Libraries in the post-scarcity era

Bodó, B.

Link to publication

Libraries in the post-scarcity era

Balázs Bodó

Abstract

In the digital era where, thanks to the ubiquity of electronic copies, the book is no longer a scarce resource, libraries find themselves in an extremely competitive environment. Several different actors are now in a position to provide low cost access to knowledge. One of these competitors are shadow libraries - piratical text collections which have now amassed electronic copies of millions of copyrighted works and provide access to them usually free of charge to anyone around the globe. While such shadow libraries are far from being universal, they are able to offer certain services better, to more people and under more favorable terms than most public or research libraries. This contribution offers insights into the development and the inner workings of one of the biggest scientific shadow libraries on the internet in order to understand what kind of library people create for themselves if they have the means and if they don't have to abide by the legal, bureaucratic and economic constraints that libraries usually face. I argue that one of the many possible futures of the library is hidden in the shadows, and those who think of the future of libraries can learn a lot from book pirates of the 21st century about how users and readers expect texts in electronic form to be stored, organized and circulated.

"The library is society's last non-commercial meeting place which the majority of the population uses." (Committee on the Public Libraries in the Knowledge Society, 2010)

"With books ready to be shared, meticulously cataloged, everyone is a librarian. When everyone is librarian, library is everywhere." – Marcell Mars, www.memoryoftheworld.org

I have spent the last few months in various libraries visiting - a library. I spent countless hours in the modest or grandiose buildings of the Harvard Libraries, the Boston and Cambridge Public Library systems, various branches of the Openbare Bibliotheek in Amsterdam, the libraries of the University of Amsterdam, with a computer in front of me, on which another library was running, a library which is perfectly virtual, which has no monumental buildings, no multi-million euro budget, no miles of stacks, no hundreds of staff, but which has, despite lacking all what apparently makes a library, millions of literary works and millions of scientific books, all digitized, all available at the click of the mouse for everyone on the earth without any charge, library or university membership. As I was sitting in these

physical spaces where the past seemed to define the present, I was wondering where I should look to find the library of the future: down to my screen or up around me.

The library on my screen was Aleph, one of the biggest of the countless piratical text collections on the internet. It has more than a million scientific works and another million literary works to offer, all free to download, without any charge or fee, for anyone on the net. I've spent months among its virtual stacks, combing through the catalogue, talking to the librarians who maintain the collection, and watching the library patrons as they used the collection. I kept going back to Aleph both as a user and as a researcher. As a user, Aleph offered me books that the local libraries around me didn't, in formats that were more convenient than print. As a researcher, I was interested in the origins of Aleph, its modus operandi, its future, and I was curious where the journey to which it has taken the book-readers, authors, publishers and libraries would end.

In this short essay I will introduce some of the findings of a two year research project conducted on Aleph. In the project I looked at several things. I reconstructed the pirate library's genesis in order to understand the forces that called it to life and shaped its development. I looked at its catalogue to understand what it has to offer and how that piratical supply of books is related to the legal supply of books through libraries and online distributors. I also acquired data on its usage, so was able to reconstruct some aspects of piratical demand. After a short introduction, in the first part of this essay I will outline some of the main findings, and in the second part will situate the findings in the wider context of the future of libraries.

Book pirates and shadow librarians

Book piracy has a fascinating history, tightly woven into the history of the printing press (Judge, 1934), into the history of censorship (Wittmann, 2004), into the history of copyright (Bently, Davis, & Ginsburg, 2010; Bodó, 2011a) and into the history of European civilization (Johns, 2010). Book piracy, in the 21st or in the mid-17th century is an activity that has deep cultural significance, because ultimately it is a story about how knowledge is circulated beyond and often against the structures of political and economic power (Bodó, 2011b), and thus it is a story about the changes this unofficial circulation of knowledge brings.

There are many different types of book pirates. Some just aim for easy money, others pursue highly ideological goals, but they are invariably powerful harbingers of change. The emergence of black markets whether they be of culture, of drugs or of arms is always a symptom, a warning sign of a friction between

supply and demand. Increased activity in the grey and black zones of legality marks the emergence of a demand which legal suppliers are unwilling or unable to serve (Bodó, 2011a). That friction, more often than not, leads to change. Earlier waves of book piracy foretold fundamental economic, political, societal or technological shifts (Bodó, 2011b): changes in how the book publishing trade was organized (Judge, 1934; Pollard, 1916, 1920); the emergence of the new, bourgeois reading class (Patterson, 1968; Solly, 1885); the decline of pre-publication censorship (Rose, 1993); the advent of the Reformation and of the Enlightenment (Darnton, 1982, 2003), or the rapid modernization of more than one nation (Khan & Sokoloff, 2001; Khan, 2004; Yu, 2000).

The latest wave of piracy has coincided with the digital revolution which, in itself, profoundly upset the economics of cultural production and distribution (Landes & Posner, 2003). However technology is not the primary cause of the emergence of cultural black markets like Aleph. The proliferation of computers and the internet has just revealed a more fundamental issue which all has to do with the uneven distribution of the access to knowledge around the globe.

