Articles from Computational Culture

Book Piracy as Peer Preservation

2014-11-09 01:11:11 matthew

Abstract

In describing the people, books, and technologies behind one of the largest "shadow" libraries in the world, we find a tension between the dynamics of sharing and preservation. The paper proceeds to contextualize contemporary book piracy historically, challenging accepted theories of peer production. Through a close analysis of one digital library's system architecture, software and community, we assert that the activities cultivated by its members are closer to that of conservationists of the public libraries movement, with the goal of preserving rather than mass distributing their collected material. Unlike common peer production models emphasis is placed on the expertise of its members as digital preservations, as well as the absorption of digital repositories. Additionally, we highlight issues that arise from their particular form of distributed architecture and community.

Literature is the secretion of civilization, poetry of the ideal. That is why literature is one of the wants of societies. That is why poetry is a hunger of the soul. That is why poets are the first instructors of the people. That is why Shakespeare must be translated in France. That is why Molière must be translated in England. That is why comments must be made on them. That is why there must be a vast public literary domain. That is why all poets, all philosophers, all thinkers, all the producers of the greatness of the mind must be translated, commented on, published, printed, reprinted, stereotyped, distributed, explained, recited, spread abroad, given to all, given cheaply, given at cost price, given for nothing. [1]

Introduction

The big money (and the bandwidth) in online media is in film, music, and software. Text is less profitable for copyright holders; it is cheaper to duplicate and easier to share. Consequently, issues surrounding the unsanctioned sharing of print material receive less press and scant academic attention. The very words, "book piracy," fail to capture the spirit of what is essentially an Enlightenment-era project, openly embodied in many contemporary "shadow[2] libraries": in the words of Victor Hugo, to establish a "vast public literary domain." Writers, librarians, and political activists from Hugo to Leo Tolstoy and Andrew Carnegie have long argued for unrestricted access to information as a form of a public good essential to civic engagement. In that sense, people participating in online book exchanges enact a role closer to that of a librarian than that of a bootlegger or a plagiarist. Whatever the reader's stance on the ethics of copyright and copyleft, book piracy should not be dismissed as mere search for free entertainment. Under the conditions of "digital disruption,"[3] when the traditional institutions of knowledge dissemination—the library, the university, the newspaper, and the publishing house—feel themselves challenged and transformed by the internet, we can look to online book sharing communities for lessons in participatory governance, technological innovation, and economic sustainability.

The primary aims of this paper are ethnographic and descriptive: to study and to learn from a library that constitutes one of the world's largest digital archives, rivaling *Google Books*, *Hathi Trust*, and *Europeana*. In approaching a "thick description" of this archive we begin to broach questions of scope and

impact. We would like to ask: Who? Where? and Why? What kind of people distribute books online? What motivates their activity? What technologies enable the sharing of print media? And what lessons can we draw from them? Our secondary aim is to continue the work of exploring the phenomenon of book sharing more widely, placing it in the context of other commons-based peer production communities like Project Gutenberg and Wikipedia. The archetypal model of peer production is one motivated by altruistic participation. But the very history of public libraries is one that combines the impulse to share and to protect. To paraphrase Jacques Derrida [4] writing in "Archive Fever," the archive shelters memory just as it shelters itself from memory. We encompass this dual dynamic under the term "peer preservation," where the logistics of "peers" and of "preservation" can sometimes work at odds to one another.

Academic literature tends to view piracy on the continuum between free culture and intellectual property rights. On the one side, an argument is made for unrestricted access to information as a prerequisite to properly deliberative democracy.[5] On this view, access to knowledge is a form of political power, which must be equitably distributed, redressing regional and social imbalances of access.[6] The other side offers pragmatic reasoning related to the long-term sustainability of the cultural sphere, which, in order to prosper, must provide proper economic incentives to content creators.[7]

It is our contention that grassroots file sharing practices cannot be understood solely in terms of access or intellectual property. Our field work shows that while some members of the book sharing community participate for activist or ideological reasons, others do so as collectors, preservationists, curators, or simply readers. Despite romantic notions to the contrary, reading is a social and mediated activity. The reader encounters texts in conversation, through a variety of physical interfaces and within an ecosystem of overlapping communities, each projecting their own material contexts, social norms, and ideologies. A technician who works in a biology laboratory, for example, might publish closed-access peer-review articles by day, as part of his work collective, and release terabytes of published material by night, in the role of a moderator for an online digital library. Our approach then, is to capture some of the complexity of such an ecosystem, particularly in the liminal areas where people, texts, and technology converge.

Ethics disclaimer

Research for this paper was conducted under the aegis of piracyLab, an academic collective exploring the impact of technology on the spread of knowledge globally.[8] One of the lab's first tasks was to discuss the ethical challenges of collaborative research in this space. The conversation involved students, faculty, librarians, and informal legal council. Neutrality, to the extent that it is possible, emerged as one of our foundational principles. To keep all channels of communication open, we wanted to avoid bias and to give voice to a diversity of stakeholders: from authors, to publishers, to distributors, whether sanctioned or not. Following a frank discussion and after several iterations, we drafted an ethics charter that continues to inform our work today. The charter contains the following provisions:

- We neither condone nor condemn any forms of information exchange.
- We strive to protect our sources and do not retain any identifying personal information.
- We seek transparency in sharing our methods, data, and findings with the widest possible audience.
- Credit where credit is due. We believe in documenting attribution thoroughly.
- We limit our usage of licensed material to the analysis of metadata, with results used for non-commercial, nonprofit, educational purposes.
- Lab participants commit to abiding by these principles as long as they remain active members of the research group.

In accordance with these principles and following the practice of scholars like Balazs Bodo [9], Eric Priest [10], and Ramon Lobato and Leah Tang [11], we redact

the names of file sharing services and user names, where such names are not made explicitly public elsewhere.

Centralization

We begin with the intuition that all infrastructure is social to an extent. Even private library collections cannot be said to reflect the work of a single individual. Collective forces shape furniture, books, and the very cognitive scaffolding that enables reading and interpretation. Yet, there are significant qualitative differences in the systems underpinning private collections, public libraries, and unsanctioned peer-to-peer information exchanges like *The Pirate Bay*, for example. Given these differences, the recent history of online book sharing can be divided roughly into two periods. The first is characterized by local, ad-hoc peer-to-peer document exchanges and the subsequent growth of centralized content aggregators. Following trends in the development of the web as a whole, shadow libraries of the second period are characterized by communal governance and distributed infrastructure.

