientific journals and upload them to file sharing networks. We need to fight for Guerilla Open Access.” (Swartz, 2008)

In January 2011 Swartz was caught downloading millions of journal articles from JSTOR via the MIT network. Though neither JSTOR nor MIT pursued a civil lawsuit against him, US federal prosecutors charged him with a number of felonies, and Swartz was threatened with up to 50 years of imprisonment and $1 million in fines. In January, 2013 he committed suicide.

Alexandra Elbakyan is a Kazakh grad student who, in 2011 decided to develop a Sci-Hub, an online service, which could provide systematic, automatized access to paywalled articles. Before Sci-Hub, various academic discussion fora, organized according to disciplinary boundaries were used to post and fulfill requests for articles. Elbakyan developed Sci-Hub to replace that tedious, ad-hoc, manual process with an automatized system (Tolkacheva, 2015). This system uses the login credentials of western scientists to download paywalled articles via proxies usually provided by libraries to allow remote access to their electronic resources. These credentials are sometimes freely shared by their owners, sometimes they were bought, either from their owners or on various online credential exchanges. At the time of writing, in mid-2016 Elbakyan is fighting a court case, and giving online interviews from an undisclosed location.

\(^9\) Gigapedia/library.nu was shut down in 2012 as a result of a court action. (Liang, 2012)
On the surface, these two individuals could not be any more different. One is a widely known hacktivist, associated with not one but two of the world's highest ranking universities, with highly influential friends and admirers in the US and global academic, political and technological establishment. The other is a virtually unknown grad student from a remote, post-Soviet republic, at the very edge of the academic, political, institutional and technological periphery. Yet, their actions, their fate is intrinsically linked, as they constitute two sides of the very same coin.

First and foremost they embody the two extremes of the structural inequality encoded in the global knowledge economies (Chon, 2010; G. Krikorian & Kapczynski, 2010). They also embody the roles which needed to cooperate to address, from the bottom-up those inequalities. Swartz and people like him are the insiders, who have access, and as such, are instrumental in smuggling the knowledge out from behind the paywalls. Without their privileges, and their willingness to share those privileges, Elbakyan’s side of the operation would not be possible. Elbakyan and people like her are the outsiders. They represent those at the wrong side of the access paywalls. They looked at the work of previous generations, and learned that it is possible to set up highly efficient clandestine knowledge distribution networks to circumvent access obstacles, evade enforcement, and build influential underground knowledge repositories under hostile circumstances (Bodó, 2015b). They know that without their side of the operation, the smuggling of knowledge would make little sense.

The Guerilla Open Access Manifesto is a powerful document of global solidarity. It is written to rally the privileged to look beyond their own, rational self-interest, and the borders of western campuses: “Information is power. But like all power, there are those who want to keep it for themselves. The world’s entire scientific and cultural heritage, published over centuries in books and journals, is increasingly being digitized and locked up by a handful of private corporations. [...] That is too high a price to pay. Forcing academics to pay money to read the work of their colleagues? Scanning entire libraries but only allowing the folks at Google to read them? Providing scientific articles to those at elite universities in the First World, but not to children in the Global South? It’s outrageous and unacceptable. Those with access to these resources—students, librarians, scientists—you have been given a privilege. You get to feed at this banquet of knowledge while the rest of the world is locked out. But you need not—indeed, morally, you cannot—keep this privilege for yourselves. You have a duty to share it with the world. And you have: trading passwords with colleagues, filling download requests for friends. [...] But all of this action goes on in the dark, hidden underground. It’s called stealing or piracy, as if sharing a wealth of knowledge were the moral equivalent of plundering a ship and murdering its crew. But sharing isn’t immoral—it’s a moral imperative. Only those blinded by greed would refuse to let a friend make a copy.”

This message to the privileged came at the exact time when piratical practices at the peripheries also reached a tipping point. Library.nu/gigapedia, the first major scholarly shadow library started in early 2005 with a few hundred e-books10. By the time Swartz published his manifesto in mid-2008, it already amassed more than 110 thousand books, and at the time of its shutdown in 2012, there were nearly half million documents in its catalog11. Similar initiatives started up around the same time in Russia, consolidating first the Russian scholarly canon, then the western free floating piratical text collections

---

10 Source: archive.org
11 Own data on file with the author.
(Bodó, 2015a). In mid-2015, Library Genesis had more than 1.3 million documents in its catalog, including the catalog it inherited from the now defunct Gigapedia/library.nu library.

The Guerilla Open Access Manifesto was published roughly at the time when the dispersed, fragmented, parallel scholarly pirate services started to consolidate at the peripheries. The key to this consolidation was the aggregation and centralization of resources: the files, the credentials, the audiences, and the volunteers, who were willing to contribute with donations, exclusive resources, personal archives, time and energy.

This consolidation was necessary to provide access to a global audience of millions, but it came at a cost. As services grew, they become more visible, and they soon grew too big to be ignored by publishers. Soon enough, all the major shadow libraries were fighting lawsuits: Library.nu/gigapedia was sued first, in 2012, LibGen, Sci-Hub and AAAARG.ORG in 2015. These lawsuits led to substantial changes in both the hidden and in the public face of scholarly piracy: the underground needed to professionalize, the public started to be more political.