Sometimes book pirates do more than just forecast and react to changes that are independent of them. Under certain conditions, they themselves can be powerful agents of change (Bodó, 2011b). Their agency rests on their ability to challenge the status quo and resist cooptation or subjugation. In that effect, digital pirates seem to be quite resilient (Giblin, 2011; Patry, 2009). They have the technological upper hand and so far they have been able to outsmart any copyright enforcement effort (Bodó, forthcoming). As long as it is not completely possible to eradicate file sharing technologies, and as long as there is a substantial difference between what is legally available and what is in demand, cultural black markets will be here to compete with and outcompete the established and recognized cultural intermediaries. Under this constant existential threat, business models and institutions are forced to adapt, evolve or die.

After the music and audiovisual industries, now the book industry has to address the issue of piracy. Piratical book distribution services are now in direct competition with the bookstore on the corner, the used book stall on the sidewalk, they compete with the Amazons of the world and, like it or not, they compete with libraries. There is, however, a significant difference between the book and the music industries. The reluctance of music rights holders to listen to the demands of their customers caused little damage beyond the markets of recorded music. Music rights holders controlled their own fates and those who wanted to experiment with alternative forms of distribution had the chance to do so. But while the rapid proliferation of book black markets may signal that the book industry suffers from similar problems as the music industry suffered a decade ago, the actions of book publishers, the policies they pursue have impact beyond the market of books and directly affect the domain of libraries.

The fate of libraries is tied to the fate of book markets in more than one way. One connection is structural: libraries emerged to remedy the scarcity in books. This is true both for the pre-print era as well as in the Gutenberg galaxy. In the era of widespread literacy and highly developed book markets, libraries offer access to books under terms publishers and booksellers cannot or would not. Libraries, to a large extent, are defined to complement the structure of the book trade. The other connection is legal. The core activities of the library (namely lending, copying) are governed by the same copyright laws that govern authors and publishers. Libraries are one of the users in the copyright system, and their existence depends on the limitations of and exceptions to the exclusive rights of the rights holders. The space that has been carved out of copyright to enable the existence of libraries has been intensely contested in the era of postmodern copyright (Samuelson, 2002) and digital technologies. This heavy legal and structural interdependence with the market means that libraries have only a limited control over their own fate in the digital domain.

Book pirates compete with some of the core services of libraries. And as is usually the case with innovation that has no economic or legal constraints, pirate libraries offer, at least for the moment, significantly better services than most of the libraries. Pirate libraries offer far more electronic books, with much less restrictions and constraints, to far more people, far cheaper than anyone else in the library domain. Libraries are thus directly affected by pirate libraries, and because of their structural interdependence with book markets, they also have to adjust to how the commercial intermediaries react to book piracy. Under such conditions libraries cannot simply count on their survival through their legacy. Book piracy must be taken seriously, not just as a threat, but also as an opportunity to learn how shadow libraries operate and interact with their users. Pirate libraries are the products of readers (and sometimes authors), academics and laypeople, all sharing a deep passion for the book, operating in a zone where there is little to no obstacle to the development of the "ideal" library. As such, pirate libraries can teach important lessons on what is expected of a library, how book consumption habits evolve, and how knowledge flows around the globe.

Pirate libraries in the digital age

The collection of texts in digital formats was one of the first activities that computers enabled: the text file is the native medium of the computer, it is small, thus it is easy to store and copy. It is also very easy to create, and as so many projects have since proved, there are more than enough volunteers who are willing to type whole books into the machine. No wonder that electronic libraries and digital text repositories were among the first "mainstream" application of computers. Combing through large stacks of matrix-

printer printouts of sci-fi classics downloaded from gopher servers is a shared experience of anyone who had access to computers and the internet before it was known as the World Wide Web.

Computers thus added fresh momentum to the efforts of realizing the age-old dream of the universal library (Battles, 2004). Digital technologies offered a breakthrough in many of the issues that previously posed serious obstacles to text collection: storage, search, preservation, access have all become cheaper and easier than ever before. On the other hand, a number of key issues remained unresolved: digitization was a slow and cumbersome process, while the screen proved to be too inconvenient, and the printer too costly an interface between the text file and the reader. In any case, ultimately it wasn't these issues that put a break to the proliferation of digital libraries. Rather, it was the realization, that there are legal limits to the digitization, storage, distribution of copyrighted works on the digital networks. That realization soon rendered many text collections in the emerging digital library scene inaccessible.

Legal considerations did not destroy this chaotic, emergent digital librarianship and the collections the ad-hoc, accidental and professional librarians put together. The text collections were far too valuable to simply delete them from the servers. Instead, what happened to most of these collections was that they retreated from the public view, back into the access-controlled shadows of darknets. Yesterday's gophers and anonymous ftp servers turned into closed, membership only ftp servers, local shared libraries residing on the intranets of various academic, business institutions and private archives stored on local hard drives. The early digital libraries turned into book piracy sites and into the kernels of today's shadow libraries.

Libraries and other major actors, who decided to start large scale digitization programs soon needed to find out that if they wanted to avoid costly lawsuits, then they had to limit their activities to work in the public domain. While the public domain is riddled with mind-bogglingly complex and unresolved legal issues, but at least it is still significantly less complicated to deal with than copyrighted and orphan works. Legally more innovative, (or as some would say, adventurous) companies, such as Google and Microsoft, who thought they had sufficient resources to sort out the legal issues soon had to abandon their programs or put them on hold until the legal issues were sorted out.