Shadow libraries of the first period resemble a private library in that they often emanate from a single authoritative source—a site of collection and distribution associated with an individual collector, sometimes explicitly. The library of Maxim Moshkov, for example, established in 1994 and still thriving at *lib.ru*, is one of the most visible collections of this kind. Despite their success, such libraries are limited in scale by the means and efforts of a few individuals. Due to their centralized architecture they are also susceptible to legal challenges from copyright owners and to state intervention. Shadow libraries responded to these problems by distributing labor, responsibility, and infrastructure, resulting in a system that is more robust, more redundant, and more resistant to any single point of failure or control.

The case of *Gigapedia* (later *library.nu*) and its related file hosting service *ifile.it* demonstrates the successes and the deficiencies of the centralized digital library model. Arguably among the largest and most popular virtual libraries online in the period of 2009-2011, the sites were operated by Irish nationals[12] on domains registered in Italy and on the island state of Niue, with servers on the territory of Germany and Ukraine. At its peak, *library.nu* (LNU) hosted more than 400,000 books and was purported to make an "estimated turnover of EUR 8 million (USD 10,602,400) from advertising revenues, donations and sales of premium-level accounts," at least according to a press release made by the International Publishers Association (IPA).[13]
Archived version of library.nu, circa 12/10/2010

Its apparent popularity notwithstanding, *LNU/Gigapedia* was supported by relatively simple architecture, likely maintained by a lone developer-administrator. The site itself consisted of a catalog of digital books and related metadata, including title, author, year of publication, number of pages, description, category classification, and a number of boolean parameters (whether the file is bookmarked, paginated, vectorized, is searchable, and has a cover). Although the books could be hosted anywhere, many in the catalog resided on the servers of a "cyberlocker" service *ifile.it*, affiliated with the main site. Not strictly a single-source archive, *LNU/Gigapedia* was nevertheless a federated entity, tied to a single site and to a single individual. On February 15, 2012, in a Munich court, the IPA, in conjunction with a consortium of international publishing houses and the help of the German law firm Lausen Rechtsanwalte,[14] served judicial cease-and-desist orders naming both sites (*Gigapedia* and *ifile.it*). Seventeen injunctions were sought in Ireland, with the consequent voluntary shut-down of both domains, which for a brief time redirected visitors first to *Google Books* and then to *Blue Latitudes*, a *New York Times* bestseller about pirates, for sale on *Amazon*.

Figure 1: Archived version of library.nu, circa 12/10/2010

The relatively brief, by library standards, existence of *LNU/Gigapedia* underscores a weakness in the federated library model. The site flourished as long as it did not attract the ire of the publishing industry. A lack of redundancy in the site's administrative structure paralleled its lack on the server level. Once the authorities were able to establish the identity of the site's operators (via *Paypal* receipts, according to a partner at Lausen Rechtsanwalte), the project was forced to shut down irrevocably.[15] The system's single point of origin proved also to be its single point of failure.

Jens Bammel, Secretary General of the IPA, called the action "an important step towards a more transparent, honest and fair trade of digital content on the Internet."[16] The rest of the internet mourned the passage of "the greatest, largest and the best website for downloading eBooks,"[17] comparing the demise of *LNU/Gigapedia* to the burning of the ancient Library of Alexandria.[18] Readers from around the world flocked to sites like *Reddit* and *TorrentFreak* to express their support and anger. For example, one reader wrote on *TorrentFreak*:

> I live in Macedonia (the Balkans), a country where the average salary is somewhere around 200eu, and I'm a student, attending a MA degree in communication sci. [...] where I come from the public library is not an option. [...] Our libraries are so poor, mostly containing 30year or older editions of books that almost never refer to the field of communication or any other contemporary science. My professors never hide that they use sites like library.nu [...] Original textbooks [...] are copy-printed handouts of some god knows how obtained original [...] For a country like Macedonia and the Balkans region generally THIS IS A APOCALYPTIC SCALE DISASTER! I really feel like the dark age is just around the corner these days.[19]

A similar comment on *Reddit* reads:

> This is the saddest news of the year…heart-breaking…shocking… I was so attached to this site…I am from a third world country where buying original books is way too expensive if we see currency exchange rates…library.nu was a sea of knowledge for me and I learnt a lot from it [...] RIP library.nu…you have ignited several minds with free knowledge.[20]

Another redditor wrote:

> This was an invaluable resource for international academics. The catalog of libraries overseas often cannot meet the needs of researchers in fields not specific to the country in which they are located. My doctoral research has taken a significant blow due to this recent shutdown [...] Please publishers, if you take away such a valuable resource, realize that you have created a gap that will be filled. This gap can either be filled by you or by us.[21]

Another concludes:

> This just makes me want to start archiving everything I can get my hands on.[22]

These anecdotal reports confirm our own experiences of studying and teaching at universities with a diverse audience of international students, who often recount a similar personal narrative. *Gigapedia* and analogous sites fulfilled an unmet need in the international market, redressing global inequities of access to information.[23]

But, being a cyberlocker-based service, *Gigapedia* did not succeed in cultivating a meaningful sense of a community (even though it supported a forum for brief periods of its existence). As Lobato and Tang [24] write in their paper on cyberlocker-based media distribution systems, cyberlockers in general "do not foster collaboration and co-creation," taking an "instrumental view of content hosted on their sites."[25] Although not strictly a cyberlocker, *LNU/Gigapedia* fit the profile of a passive, non-transformative site by these criteria. For Lobato and Tang, the rapid disappearance of many prominent cyberlocker sites underscores the "structural instability" of "fragile file-hosting ecology."[26] In our case, it would be more precise to say that cyberlocker architecture highlights rather the structural instability of centralized media archives, and not of file sharing communities in general. Although bereaved readers were concerned about the irrevocable loss of a valuable resource, digital libraries that followed built a model of file sharing that is more resilient, more transparent, and more participatory than their *LNU/Gigapedia* predecessors.

Distribution

In parallel with the development of *LNU/Gigapedia*, a group of Russian enthusiasts were working on a meta-library of sorts, under the name of *Aleph*. Records of *Aleph's* activity go back at least as far as 2009. Colloquially known as "prospectors," the volunteer members of *Aleph* compiled library collections widely available on the gray market, with an emphasis on academic and technical literature in Russian and English.
DVD case cover of "Traum's library" advertising "more than 167,000 books" in fb2 format. Similar DVDs sell for around 1,000 RUB ($25-30 US) on the streets of Moscow.