The professionalization of the pirate services

Most of the shadow libraries started out and developed as semi-amateur operations run by a number of volunteers, in their free time. Lawsuits are an existential threat to these services. Some shadow libraries, who did not have the will or the resources to respond to the new circumstances just pulled the plug, even if they weren’t targeted directly. Those who decided to fight, had to prepare for the hard times ahead, and plan around the pending seizure of the domain names and servers, problems in soliciting donations, etc. The loose network of volunteers, who operated these services also had to change, and think about stronger centralization, tighter organization, better operations security, increased redundancy, increased obscurity. They started donation campaigns to fill up cash reserves and cover the costs of technical and legal defense; they set up new domain names, and new servers; and they started to make redundant copies of their services on Tor and other enhanced privacy networks. As the discussions of one of the biggest shadow libraries illustrate, the increased exposure and the higher threat levels also led to the rough codification of informal, unwritten community norms; the establishment of basic organizational units; the separation of tasks and responsibilities; and in general, a more formalized and structured approach to operating the site and taking decisions over its course.\(^\text{12}\)

The costs and benefits of public exposure

Besides the legal troubles, and the engineering challenges, shadow libraries had to deal with the uninvited attention and public scrutiny. Neither Swartz, nor Elbakyan wished or planned to become the poster child of scholarly piracy. Their different responses to the situation highlight the perils of being identified as a pirate.

\(^{12}\) On file with the author.
Swartz, the author of GOA Manifesto was a powerful advocate of all kinds of openness in all of his life. When he was caught downloading millions of articles, US prosecutors assumed that he wanted to release them to the public, just as he suggested in his writing. However, Swartz, probably at the advice of his lawyers remained silent during the investigation. To minimize his legal liabilities and exposure, he said nothing about what his aims were with the articles, and decided not to use the court case to advocate his ideas on (guerilla) open access.

The pause in the debate on Guerilla Open Access ended with his suicide. The “official” reactions were rather mute. The intense soul searching among the friends of Swartz in the US academic and political establishment ended in a moral impasse, where the establishment could neither condone Swartz’s acts, nor were in the position to squarely put the blame on him (Abelson, Diamond, Grosso, & Pfeiffer, 2013; Lessig, 2013; Samuelson, 2013). Aaron’s Law, an amendment to the Computer Fraud and Abuse Act to prevent prosecutorial overreach similar to what happened in Swartz’s case was introduced in the US Senate in April, 2015, with no subsequent action as of mid-2016. Though the US took huge steps to remove access obstacles for publicly funded research (Holdren, 2013; “Notes on the Fair Access to Science and Technology Research Act,” 2015), none of these initiative reference Swartz, or his legacy.

On the other hand, the community response to his suicide was highly transformative. Honoring Swartz’s hacktivist legacy, the first memorial hackathon was organized a few days after his death, of all the places, in Tehran, Iran. Soon, dozens of others followed, organized by hacker spaces, universities, civil society organizations around the world from Bangalore, India to Stanford University California. By 2016, both his legacy is remembered every year with various grassroots-organized events that already resulted in the development of numerous software tools in the domains of privacy, learning, data sharing, transparency, as well as activist projects in the field of law and politics.

Elbakyan’s approach to the court case and to public exposure is the complete opposite to Swartz’s. After being identified in court documents as the person responsible for Sci-Hub, she immediately went on the offensive and started to explain her actions, motives, in court documents (Elbakyan, 2015), blog posts (Elbakyan, n.d.), interviews (Belluz, 2016; Bonik & Schale, 2015; Ernesto, 2015), online discussions (Cochran, 2016). She even released logs that enabled the analysis of the usage of Sci-Hub (Bohannon, 2016). She does not seems to care about the impact of her public activities on her legal case. Her words in effect amount to a guilty plea, reducing her chances to win the legal case to virtually nil. On the other hand, her engagement managed to open up and merge the different debates on the perilous position of research libraries, on the individual responsibilities of scientists, on commercial publishers, on the limits of government intervention, on the role of copyright in science, on global knowledge inequalities, etc. Her presence in the debates made it impossible to pretend that shadow libraries, and the problems they address are not there, or they are marginal, or that there is a straightforward solution to them. It forced various stakeholders to come out forward, and clearly state where they stand in the debate, and what their solution would be.

13 https://www.noisebridge.net/wiki/Worldwide_Aaron_Swartz_Memorial_Hackathon_Series
14 http://www.aaronswartzday.org/
Guerilla Open access – where is it heading?

A year into this debate it is clear that the world of scholarly publishing and open access is at an impasse. Open Access will soon become a statutory obligation in the two biggest science producers in the world, but it is unclear what the next steps would be. The status quo is as entrenched as ever. The recent history of copyright enforcement suggests that despite all the lawsuits, shadow libraries are also here to stay.