There were, however, a large group of disenfranchised readers, library patrons, authors and users who decided to ignore the legal problems and set out to build the best library that could possibly be built using the digital technologies. Despite the increased awareness of rights holders to the issue of digital book piracy, more and more communities around text collections started defy the legal constraints and to operate and use more or less public piratical shadow libraries.

Aleph[1]

Aleph[2] is a meta-library, and currently one of the biggest online piratical text collections on the internet. The project started on a Russian bulletin board devoted to piracy in around 2008 as an effort to integrate various free-floating text collections that circulated online, on optical media, on various public and private ftp servers and on hard-drives. Its aim was to consolidate these separate text collections, many of which were created in various Russian academic institutions, into a single, unified catalog, standardize the technical aspects, add and correct missing or incorrect metadata, and offer the resulting catalogue, computer code and the collection of files as an open infrastructure.

From Russia with love

It is by no means a mistake that Aleph was born in Russia. In post-Soviet Russia the unique constellation of several different factors created the necessary conditions for the digital librarianship movement that ultimately led to the development of Aleph. A rich literary legacy, the Soviet heritage, the pace with which various copying technologies penetrated the market, the shortcomings of the legal environment and the informal norms that stood in for the non-existent digital copyrights all contributed to the emergence of the biggest piratical library in the history of mankind.

Russia cherishes a rich literary tradition, which suffered and endured extreme economic hardships and political censorship during the Soviet period (Ermolaev, 1997; Friedberg, Watanabe, & Nakamoto, 1984; Stelmakh, 2001). The political transformation in the early 1990's liberated authors, publishers, librarians and readers from much of the political oppression, but it did not solve the economic issues that stood in the way of a healthy literary market. Disposable income was low, state subsidies were limited, the dire economic situation created uncertainty in the book market. The previous decades, however, have taught authors and readers how to overcome political and economic obstacles to access to books. During the Soviet times authors, editors and readers operated clandestine samizdat distribution networks, while informal book black markets, operating in semi-private spheres, made uncensored but hard to come by books accessible (Stelmakh, 2001). This survivalist attitude and the skills that came with it became handy in the post-Soviet turmoil, and were directly transferable to the then emerging digital technologies.

---

[1] I have conducted extensive research on the origins of Aleph, on its catalogue and its users. The detailed findings, at the time of writing this contribution are being prepared for publication. The following section is brief summary of those findings and is based upon two forthcoming book chapters on Aleph in a report, edited by Joe Karaganis, on the role of shadow libraries in the higher education systems of multiple countries.

[2] Aleph is a pseudonym chosen to protect the identity of the shadow library in question.

Russia is not the only country with a significant informal media economy of books, but in most other places it was the photocopy machine that emerged to serve such book grey/black markets. In pre-1990 Russia and in other Eastern European countries the access to this technology was limited, and when photocopiers finally became available, computers were close behind them in terms of accessibility. The result of the parallel introduction of the photocopier and the computer was that the photocopy technology did not have time to lock in the informal market of texts. In many countries where the photocopy machine preceded the computer by decades, copy shops still capture the bulk of the informal production and distribution of textbooks and other learning material. In the Soviet-bloc PCs instantly offered a less costly and more adaptive technology to copy and distribute texts.

Russian academic and research institutions were the first to have access to computers. They also had to somehow deal with the frustrating lack of access to up-to-date and affordable western works to be used in education and research (Abramitzky & Sin, 2014). This may explain why the first batch of shadow libraries started in a number of academic/research institutions such as the Department of Mechanics and Mathematics (MexMat) at Moscow State University. The first digital librarians in Russia were mathematicians, computer scientists and physicists, working in those institutions.

As PCs and internet access slowly penetrated Russian society, an extremely lively digital librarianship movement emerged, mostly fuelled by enthusiastic readers, book fans and often authors, who spared no effort to make their favorite books available on FIDOnet, a popular BBS system in Russia. One of the central figures in these tumultuous years, when typed-in books appeared online by the thousands, was Maxim Moshkov, a computer scientist, alumnus of the MexMat, and an avid collector of literary works. His digital library, lib.ru was at first mostly a private collection of literary texts, but soon evolved into the number one text repository which everyone used to depose the latest digital copy on a newly digitized book (Мошков, 1999). Eventually the library grew so big that it had to be broken up. Today it only hosts the Russian literary classics. User generated texts, fan fiction and amateur production was spin off into the aptly named samizdat.lib.ru collection, low brow popular fiction, astrology and cheap romance found its way into separate collections, and so did the collection of academic/scientific books, which started an independent life under the name of Kolkhoz. Kolkhoz, which borrowed its name from the commons based agricultural cooperative of the early Soviet era, was both a collection of scientific texts, and a community of amateur librarians, who curated, managed and expanded the collection.

Moshkov and his library introduced several important norms into the bottom-up, decentralized, often anarchic digital library movement that swept through the Russian internet in the late 1990's, early 2000's. First, lib.ru provided the technological blueprint for any future digital library. But more importantly,

Moshkov's way of handling the texts, his way of responding to the claims, requests, questions, complaints of authors and publishers paved the way to the development of copynorms (Schultz, 2007) that continue to define the Russian digital library scene until today. Moshkov was instrumental in the creation of an enabling environment for the digital librarianship while respecting the claims of authors, during times when the formal copyright framework and the enforcement environment was both unable and unwilling to protect works of authorship (Elst, 2005; Sezneva, 2012).