At its inception, *Aleph* aggregated several "home-grown" archives, already in wide circulation in universities and on the gray market. These included:

- *KoLXo3*, a collection of scientific texts that was at one time distributed on 20 DVDs, overlapping with early Gigapedia efforts;
- *mexmat*, a library collected by the members of Moscow State University's Department of Mechanics and Mathematics for internal use, originally distributed through private FTP servers;
- *Homelab*, *Ihtik*, and *Ingsat* libraries;
- the Foreign Fiction archive collected from IRC #*** 2003.09-2011.07.09 and the Internet Library;
- the *Great Science Textbooks* collection and, later, over 20 smaller miscellaneous archives.[27]

In retrospect, we can categorize the founding efforts along three parallel tracks: 1) as the development of "front-end" server software for searching and downloading books, 2) as the organization of an online forum for enthusiasts willing to contribute to the project, and 3) the collection effort required to expand and maintain the "back-end" archive of documents, primarily in .pdf and .djvu formats.[28] "What do we do?" writes one of the early volunteers (in 2009) on the topic of "Outcomes, Goals, and Scope of the Project." He answers: "we loot sites with ready-made collections," "sort the indices in arbitrary normalized formats," "for uncatalogued books we build a 'technical index': name of file, size, hashcode," "write scripts for database sorting after the initial catalog process," "search the database," "use the database for the construction of an accessible catalog," "build torrents for

the distribution of files in the collection."[29] But, "everything begins with the forum," in the words of another founding member.[30] *Aleph*, the very name of the group, reflects the aspiration to develop a "platform for the inception of subsequent and more user-friendly" libraries–a platform "useful for the developer, the reader, and the librarian."[31]

Aleph's *anatomy*

Figure 2: DVD case cover of "Traum's library" advertising "more than 167,000 books

What is *Aleph*? Is it a collection of books? A community? A piece of software? What makes a library? When attempting to visualize Aleph's constituents (Figure 3), it seems insufficient to point to books alone, or to social structure, or to technology in the absence of people and content. Taking a systems approach to description, we understand a library to comprise an assemblage of books, people, and infrastructure, along with their corresponding words and texts, rules and institutions, and shelves and servers.[32] In this light, *Aleph*'s iteration on *LNU/Gigapedia* lies not in technological advancement alone, but in system architecture, on all levels of analysis.

Where the latter relied on proprietary server applications, *Aleph* built software that enabled others to mirror and to serve the site in its entirety. The server was written by d* from www.l*.com (Bet), utilizing a codebase common to several similar large book-sharing communities. The initial organizational efforts happened on a sub-forum of a popular torrent tracker (*RR*). Fifteen founding members reached early consensus to start hashing document filenames (using the MD5 message-digest algorithm), rather than to store files as is, with their appropriate .pdf or .mobi extensions.[33] Bit-wise hashing was likely chosen as a (computationally) cheap way to de-duplicate documents, since two identical files would hash into an identical string. Hashing the filenames was hoped to have the side-effect of discouraging direct (file system-level) browsing of the archive.[34] Instead, the books were meant to be accessed through the front-end "librarian" interface, which added a layer of meta-data and search tools. In other words, the group went out of its way to distribute *Aleph* as a library and not merely as a large aggregation of raw files.
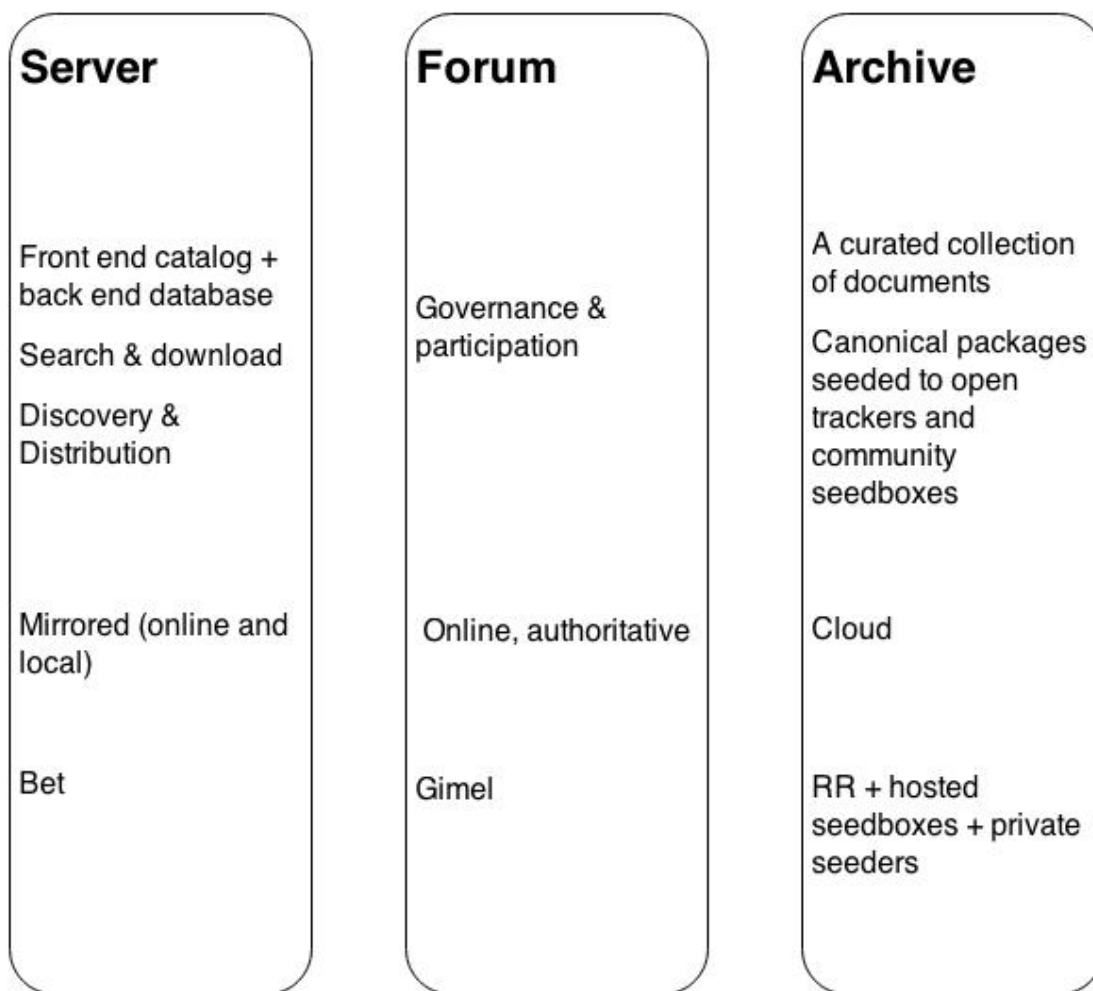
Server	Forum	Archive
Front end catalog + back end database	Governance & participation	A curated collection of documents
Search & download		Canonical packages seeded to open trackers and community seedboxes
Discovery & Distribution		
Mirrored (online and local)	Online, authoritative	Cloud
Bet	Gimel	RR + hosted seedboxes + private seeders