This situation is obviously not sustainable, but it is also not clear how it could be resolved. Shadow libraries will continue to play a central role in the global domain of scholarly communications as long as the inequalities in terms of center and periphery, legacy works and new production, access as reader and access as writer remain. Shadow libraries provide equally permissive terms of access to any content, regardless of its copyright status, to anyone, regardless of individual or institutional wealth, privilege, or status in the geographical and power hierarchies. As long as this equality is not offered, either voluntarily by authors and publishers, or through statutory means by governments, shadow libraries will play a role in the domain of scholarly communication.

How central and transformative that role will be depends on three additional factors: the long term alliance of the different agents at the center and at the periphery; the technical impunity of the services; and the legal impunity of open access guerillas.

The Open Access Guerilla Cookbook (Williwaw, 2013) lists all the different skills, resources and expertise a shadow library needs for its existence. Most these roles require individuals to be engaged in copyright piracy, which is rarely seen as a legitimate form of activity, especially in the US, where institutional norms and the legal culture simply does not allow copyright infringement to be interpreted as a condonable course of action. The European cultural environment seems a little bit more permissive, especially after pirate parties and popular movements were able to transform discussions on copyright infringement into a wider digital/cultural environmentalist political/social movements (Burkart, 2014). But still, if Open Access Guerillas want to ensure the support and solidarity of the enforcement-weary, law abiding mainstream, they may need to steer the discussion away from copyright infringement, towards other issues, such as the commercial appropriability of scholarly knowledge.

The current technology landscape, especially the widespread use of privacy enhancing technologies (PETs) may also be useful to foster solidarity. PETs enable individual scientists to maintain multiple identities, some of them respectable and legitimate, others less so (Bodó, 2014a). Most scholars probably already follow an opportunistic mix of scholarly communication strategies, in which seemingly contradictory practices, such as getting published in a first tier journal, even if it is paywalled; using pre-print services, such as SSRN to distribute manuscripts; showing support for OA; sharing scanned

---

15 The Advocate (who promotes the cause of open access), The Prospector (who seeks out closed databases worth opening up), the Scribe (who is willing to share his/her own private collection of documents), the Courier (who moves and distributes collections across networks, domains, technologies), the Innkeeper (who provides the safe and secure places of communication and coordination); the Armorer (who write and update the scrapers to liberate content), the Sapper (who explores ways to infiltrate the security of archives and databases and enable outsiders direct access), the Traitor (who has legitimate and legal access to content and is willing to share that with others), and Custodian (who do the canonical preservation).
literature with students; occasionally helping out a colleague on the #IcanHazPDF channel; and using LibGen and Sci-Hub with or without bad conscience can peacefully coexist. PETs allow the anonymous engagement in any of these practices with a few simple clicks, significantly lowering the threshold of participation.

The second factor is the technical impunity of shadow libraries. Again, PETs will prove to be key in ensuring that shadow libraries remain accessible in face of domain name and server seizures, filtering and other technical enforcement measures (Bodó, 2014b). At this point it is already clear that the technical impunity of piracy depends on the outcome in the digital privacy / encryption debates. If in those debates the idea of universally strong encryption prevails, PETs will always provide a technological safe haven for piracy.

The last factor is the legal impunity of shadow librarians. A lost court case can have devastating consequences for the individual found guilty of copyright infringement, and often to the services they created. Swartz was facing excessive punishment in the hands of US prosecutors and he escaped to suicide. The admins of Gigapedia/library.nu got sued and shut down the service (Kelty, 2012). The administrators of the Pirate Bay spent several months in prison, and while the site is still up, it is slowly sinking into irrelevance. Elbakyan seems to believe that she is beyond the reach of US jurisdiction. It remains to be seen whether that is truly the case, but even if she is able to evade the legal consequences of her actions, this comes at the devastating cost of not being able to travel to the scholarly centers any more.

There is a good chance that shadow libraries will prove to be irrepressible. Will commercial publishers voluntarily develop services to outcompete pirates, the way music and audiovisual piracy forced the entertainment business to offer low cost all-you can eat subscription services? It might be, but the scholarly communication market is substantially different from the entertainment market. Piratical scholars are not the customers of commercial publishers. It is the libraries who pay the subscriptions and who have to negotiate between the local scholarly communities they serve, the pirates, who use their services without authorization, and commercial publishers trying to stop the content leakages (Ruff, 2016). Currently “[p]art of being an academic librarian today involves providing uncompensated copyright enforcement for publishing interests, in order to reinforce values you do not even believe in” (Banks, 2016). Under such conditions it is unclear what the right compromise, acceptable for all involved, would really be.

But change does not need to happen organically. The OA movement has already proved that governments can be persuaded to step in, and overcome the passivity of scholars, and the pushback from publishers. After the success of the government mandated Open Access, the liberation of past research may be the next logical step. Because as the history of book piracy suggests, the best antidote to piracy is to create conditions in which they are redundant and obsolete (Bodó, 2011).
References


Moody, G. (2016). Open access: All human knowledge is there—so why can’t everybody access it? Retrieved from http://arstechnica.co.uk/science/2016/06/what-is-open-access-free-sharing-of-all-human-knowledge/