Guerilla Open Access

Around the time of the late 2000's when Aleph started to merge the Kolkhoz collection with other, free-floating texts collections, two other notable events took place. It was in 2008 when Aaron Swartz penned his Guerilla Open Access Manifesto (Swartz, 2008), in which he called for the liberation and sharing of scientific knowledge. Swartz forcefully argued that scientific knowledge, the production of which is mostly funded by the public and by the voluntary labor of academics, cannot be locked up behind corporate paywalls set up by publishers. He framed the unauthorized copying and transfer of scientific works from closed access text repositories to public archives as a moral act, and by doing so, he created an ideological framework which was more radical and promised to be more effective than either the creative commons (Lessig, 2004) or the open access (Suber, 2013) movements that tried to address the access to knowledge issues in a more copyright friendly manner. During interviews, the administrators of Aleph used the very same arguments to justify the raison d'être of their piratical library. While it seems that Aleph is the practical realization of Swartz's ideas, it is hard to tell which served as an inspiration for the other.

It was also in around the same time when another piratical library, gigapedia/library.nu started its operation, focusing mostly on making freely available English language scientific works (Liang, 2012). Until its legal troubles and subsequent shutdown in 2012, gigapedia/library.nu was the biggest English language piratical scientific library on the internet amassing several hundred thousand books, including high-quality proofs ready to print and low resolution scans possibly prepared by a student or a lecturer. During 2012 the mostly Russian-language and natural sciences focused Alephs absorbed the English language, social sciences rich gigapedia/library.nu, and with the subsequent shutdown of gigapedia/library.nu Aleph became the center of the scientific shadow library ecosystem and community.

Aleph by numbers

By adding pre-existing text collections to its catalogue Aleph was able to grow at an astonishing rate. Aleph added, on average 17.500 books to its collection each month since 2009, and as a result, by April 2014 is has more than 1.15 million documents. Nearly two thirds of the collection is in English, one fifth of the documents is in Russian, while German works amount to the third largest group with 8.5% of the collection. The rest of the major European languages, like French or Spanish have less than 15000 works each in the collection.

More than 50 thousand publishers have works in the library, but most of the collection is published by mainstream western academic publishers. Springer published more than 12% of the works in the collection, followed by the Cambridge University Press, Wiley, Routledge and Oxford University Press, each having more than 9000 works in the collection.

Most of the collection is relatively recent, more than 70% of the collection being published in 1990 or after. Despite the recentness of the collection, the electronic availability of the titles in the collection is limited. While around 80% of the books that had an ISBN number registered in the catalogue[3] was available in print either as a new copy or a second hand one, only about one third of the titles were available in e-book formats. The mean price of the titles still in print was 62 USD according to the data gathered from Amazon.com.

The number of works accessed through of Aleph is as impressive as its catalogue. In the three months between March and June, 2012, on average 24.000 documents were downloaded every day from one of its half-a-dozen mirrors.[4] This means that the number of documents downloaded daily from Aleph is probably in the 50 to 100.000 range. The library users come from more than 150 different countries. The biggest users in terms of volume were the Russian Federation, Indonesia, USA, India, Iran, Egypt, China, Germany and the UK. Meanwhile, many of the highest per-capita users are Central and Eastern European countries.

What Aleph is and what it is not

Aleph is an example of the library in the post scarcity age. It is founded on the idea that books should no longer be a scarce resource. Aleph set out to remove both sources of scarcity: the natural source of

---

[3] Market availability data is only available for that 40% of books in the Aleph catalogue that had an ISBN number on file. The titles without a valid ISBN number tend to be older, Russian language titles, in general with low expected print and e-book availability.

[4] Download data is based on the logs provided by one of the shadow library services which offers the books in Aleph's catalogue as well as other works also free and without any restraints or limitations.

scarcity in physical copies is overcome through distributed digitization; the artificial source of scarcity created by copyright protection is overcome through infringement. The liberation from both constraints is necessary to create a truly scarcity free environment and to release the potential of the library in the post-scarcity age.

Aleph is also an ongoing demonstration of the fact that under the condition of non-scarcity, the library can be a decentralized, distributed, commons-based institution created and maintained through peer production (Benkler, 2006). The message of Aleph is clear: users left to their own devices, can produce a library by themselves for themselves. In fact, users are the library. And when everyone has the means to digitize, collect, catalogue and share his/her own library, then the library suddenly is everywhere. Small individual and institutional collections are aggregated into Aleph, which, in turn is constantly fragmented into smaller, local, individual collections as users download works from the collection. The library is breathing (Battles, 2004) books in and out, but for the first time, this circulation of books is not a zero sum game, but a cumulative one: with every cycle the collection grows.

On the other hand Aleph may have lots of books on offer, but it is clear that it is neither universal in its scope, nor does it fulfill all the critical functions of a library. Most importantly Aleph is disembedded from the local contexts and communities that usually define the focus of the library. While it relies on the availability of local digital collections for its growth, it has no means to play an active role in its own development. The guardians of Aleph can prevent books from entering the collection, but they cannot pay, ask or force anyone to provide a title if it is missing. Aleph is reliant on the weak copy-protection technologies of official e-text repositories and the goodwill of individual document submitters when it comes to the expansion of the collection. This means that the Aleph collection is both fragmented and biased, and it lacks the necessary safeguards to ensure that it stays either current or relevant.