Figure 3: Aleph's anatomy

Site volunteers coordinate their efforts asynchronously, by means of a simple online forum (using *phpBB* software), open to all interested participants. Important issues related to the governance of the project—decisions about new hardware upgrades, software design, and book acquisition—receive public airing. For example, at one point, the site experienced increased traffic from *Google* searches. Some senior members welcomed the attention, hoping to attract new volunteers. Others worried increased visibility would bring unwanted scrutiny. To resolve the issue, a member suggested delisting the website by altering the robots.txt configuration file and thereby blocking *Google* crawlers.[35] Consequently, the site would become invisible to *Google*, while remaining freely accessible via a direct link. Early conversations on *RR*, reflect a consistent concern about the archive's longevity and its vulnerability to official sanctions. Rather than following the cyber-locker model of distribution, the prospectors decided to release canonical versions of the library in chunks, via *BitTorrent*—a distributed protocol for file sharing. Another decision was made to "store" the library on open trackers (like *The Pirate Bay*), rather than tying it to a closed, by-invitation-only community. Although *LN/Gigapedia* was already decentralized to an extent, the archeology of the community discussion reveals a multitude of concious choices that work to further atomize *Aleph* and to decentralize it along the axes of the collection, governance, and engineering.

By March of 2009 these efforts resulted in approximately 79k volumes or around 180gb of data.[36] By December of the same year, the moderators began talking about a terabyte, 2tb in 2010, and around 7tb by 2011.[37] By 2012, the core group of "prospectors" grew to 1,000 registered users. *Aleph*'s main mirror received over a million page views per month and about 40,000 unique visits per day.[38] An online eBook piracy report estimates a combined total of a million unique visitors per day for *Aleph* and its mirrors.[39]

As of January 2014, the *Aleph* catalog contains over a million books (1,021,000)

and over 15 million academic articles, "weighing in" at just under 10tb. Most remarkably, one of the world's largest digital libraries operates on an annual budget of $1,900 US.[40]

Vulnerability
Distributed architecture gives *Aleph* significant advantages over its federated predecessors. Were *Aleph* servers to go offline the archive would survive "in the cloud" of the *BitTorrent* network. Should the forum (*Bet*) close, another online forum could easily take its place. And were *Aleph* library portal itself go dark, other mirrors would (and usually do) quickly take its place.

But the decentralized model of content distribution is not without its challenges. To understand them, we need to review some of the fundamentals behind the *BitTorrent* protocol. At its bare minimum (as it was described in the original specification by Bram Cohen) the protocol involves a "seeder," someone willing to share something it its entirety; a "leecher," someone downloading shared data; and a torrent "tracker" that coordinates activity between seeders and leechers.[41]

Imagine a music album sharing agreement between three friends, where, initially, only one holds a copy of some album: for example, Nirvana's *Nevermind*. Under the centralized model of file sharing, the friend holding the album would transmit two copies, one to each friend. The power of *BitTorrent* comes from shifting the burden of sharing from a single seeder (friend one) to a "swarm"" of leechers (friends two and three). On this model, the first leecher joining the network (friend two, in our case) would begin to get his data from the seeder directly, as before. But the second leecher would receive some bits from the seeder and some from the first leecher, in a non-linear, asynchronous fashion. In our example, we can imagine the remaining friend getting some songs from the first friend and some from the second. The friend who held the album originally now transmitted something less than two full copies of the album, since the other two friends exchanged some bits of information between themselves, lessening the load on the original album holder.

When downloading from the *BitTorrent* network, a peer may receive some bits from the beginning of the document, some from the middle, and some from the end, in parts distributed among the members of the swarm. A local application called the ""client" is responsible for checking the integrity of the pieces and for reassembling the them into a coherent whole. A torrent ""tracker" coordinates the activity between peers, keeping track of who has what where. Having received the whole document, a leecher can, in turn, become a seeder by sharing all of his downloaded bits with the remaining swarm (who only have partial copies). The leecher can also take the file offline, choosing not to share at all.[42]

The original protocol left torrent trackers vulnerable to charges of aiding and abetting copyright infringement.[43] Early in 2008, Cohen extended *BitTorrent* to make use of ""distributed sloppy hash tables" (DHT) for storing peer locations without resorting to a central tracker. Under these new guidelines, each peer would maintain a small routing table pointing to a handful of nearby peer locations. In effect, DHT placed additional responsibility on the swarm to become a tracker of sorts, however ""sloppy" and imperfect. By November of of 2009, *Pirate Bay* announced its transition away from tracking entirely, in favor of DHT and the related PEX and Magnetic Links protocols. At the time they called it, ""world's most resilient tracking."[44]

Despite these advancements, the decentralized model of file sharing remains susceptible to several chronic ailments. The first follows from the fact that ad-hoc distribution networks privilege popular material. A file needs to be actively traded to ensure its availability. If nobody is actively sharing and downloading Nirvana's *Nevermind*, the album is in danger of fading out of the cloud. As one member wrote succinctly on *Gimel* forums, ""unpopular files are in danger of become inaccessible."[45] This dynamic is less of a concern for Hollywood blockbusters, but more so for ""long tail" specialized materials of

the sort found in *Aleph*, and indeed, for *Aleph* itself as a piece of software distributed through the network. *Aleph* combats the problem of fading torrents by renting ""seedboxes"–servers dedicated to keeping the *Aleph* seeds containing the archive alive, preserving the availability of the collection. The server in production as of 2014 can serve up to 12tb of data speeds of 100-800 megabits per second. Other file sharing communities address the issue by enforcing a certain download to upload ratio on members of their network.