Aleph, with all its strengths and weaknesses carries an important lesson for the discussions on the future of libraries. In the next section I'll try situate these lessons in the wider context of the library in the post scarcity age.

The future of the library

There is hardly a week without a blog post, a conference, a workshop or an academic paper discussing the future of libraries. While existing libraries are buzzing with activity, librarians are well aware that they need to re-define themselves and their institutions, as the book collections around which libraries were organized slowly go the way the catalogue has gone: into the digital realm. It would be impossible to give

a faithful summary of all the discussions on the future of libraries is such a short contribution. There are, however, a few threads, to which the story of Aleph may contribute.

Competition

It is very rare to find the two words: libraries and competition in the same sentence. No wonder: libraries enjoyed a near perfect monopoly in their field of activity. Though there may have been many different local initiatives that provided free access to books, as a specialized institution to do so, the library was unmatched and unchallenged. This monopoly position has been lost in a remarkably short period of time due to the internet and the rapid innovations in the legal e-book distribution markets. Textbooks can be rented, e-books can be lent, a number of new startups and major sellers offer flat rate access to huge collections. Expertise that helps navigate the domains of knowledge is abundant, there are multiple authoritative sources of information and meta-information online. The search box of the library catalog is only one, and not even the most usable of all the different search boxes one can type a query in[5]. Meanwhile there are plenty of physical spaces which offer good coffee, an AC plug, comfortable chairs and low levels of noise to meet, read and study from local cafes via hacker- and maker spaces, to co-working offices. Many library competitors have access to resources (human, financial, technological and legal) way beyond the possibilities of even the richest libraries. In addition, publishers control the copyrights in digital copies which, absent of well fortified statutory limitations and exceptions, prevent libraries keeping up with the changes in user habits and with the competing commercial services.

Libraries definitely feel the pressure. *"Libraries' offers of materials […] compete with many other offers that aim to attract the attention of the public. […] It is no longer enough just to make a good collection available to the public."* (Committee on the Public Libraries in the Knowledge Society, 2010) As a response, libraries have developed different strategies to cope with this challenge. The common thread in the various strategy documents is that they try to redefine the library as a node in the vast network of institutions that provide knowledge, enable learning, facilitate cooperation and initiate dialogues. Some of the strategic plans redefine the library space as an "*independent medium to be developed"* (Committee on the Public Libraries in the Knowledge Society, 2010), and advise libraries to transform themselves into culture and community centers which establish partnerships with citizens, communities and with other public and private institutions. Some librarians propose even more radical ways of keeping the library

---

[5] ArXiv, SSRN, RePEc, PubMed Central, Google Scholar, Google Books, Amazon, Mendeley, Citavi, ResearchGate, Goodreads, LibraryThing, Wikipedia, Yahoo Answers, Khan Academy, specialized twitter and other social media accounts are just a few of the available discovery services.

relevant by, for example, advocating more opening hours without staff and hosting more user-governed activities.

In the research library sphere, the Commission on the Future of the Library, a task force set up by the University of California Berkeley defined the values the university research library will add in the digital age as *"1) Human expertise; 2) Enabling infrastructure; and 3) Preservation and dissemination of knowledge for future generations."* (Commission on the Future of the Library, 2013). This approach is from among the more conservative ones, still relying on the hope that libraries can offer something unique that no one else is able to provide. Others, working at the Association of Research Libraries are more like their public library counterparts, defining the future role of the research libraries as a *"convener of 'conversations' for knowledge construction, an inspiring host; a boundless symposium; an incubator; a 3rd space both physically and virtually; a scaffold for independence of mind; and a sanctuary for freedom of expression, a global entrepreneurial engine"* (Pendleton-Jullian, Lougee, Wilkin, & Hilton, 2014), in other words, as another important, but in no way unique node in the wider network of institutions that creates and distributes knowledge.

Despite the differences in priorities, all these recommendations carry the same basic message. The unique position of libraries in the center of a book-based knowledge economy, on the top of the paper-bound knowledge hierarchy is about to be lost. As libraries are losing their monopoly of giving low cost, low restrictions access to books which are scarce by nature, and they are losing their privileged and powerful position as the guardians of and guides to the knowledge stored in the stacks. If they want to survive, they need to find their role and position in a network of institutions, where everyone else is engaged in activities that overlap with the historic functions of the library. Just like the books themselves, the power that came from the privileged access to books is in part dispersed among the countless nodes in the knowledge and learning networks, and in part is being captured by those who control the digital rights to digitize and distribute books in the digital era.

One of the main reasons why libraries are trying to redefine themselves as providers of ancillary services is because the lack of digital lending rights prevents them from competing on their own traditional home turf - in giving free access to knowledge. The traditional legal limitations and exceptions to copyright that enabled libraries to fulfill their role in the analogue world do not apply in the digital realm. In the European Union, the Infosoc Directive ("Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society," 2001) allows for libraries to create digital copies for preservation, indexing and similar purposes and allows for the display of digital copies on their premises for research and personal study (Triaille et al., 2013). While in theory these rights provide for

the core library services in the digital domain, their practical usefulness is rather limited, as off-premises e-lending of copyrighted works is in most cases[6] only possible through individual license agreements with publishers.

Under such circumstances libraries complain that they cannot fulfill their public interest mission in the digital era. What libraries are allowed to do under their own under current limitations and exceptions, is seen as inadequate for what is expected of them. But to do more requires the appropriate e-lending licenses from rights holders. In many cases, however, libraries simply cannot license digitally for e-lending. In those cases when licensing is possible, they see transaction costs as prohibitively high; they feel that their bargaining positions vis-à-vis rightholders is unbalanced; they do not see that the license terms are adapted to libraries' policies, and they fear that the licenses provide publishers excessive and undue influence over libraries (*Report on the responses to the Public Consultation on the Review of the EU Copyright Rules*, 2013).

What is more, libraries face substantial legal uncertainties even where there are more-or-less well defined digital library exceptions. In the EU, questions such as whether the analogue lending rights of libraries extend to e-books, whether an exhaustion of the distribution right is necessary to enjoy the lending exception, and whether licensing an e-book would exhaust the distribution right are under consideration by the Court of Justice of the European Union in a Dutch case (Rosati, 2014b). And while in another case (Case C-117/13 Technische Universität Darmstadt v Eugen Ulmer KG) the CJEU reaffirmed the rights of European libraries to digitize books in their collection if that is necessary to give access to them in digital formats on their premises, it also created new uncertainties by stating that libraries may not digitize their entire collections (Rosati, 2014a).

US libraries face a similar situation, both in terms of the narrowly defined exceptions in which libraries can operate, and the huge uncertainty regarding the limits of fair use in the digital library context. US rights holders challenged both Google's (Authors Guild v Google) and the libraries (Authors Guild v HathiTrust) rights to digitize copyrighted works. While there seems to be a consensus of courts that the mass digitization conducted by these institutions was fair use (Diaz, 2013; Rosati, 2014c; Samuelson, 2014), the accessibility of the scanned works is still heavily limited, subject to licenses from publishers, the existence of print copies at the library and the institutional membership held by prospective readers. While in the highly competitive US e-book market many commercial intermediaries offer e-lending

---

[6] The notable exception being orphan works which are presumed to be still copyrighted, but without an identifiable rights owner. In the EU, the Directive 2012/28/EU on certain permitted uses of orphan works in theory eases access to such works, but in practice its practical impact is limited by the many constraints among its provisions. Lacking any orphan works legislation and the Google Book Settlement still in limbo, the US is even farther from making orphan works generally accessible to the public.

licenses to e-book catalogues of various sizes, these arrangements also carry the danger of a commercial lock-in of the access to digital works, and render libraries dependent upon the services of commercial providers who may or may not be the best defenders of public interest (OECD, 2012).

Shadow libraries like Aleph are called into existence by the vacuum that was left behind by the collapse of libraries in the digital sphere and by the inability of the commercial arrangements to provide adequate substitute services. Shadow libraries are pooling distributed resources and expertise over the internet, and use the lack of legal or technological barriers to innovation in the informal sphere to fill in the void left behind by libraries.

What can Aleph teach us about the future of libraries?

The story of Aleph offers two, closely interrelated considerations for the debate on the future of libraries: a legal and an organizational one. Aleph operates beyond the limits of legality, as almost all of its activities are copyright infringing, including the unauthorized digitization of books, the unauthorized mass downloads from e-text repositories, the unauthorized acts of uploading books to the archive, the unauthorized distribution of books, and, in most countries, the unauthorized act of users' downloading books from the archive. In the debates around copyright infringement, illegality is usually interpreted as a necessary condition to access works for free. While this is undoubtedly true, the fact that Aleph provides no-cost access to books seems to be less important than the fact that it provides an access to them in the first place.

Aleph is a clear indicator of the volume of the demand for current books in digital formats in developed and in developing countries. The legal digital availability, or rather, unavailability of its catalogue also demonstrates the limits of the current commercial and library based arrangements that aim to provide low cost access to books over the internet. As mentioned earlier, Aleph's catalogue is mostly of recent books, meaning that 80% of the titles with a valid ISBN number are still in print and available as a new or used print copy through commercial retailers. What is also clear, that around 66% of these books are yet to be made available in electronic format. While publishers in theory have a strong incentive to make their most recent titles available as e-books, they lag behind in doing so.

This might explain why one third of all the e-book downloads in Aleph are from highly developed Western countries, and two third of these downloads are of books without a kindle version. Having access to print copies either through libraries or through commercial retailers is simply not enough anymore. Developing countries are a slightly different case. There, compared to developed countries, twice as many

of the downloads (17% compared to 8% in developed countries) are of titles that aren't available in print at all. Not having access to books in print seems to be a more pressing problem for developing countries than not having access to electronic copies. Aleph thus fulfills at least two distinct types of demand: in developed countries it provides access to missing electronic versions, in developing countries it provides access to missing print copies.

The ability to fulfill an otherwise unfulfilled demand is not the only function of illegality. Copyright infringement in the case of Aleph has a much more important role: it enables the peer production of the library. Aleph is an open source library. This means that every resource it uses and every resource it creates is freely accessible to anyone for use without any further restrictions. This includes the server code, the database, the catalogue and the collection. The open source nature of Aleph rests on the ideological claim that the scientific knowledge produced by humanity, mostly through public funds should be open for anyone to access without any restrictions. Everything else in and around Aleph stems from this claim, as they replicate the open access logic in all the other aspects of Aleph's operation. Aleph uses the peer produced Open Library to fetch book metadata, it uses the bittorrent and ed2k P2P networks to store and make books accessible, it uses Linux and MySQL to run its code, and it allows its users to upload books and edit book metadata. As a consequence of its open source nature, anyone can contribute to the project, and everyone can enjoy its benefits.