The lack of true anonymity is the second problem intrinsic to the *BitTorrent* protocol. Peers sharing bits directly cannot but avoid exposing their IP address (unless these are masked behind virtual private networks or TOR relays). A ""Sybil" attack becomes possible when a malicious peer shares bits in bad faith, with the intent to log IP addresses.[46] Researchers exploring this vector of attack were able to harvest more than 91,000 IP addresses in less than 24 hours of sharing a popular television show.[47] They report that more than 9% of requests made to their servers indicated ""modified clients," which are likely also to be running experiments in the DHT. Legitimate copyright holders and copyright ""trolls" alike have used this vulnerability to bring lawsuits against individual sharers in court.[48]

These two challenges are further exacerbated in the case of *Aleph*, which uses *BitTorrent* to distribute large parts of its own architecture. These parts are relatively large–around 40-50GB each. Long-term sustainability of *Aleph* as a distributed system therefore requires a rare participant: one interested in downloading the archive as a whole (as opposed to downloading individual books), one who owns the hardware to store and transmit terabytes of data, and one possessing the technical expertise to do so safely.

Peer preservation

In light of the challenges and the effort involved in maintaining the archive, one would be remiss to describe *Aleph* merely in terms of book piracy, understood in conventional terms of financial gain, theft, or profiteering. Day-to-day labor of the core group is much more comprehensible as a mode of commons-based peer production, which is, in the canonical definition, work made possible by a ""networked environment," ""radically decentralized, collaborative, and non-proprietary; based on sharing resources and outputs among widely distributed, loosely connected individuals who cooperate with each other without relying on either market signals or managerial commands."[49] *Aleph* answers the definition of peer production, resembling in many respects projects like *Linux*, *Wikipedia*, and *Project Gutenberg*.

Yet, *Aleph* is also patently a library. Its work can and should be viewed in the broader context of Enlightenment ideals: access to literacy, universal education, and the democratization of knowledge. The very same ideals gave birth to the public library movement as a whole at the turn of the 20th century, in the United States, Europe, and Russia.[50] Parallels between free library movements of the early 20th and the early 21st centuries point to a social dynamic that runs contrary to the populist spirit of commons-based peer production projects, in a mechanism that we describe as peer preservation. The idea encompasses conflicting drives both to share and to hoard information.

The roots of many public libraries lie in extensive private collections. Bodleian Library at Oxford, for example, traces its origins back to the collections of Thomas Cobham, Bishop of Worcester, Humphrey, Duke of Gloucester, and to Thomas Bodley, himself an avid book collector. Similarly, Poland's Zaluski Library, one of Europe's oldest, owes its existence to the collecting efforts of the Zaluski brothers, both bishops and bibliophiles.[51] As we mentioned earlier, *Aleph* too began its life as an aggregator of collections, including the personal libraries of Moshkov and Traum. When books are scarce, private libraries are a sign of material wealth and prestige. In the digital realm, where the cost of media acquisition is low, collectors amass social capital. *Aleph* extends its collecting efforts on *RR*, a much larger, moderated torrent exchange forum and tracker. *RR* hosts a number of sub-forums dedicated to the exchange of

software, film, music, and books (where members of *Aleph* often make an appearance). In the exchange economy of symbolic goods, top collectors are known by their standing in the community, as measured by their seniority, upload and download ratios, and the number of ""releases." A release is more than just a file: it must not duplicate items in the archive and follows strict community guidelines related to packaging, quality, and meta-data accompanying the document. Less experienced members of the community treat high status numbers with reverence and respect.

According to a question and answer session with an official *RR* representative, *RR* is not particularly friendly to new users.[52] In fact, high barriers to entry are exactly what differentiates *RR* from sites like *The Pirate Bay* and other unmoderated, open trackers. *RR* prides itself on the ""quality of its moderation." Unlike *Pirate Bay*, *RR* sees itself as a ""media library," where content is ""organized and properly shelved." To produce an acceptable book ""release" one needs to create a package of files, including well-formatted meta-data (following strict stylistic rules) in the header, the name of the book, an image of its cover, the year of release, author, genre, publisher, format, language, a required description, and screenshots of a sample page. The files must be named according to a convention, be ""of the same kind" (that is belong to the same collection), and be of the right size. Home-made scans are discouraged and governed by a 1,000-words instruction manual. Scanned books must have clear attribution to the releaser responsible for scanning and processing.

More than that, guidelines indicate that smaller releases should be expected to be ""absorbed" into larger ones. In this way, a single novel by Charles Dickens can and will be absorbed into his collected works, which might further be absorbed into ""Novels of 19th Century," and then into ""Foreign Fiction" (as a hypothetical, but realistic example). According to the rules, the collection doing the absorbing must be ""at least 50% larger than the collection it is absorbing." Releases are further governed by a subset or rules particular to the forum subsections (e.g. journals, fiction, documentation, service manuals, etc.).[53]

All this to say that although barriers to acquisition are low, the barriers to active participation are high and continually *increase with time*. The absorption of smaller collections by larger favors the veterans. Rules and regulations grow in complexity with the maturation of the community, further widening the rift between senior and junior peers. We are then witnessing something like the institutionalization of a professional ""librarian" class, whose task it is to protect the collection from the encroachment of low-quality contributors. Rather than serving the public, a librarian's primary commitment is to the preservation of the archive as a whole. Thus what starts as a true peer production project, may, in the end, grow to erect solid walls to peering. This dynamic is already embodied in the history of public libraries, where amateur librarians of the late 19th century eventually gave way to their modern degree-holding counterparts. The conflicting logistics of access and preservation may lead digital library development along a similar path.

The expression of this dual push and pull dynamic in the observed practices of peer preservation communities conforms to Derrida's insight into the nature of the archive. Just as the walls of a library serve to shelter the documents within, they also isolate the collection from the public at large. Access and preservation, in that sense, subsist at opposite and sometime mutually exclusive ends of the sharing spectrum. And it may be that this dynamic is particular to all peer production communities, like *Wikipedia*, which, according to recent studies, saw a decline in new contributors due to increasingly strict rule enforcement.[54] However, our results are merely speculative at the moment. The analysis of a large dataset we have collected as corollary to our field work online may offer further evidence for these initial intuitions. In the meantime, it is not enough to conclude that brick-and-mortar libraries should learn from these emergent, distributed architectures of peer preservation. If the future of *Aleph* is leading to increased institutionalization, the community may soon face the fate embodied by its own procedures: the

absorption of smaller, wonderfully messy, ascending collections into larger, more established, and more rigid social structures.

Biographies

Dennis Tenen teaches in the fields of new media and digital humanities at Columbia University, Department of English and Comparative Literature. His research often happens at the intersection of people, texts, and technology. He is currently writing a book on minimal computing, called *Plain Text*.

Maxwell Foxman is an adjunct professor at Marymount Manhattan College and a PhD candidate in Communications at Columbia University, where he studies the use and adoption of digital media into everyday life. He has written on failed social media and on gamification in electoral politics, newsrooms, and mobile media.

References

Allen, Elizabeth Akers, and James Phinney Baxter. *Dedicatory Exercises of the Baxter Building*. Auburn, Me: Lakeside Press, 1889.

Anonymous author. "Library.nu: Modern era's 'Destruction of the Library of Alexandria.'" *Breaking Culture*. Last edited on February 16, 2012 and archived on archived on January 14, 2014. http://breakingculture.tumblr.com/post/17697325088/gigapedia-rip.

Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press, 2006.

Bittorrent.org. "The BitTorrent Protocol Specification." Last modified October 20, 2012 and archived on June 13, 2014. http://www.bittorrent.org/beps/bep_0003.html.

Bodo, Balazs. "Set the Fox to Watch the Geese: Voluntary IP Regimes in Piratical File-Sharing Communities." In *Piracy: Leakages from Modernity*. Litwin Books, LLC, 2012.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. The MIT Press, 1999.

Calandrillo, Steve P. "Economic Analysis of Property Rights in Information: Justifications and Problems of Exclusive Rights, Incentives to Generate Information, and the Alternative of a Government-Run Reward System, an." *Fordham Intellectual Property, Media & Entertainment Law Journal* 9 (1998): 301.

Calhoun, Craig. "Information Technology and the International Public Sphere." *In Shaping the Network Society: the New Role of Civil Society in Cyberspace*, edited by Douglas Schuler and Peter Day, 229–52. MIT Press, 2004.

Castells, Manuel. "Communication, Power and Counter-Power in the Network Society." *International Journal of Communication* 1 (2007): 238–66.

Cholez, Thibault, Isabelle Chrisment, and Olivier Festor. "Evaluation of Sybil Attacks Protection Schemes in KAD." In *Scalability of Networks and Services*, edited by Ramin Sadre and Aiko Pras, 70–82. Lecture Notes in Computer Science 5637. Springer Berlin Heidelberg, 2009.

Cohen, Bram. *Incentives Build Robustness in BitTorrent*, May 22, 2003. http://www.bittorrent.org/bittorrentecon.pdf.

Cohen, Julie. "Creativity and Culture in Copyright Theory." *U.C. Davis Law Review* 40 (2006): 1151.

Day, Brian R. *In Defense of Copyright: Creativity, Record Labels, and the Future of*

Music. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, May 2010.

Derrida, Jacques. "Archive Fever: a Freudian Impression." *Diacritics* 25, no. 2 (July 1995): 9–63.

DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. "Social Implications of the Internet." *Annual Review of Sociology* 27 (January 2001): 307–36.

Edwards, Paul N. "Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems." In *Modernity and Technology*, 185–225, 2003.

———. "Y2K: Millennial Reflections on Computers as Infrastructure." *History and Technology* 15, no. 1-2 (1998): 7–29.

Edwards, Paul N., Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams. "Introduction: an Agenda for Infrastructure Studies." *Journal of the Association for Information Systems* 10, no. 5 (2009): 364–74.

Ernesto. "US P2P Lawsuit Shows Signs of a 'Pirate Honeypot'." Technology. *TorrentFreak*. Last edited in June 2011 and archived on January 14, 2014. http://torrentfreak.com/u-s-p2p-lawsuit-shows-signs-of-a-pirate-honeypot-110601/.

Gauravaram, Praveen, and Lars R. Knudsen. "Cryptographic Hash Functions." In *Handbook of Information and Communication Security*, edited by Peter Stavroulakis and Mark Stamp, 59–79. Springer Berlin Heidelberg, 2010.

Greenwood, Thomas. *Public Libraries: a History of the Movement and a Manual for the Organization and Management of Rate Supported Libraries*. Simpkin, Marshall, Hamilton, Kent, 1890.

Halfaker, Aaron, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. "The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline." *American Behavioral Scientist*, December 2012, 0002764212469365.

Harris, Michael H. *History of Libraries of the Western World*. Fourth Edition. Lanham, Md.; London: Scarecrow Press, 1999.

Hughes, Justin. "Philosophy of Intellectual Property, the." *Georgetown Law Journal* 77 (1988): 287. http://heinonline.org/HOL/Page?handle=hein.journals/glj77&id=309&div=&collection=journals.

Hugo, Victor. *Works of Victor Hugo*. New York: Nottingham Society, 1907.

International Publishers Association. "Publishers Strike Major Blow against Internet Piracy." Last modified February 15, 2012. http://www.internationalpublishers.org/ipa-press-releases/286-publishers-strike-major-blow-against-internet-piracy.

Johnson, Simon for Reuters.com. "Pirate Bay Copyright Test Case Begins in Sweden." Last edited on February 16, 2009 and archived on August 4, 2014. http://uk.reuters.com/article/2009/02/16/tech-us-sweden-piratebay-idUKTRE51F3K120090216.]

Karaganis, Joe, ed. *Media Piracy in Emerging Economies*. Social Science Research Network, March 2011. http://piracy.americanassembly.org/the-report/..

Landes, William M., and Richard A. Posner. *The Economic Structure of Intellectual Property Law*. Harvard University Press, 2003.

Larkin, Brian. "Degraded Images, Distorted Sounds: Nigerian Video and the Infrastructure of Piracy." *Public Culture* 16, no. 2 (2004): 289–314.

———. "Pirate Infrastructures." In *Structures of Participation in Digital Culture*, edited by Joe Karaganis, 74–87. New York: SSRC, 2008.

Lessig, Lawrence. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. The Penguin Press, 2004.

Liang, Lawrence. "Shadow Libraries E-Flux," last edited 2012 and archived on October 14, 2014. http://www.e-flux.com/journal/shadow-libraries/.

Lobato, Ramon, and Leah Tang. "The Cyberlocker Gold Rush: Tracking the Rise of File-Hosting Sites as Media Distribution Platforms." *International Journal of Cultural Studies*, November 2013.