It is hard to quantify the impact of this piratical open access library on education, science and research in various local contexts where Aleph is the prime source of otherwise inaccessible books. But it is relatively easy to measure the consequences of openness at the level of the Aleph, the library. The collection of Aleph was created mostly by those individuals and communities who decided to digitize books by themselves for their own use. While any single individual is only capable of digitizing a few books at the maximum, the small contributions quickly add up. To digitize the 1.15 million documents in the Aleph collection would require an investment of several hundred million Euros, and a substantial subsequent investment in storage, collection management and access provision (Poole, 2010). Compared to these figures the costs associated with running Aleph is infinitesimal, as it survives on the volunteer labor of a few individuals, and annual donations in the total value of a few thousand dollars. The hundreds of thousands who use Aleph on a more or less regular basis have an immense amount of resources, and by disregarding the copyright laws Aleph is able to tap into those resources and use them for the development of the library. The value of these resources and of the peer produced library is the difference between the actual costs associated with Aleph, and the investment that would be required to create something remotely similar.

The decentralized, collaborative mass digitization and making available of current, thus most relevant scientific works is only possible at the moment through massive copyright infringement. It is debatable whether the copyrighted corpus of scientific works should be completely open, and whether the blatant disregard of copyrights through which Aleph achieved this openness is the right path towards a more openly accessible body of scientific knowledge. It is also yet to be measured what effects shadow libraries may have on the commercial intermediaries and on the health of scientific publishing and science in general. But Aleph, in any case, is a case study in the potential benefits of open sourcing the library.

Conclusion

If we can take Aleph as an expression of what users around the globe want from a library, then the answer is that there is a strong need for a universally accessible collection of current, relevant (scientific) books in restrictions-free electronic formats. Can we expect any single library to provide anything even remotely similar to that in the foreseeable future? Does such a service have a place in the future of libraries? It is as hard to imagine the future library with such a service as without.

While the legal and financial obstacles to the creation of a scientific library with as universal reach as Aleph may be difficult the overcome, other aspects of it may be more easily replicable. The way Aleph operates demonstrates the amount of material and immaterial resources users are willing to contribute to build a library that responds to their needs and expectations. If libraries plan to only 'host' user-governed activities, it means that the library is still imagined to be a separate entity from its users. Aleph teaches us that this separation can be overcome and users can constitute a library. But for that they need opportunities to participate in the production of the library: they need the right to digitize books and copy digital books to and from the library, they need the opportunity to participate in the cataloging and collection building process, they need the opportunity to curate and program the collection. In other words users need the chance to be librarians in the library if they wish to do so, and so libraries need to be able to provide access not just to the collection but to their core functions as well. The walls that separate librarians from library patrons, private and public collections, insiders and outsiders can all prevent the peer production of the library, and through that, prevent the future that is the closest to what library users think of as ideal.

References

Abramitzky, R., & Sin, I. (2014). *Book Translations as Idea Flows: The Effects of the Collapse of Communism on the Diffusion of Knowledge* (No. w20023). Retrieved from http://papers.ssrn.com/abstract=2421123

Battles, M. (2004). *Library: An unquiet history*. WW Norton & Company.

Benkler, Y. (2006). *The wealth of networks : how social production transforms markets and freedom*. New Haven: Yale University Press.

Bently, L., Davis, J., & Ginsburg, J. C. (Eds.). (2010). *Copyright and Piracy An Interdisciplinary Critique*. Cambridge University Press.

Bodó, B. (2011a). *A szerzői jog kalózai*. Budapest: Typotex.

Bodó, B. (2011b). Coda: A Short History of Book Piracy. In J. Karaganis (Ed.), *Media Piracy in Emerging Economies*. New York: Social Science Research Council.

Bodó, B. (forthcoming). Piracy vs privacy–the analysis of Piratebrowser. *IJOC*.

Commission on the Future of the Library. (2013). Report of the Commission on the Future of the UC Berkeley Library. Berkeley: UC Berkeley.

Committee on the Public Libraries in the Knowledge Society. (2010). The Public Libraries in the Knowledge Society. Copenhagen: Kulturstyrelsen.

Darnton, R. (1982). *The literary underground of the Old Regime*. Cambridge, Mass: Harvard University Press.

Darnton, R. (2003). The Science of Piracy: A Crucial Ingredient in Eighteenth-Century Publishing. *Studies on Voltaire and the Eighteenth Century*, *12*, 3–29.

Diaz, A. S. (2013). Fair Use & Mass Digitization: The Future of Copy-Dependent Technologies after Authors Guild v. Hathitrust. *Berkeley Technology Law Journal*, *23*.

Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society. (2001). *Official Journal L*, *167*, 10–19.

Elst, M. (2005). *Copyright, freedom of speech, and cultural policy in the Russian Federation*. Leiden/Boston: Martinus Nijhoff.

Ermolaev, H. (1997). *Censorship in Soviet Literature: 1917-1991*. Rowman & Littlefield.

Friedberg, M., Watanabe, M., & Nakamoto, N. (1984). The Soviet Book Market: Supply and Demand. *Acta Slavica Iaponica*, *2*, 177–192.