Losowsky, Andrew. "Book Downloading Site Targeted in Injunctions Requested by 17 Publishers." *Huffington Post*, last edited on February 2012 and archived on October 14, 2014. http://www.huffingtonpost.com/2012/02/15/librarynu-book-downloading-injunction_n_1280383.html.

Papacharissi, Zizi. "The Virtual Sphere the Internet as a Public Sphere." *New Media & Society* 4, no. 1 (February 2002): 9–27.

Priest, Eric. "The Future of Music and Film Piracy in China." *Berkeley Technology Law Journal* 21 (2006): 795.

Salmon, Ricardo, Jimmy Tran, and Abdolreza Abhari. "Simulating a File Sharing System Based on BitTorrent." In *Proceedings of the 2008 Spring Simulation Multiconference*, 21:1–:5. SpringSim '08. San Diego, CA, USA: Society for Computer Simulation International, 2008.

Shirky, Clay. *Here Comes Everybody: the Power of Organizing Without Organizations*. New York: Penguin Press, 2008.

Star, Susan Leigh, and Geoffrey C. Bowker. "How to Infrastructure." In *Handbook of New Media: Social Shaping and Social Consequences of ICTs*, Updated Student Edition., 230–46. SAGE Publications Ltd, 2010.

Stuart, Mary. "Creating a National Library for the Workers' State: the Public Library in Petrograd and the Rumiantsev Library Under Bolshevik Rule." *The Slavonic and East European Review* 72, no. 2 (April 1994): 233–58.

———. "'The Ennobling Illusion': the Public Library Movement in Late Imperial Russia." *The Slavonic and East European Review* 76, no. 3 (July 1998): 401–40.

———. "The Evolution of Librarianship in Russia: the Librarians of the Imperial Public Library, 1808-1868." *The Library Quarterly* 64, no. 1 (January 1994): 1–29.

Timpanaro, J.P., T. Cholez, I Chrisment, and O. Festor. "BitTorrent's Mainline DHT Security Assessment." In *2011 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 1–5, 2011.

TPB. "Worlds most resiliant tracking." Last edited November 17, 2009 and archived on August 4, 2014. thepiratebay.se/blog/175

Vik. "Gigapedia: The greatest, largest and the best website for downloading eBooks." Emotionallyspeaking.com. Last edited on August 10, 2009 and archived on July 15, 2012. http://archive.is/g205">http://vikas-gupta.in/2009/08/10/gigapedia-the-greatest-largest-and-the-best-website-for-downloading-free-e-books/.

1. Victor Hugo, *Works of Victor Hugo* (New York: Nottingham Society, 1907), 230. (up)
2. Lawrence Liang, "Shadow Libraries E-Flux," 2012. (up)
3. McKendrick, Joseph. *Libraries: At the Epicenter of the Digital Disruption, The Library Resource Guide Benchmark Study on 2013/14 Library Spending Plans* (Unisphere Media, 2013). (up)
4. "Archive Fever: a Freudian Impression," *Diacritics* 25, no. 2 (July 1995): 9–

63. (up)

5. Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven: Yale University Press, 2006), 92; Paul DiMaggio et al., "Social Implications of the Internet," *Annual Review of Sociology* 27 (January 2001): 320; Zizi Papacharissi "The Virtual Sphere the Internet as a Public Sphere," *New Media & Society* 4.1 (2002): 9–27; Craig Calhoun "Information Technology and the International Public Sphere," in *Shaping the Network Society: the New Role of Civil Society in Cyberspace*, ed. Douglas Schuler and Peter Day (MIT Press, 2004), 229–52. (up)

6. Benkler, *The Wealth of Networks*, 442; Manuel Castells, "Communication, Power and Counter-Power in the Network Society," *International Journal of Communication* (2007): 251; Lawrence Lessig *Free Culture:How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (The Penguin Press, 2004); Clay Shirky Here Comes Everybody: the Power of Organizing Without Organizations (New York: Penguin Press, 2008), 153. (up)

7. Brian R. Day "In Defense of Copyright: Creativity, Record Labels, and the Future of Music," *Seton Hall Journal of Sports and Entertainment Law*, 21.1 (2011); William M. Landes and Richard A. Posner, *The Economic Structure of Intellectual Property Law* (Harvard University Press, 2003). For further discussion see Steve P. Calandrillo, "Economic Analysis of Property Rights in Information: Justifications and Problems of Exclusive Rights, Incentives to Generate Information, and the Alternative of a Government-Run Reward System" *Fordham Intellectual Property, Media & Entertainment Law Journal* 9 (1998): 306; Julie Cohen, "Creativity and Culture in Copyright Theory," *U.C. Davis Law Review* 40 (2006): 1151; Justin Hughes "Philosophy of Intellectual Property," *Georgetown Law Journal* 77 (1988): 303. (up)

8. piracylab.org. (up)

9. "Set the Fox to Watch the Geese: Voluntary IP Regimes in Piratical File-Sharing Communities, in *Piracy: Leakages from Modernity* (Litwin Books, LLC, 2012). (up)

10. "The Future of Music and Film Piracy in China," *Berkeley Technology Law Journal* 21 (2006): 795. (up)

11. "The Cyberlocker Gold Rush: Tracking the Rise of File-Hosting Sites as Media Distribution Platforms," *International Journal of Cultural Studies*, (2013). (up)

12. The injunctions name I* and F* N* (also known as Smiley). (up)

13. "Publishers Strike Major Blow against Internet Piracy" last modified February 15, 2012 and archived on January 10, 2014, http://www.internationalpublishers.org/ipa-press-releases/286-publishers-strike-major-blow-against-internet-piracy. (up)

14. Including the German Publishers and Booksellers Association, Cambridge University Press, Georg Thieme, Harper Collins, Hogrefe, Macmillan Publishers Ltd., Cengage Learning, Elsevier, John Wiley & Sons, The McGraw-Hill Companies, Pearson Education Ltd., Pearson Education Inc., Oxford University Press, Springer, Taylor & Francis, C.H. Beck as well as Walter De Gruyter. The legal proceedings are also supported by the Association of American Publishers (AAP), the Dutch Publishers Association (NUV), the Italian Publishers Association (AIE) and the International Association of Scientific Technical and Medical Publishers (STM). (up)

15. Andrew Losowsky, "Book Downloading Site Targeted in Injunctions Requested by 17 Publishers," *Huffington Post*, accessed on September 1, 2014, http://www.huffingtonpost.com/2012/02/15/librarynu-book-downloading-injunction_n_1280383.html. (up)