Giblin, R. (2011). *Code Wars: 10 Years of P2P Software Litigation*. Cheltenham, UK ; Northampton, MA: Edward Elgar Publishing.

Johns, A. (2010). *Piracy: The Intellectual Property Wars from Gutenberg to Gates*. University Of Chicago Press.

Judge, C. B. (1934). *Elizabethan book-pirates*. Cambridge: Harvard University Press.

Khan, B. Z. (2004). Does Copyright Piracy Pay? The Effects Of U.S. International Copyright Laws On The Market For Books, 1790-1920. Cambridge, MA: National Bureau Of Economic Research.

Khan, B. Z., & Sokoloff, K. L. (2001). The early development of intellectual property institutions in the United States. *Journal of Economic Perspectives*, *15*(3), 233–246.

Landes, W. M., & Posner, R. A. (2003). *The economic structure of intellectual property law*. Cambridge, Mass.: Harvard University Press.

Lessig, L. (2004). *Free culture : how big media uses technology and the law to lock down culture and control creativity*. New York: Penguin Press.

Liang, L. (2012). Shadow Libraries. *e-flux*. Retrieved from http://www.e-flux.com/journal/shadow-libraries/

Patry, W. F. (2009). *Moral panics and the copyright wars*. New York: Oxford University Press.

Patterson, L. R. (1968). *Copyright in historical perspective* (p. vii, 264 p.). Nashville,: Vanderbilt University Press.

Pendleton-Jullian, A., Lougee, W. P., Wilkin, J., & Hilton, J. (2014). Strategic Thinking and Design—Research Library in 2033—Vision and System of Action—Part One. Colombus, OH: Association of Research Libraries. Retrieved from http://www.arl.org/about/arl-strategic-thinking-and-design/arl-membership-refines-strategic-thinking-and-design-at-spring-2014-meeting

Pollard, A. W. (1916). The Regulation Of The Book Trade In The Sixteenth Century. *Library*, *s3-VII*(25), 18–43.

Pollard, A. W. (1920). *Shakespeare's fight with the pirates and the problems of the transmission of his text*. Cambridge [Eng.]: The University Press.

Poole, N. (2010). *The Cost of Digitising Europe's Cultural Heritage - A Report for the Comité des Sages of the European Commission*. Retrieved from http://nickpoole.org.uk/wp-content/uploads/2011/12/digiti_report.pdf

Report on the responses to the Public Consultation on the Review of the EU Copyright Rules. (2013). European Commission, Directorate General for Internal Market and Services.

Rosati, E. (2014a). Copyright exceptions and user rights in Case C-117/13 Ulmer: a couple of observations. *IPKat*. Retrieved October 08, 2014, from http://ipkitten.blogspot.co.uk/2014/09/copyright-exceptions-and-user-rights-in.html

Rosati, E. (2014b). Dutch court refers questions to CJEU on e-lending and digital exhaustion, and another Dutch reference on digital resale may be just about to follow. *IPKat*. Retrieved October 08, 2014, from http://ipkitten.blogspot.co.uk/2014/09/dutch-court-refers-questions-to-cjeu-on.html

Rosati, E. (2014c). Google Books' Library Project is fair use. *Journal of Intellectual Property Law & Practice*, *9*(2), 104–106.

Rose, M. (1993). *Authors and owners : the invention of copyright*. Cambridge, Mass: Harvard University Press.

Samuelson, P. (2002). Copyright and freedom of expression in historical perspective. *J. Intell. Prop. L.*, *10*, 319.

Samuelson, P. (2014). Mass Digitization as Fair Use. *Communications of the ACM*, *57*(3), 20–22.

Schultz, M. F. (2007). Copynorms: Copyright Law and Social Norms. *Intellectual Property And Information Wealth v01*, *1*, 201.

Sezneva, O. (2012). The pirates of Nevskii Prospekt: Intellectual property, piracy and institutional diffusion in Russia. *Poetics*, *40*(2), 150–166.

Solly, E. (1885). Henry Hills, the Pirate Printer. *Antiquary*, *xi*, 151–154.

Stelmakh, V. D. (2001). Reading in the Context of Censorship in the Soviet Union. *Libraries & Culture*, *36*(1), 143–151.

Suber, P. (2013). *Open Access* (Vol. 1). Cambridge, MA: The MIT Press. doi:10.1109/ACCESS.2012.2226094

Swartz, A. (2008). Guerilla Open Access Manifesto. *Aaron Swartz*. Retrieved from https://archive.org/stream/GuerillaOpenAccessManifesto/Goamjuly2008_djvu.txt

Triaille, J.-P., Dusollier, S., Depreeuw, S., Hubin, J.-B., Coppens, F., & Francquen, A. de. (2013). *Study on the application of Directive 2001/29/EC on copyright and related rights in the information society (the "Infosoc Directive")*. European Union.

Wittmann, R. (2004). Highwaymen or Heroes of Enlightenment? Viennese and South German Pirates and the German Market. *Paper presented at the History of Books and Intellectual History conference*. Princeton University.

Yu, P. K. (2000). From Pirates to Partners: Protecting Intellectual Property in China in the Twenty-First Century. *American University Law*, *50*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=245548

Мошков, М. (1999). Что вы все о копирайте. Лучше бы книжку почитали (Библиотеке копирайт не враг). *Компьютерры*, (300).