16. International Publishers Association. (up)

17. Vik, "Gigapedia: The greatest, largest and the best website for downloading eBooks," Emotionallyspeaking.com, last edited on August 10, 2009 and archived on July 15, 2012, http://archive.is/g205">http://vikas-gupta.in/2009/08/10/gigapedia-the-greatest-largest-and-the-best-website-for-downloading-free-e-books/. (up)

18. Anonymous author, "Library.nu: Modern era's 'Destruction of the Library

of Alexandria,'" *Breaking Culture* (on tublr.com), last edited on February 16, 2012 and archived on January 14, 2014, http://breakingculture.tumblr.com/post/17697325088/gigapedia-rip. (up)

19. http://torrentfreak.com/book-publishers-shut-down-library-nu-and-ifile-it-120215 archived on January 10, 2014. (up)

20. http://www.reddit.com/r/trackers/comments/ppfwc/librarynu_admin_the_website_is_shutting_down_due archived on January 10, 2014. (up)

21. http://www.reddit.com/r/trackers/comments/ppfwc/librarynu_admin_the_website_is_shutting_down_due orchived on January 10, 2014. (up)

22. www.reddit.com/r/trackers/comments/ppfwc/librarynu_admin_the_website_is_shutting_down_due archived on January 10, 2014. (up)

23. This point is made at length in the report on media piracy in emerging economies, released by the American Assembly in 2011. See Joe Karaganis, ed. *Media Piracy in Emerging Economies* (Social Science Research Network, March 2011), http://piracy.americanassembly.org/the-report/, I. (up)

24. Lobato and Tang, "The Cyberlocker Gold Rush." (up)

25. Lobato and Tang, "The Cyberlocker Gold Rush," 9. (up)

26. Lobato and Tang, "The Cyberlocker Gold Rush," 7. (up)

27. GIMEL/viewtopic.php?f=8&t=169; GIMEL/viewtopic.php?f=17&t=299. (up)

28. GIMEL/viewtopic.php?f=17&t=299. (up)

29. GIMEL/viewtopic.php?f=8&t=169. All quotes translated from Russian by the authors, unless otherwise noted. (up)

30. GIMEL/viewtopic.php?f=8&t=6999&p=41911. (up)

31. GIMEL/viewtopic.php?f=8&t=757. (up)

32. In this sense, we see our work as complementary to but not exhausted by infrastructure studies. See Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (The MIT Press, 1999); Paul N. Edwards, "Y2K: Millennial Reflections on Computers as Infrastructure," *History and Technology* 15.1-2 (1998): 7–29; Paul N. Edwards, "Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems," in *Modernity and Technology*, 2003, 185–225; Paul N. Edwards et al., "Introduction: an Agenda for Infrastructure Studies," *Journal of the Association for Information Systems* 10.5 (2009): 364–74; Brian Larkin "Degraded Images, Distorted Sounds: Nigerian Video and the Infrastructure of Piracy," *Public Culture* 16.2 (2004): 289–314; Brian Larkin "Pirate Infrastructures," in *Structures of Participation in Digital Culture*, ed. Joe Karaganis (New York: SSRC, 2008), 74–87; Susan Leigh Star and Geoffrey C. Bowker, "How to Infrastructure," in *Handbook of New Media: Social Shaping and Social Consequences of ICTs*, (SAGE Publications Ltd, 2010), 230–46. (up)

33. For information on cryptographic hashing see Praveen Gauravaram and Lars R. Knudsen, "Cryptographic Hash Functions," in *Handbook of Information and Communication Security*, ed. Peter Stavroulakis and Mark Stamp (Springer Berlin Heidelberg, 2010), 59–79. (up)

34. See GIMEL/viewtopic.php?f=8&t=55kj and GIMEL/viewtopic.php?f=8&t=18&sid=936. (up)

35. GIMEL/viewtopic.php?f=8&t=714. (up)

36. GIMEL/viewtopic.php?f=8&t=47. (up)

37. GIMEL/viewtopic.php?f=17&t=175&hilit=RR&start=25. (up)

38. GIMEL/viewtopic.php?f=17&t=104&start=450. (up)

39. URL redacted; These numbers should be taken as a very rough estimate because 1) we do not consider Alexa to be a reliable source for web traffic and 2) some of the other figures cited in the report are suspicious. For example, *Aleph* has a relatively small archive of foreign fiction, at odds with the reported figure of 800,000 volumes. (up)

40. GIMEL/viewtopic.php?f=17&t=7061. (up)

41. "The BitTorrent Protocol Specification," last modified October 20, 2012 and archived on June 13, 2014, http://www.bittorrent.org/beps/bep_0003.html. (up)

42. For more information on BitTorrent, see Bram Cohen, *Incentives Build Robustness in BitTorrent*, last modified on May 22, 2003, http://www.bittorrent.org/bittorrentecon.pdf; Ricardo Salmon, Jimmy Tran, and Abdolreza Abhari, "Simulating a File Sharing System Based on BitTorrent," in *Proceedings of the 2008 Spring Simulation Multiconference*,

SpringSim '08 (San Diego, CA, USA: Society for Computer Simulation International, 2008), 21:1–5. (up)

43. In 2008 *The Pirate Bay* co-founders Peter Sunde, Gottfrid Svartholm Warg, Fredrik Neij, and Carl Lundstromwere were charged with "conspiracy to break copyright related offenses" in Sweden. See Simon Johnson for Reuters.com, "Pirate Bay Copyright Test Case Begins in Sweden," last edited on February 16, 2009 and archived on August 4, 2014, http://uk.reuters.com/article/2009/02/16/tech-us-sweden-piratebay-idUKTRE51F3K120090216. (up)

44. TPB, "Worlds most resiliant tracking," last edited November 17, 2009 and archived on August 4, 2014, thepiratebay.se/blog/175. (up)

45. GIMEL/viewtopic.php?f=8&t=6999. (up)

46. Thibault Cholez, Isabelle Chrisment, and Olivier Festor "Evaluation of Sybil Attacks Protection Schemes in KAD," in *Scalability of Networks and Services*, ed. Ramin Sadre and Aiko Pras, Lecture Notes in Computer Science 5637 (Springer Berlin Heidelberg, 2009), 70–82. (up)

47. J.P. Timpanaro et al., "BitTorrent's Mainline DHT Security Assessment," in *2011 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2011, 1–5. (up)

48. Ernesto, "US P2P Lawsuit Shows Signs of a 'Pirate Honeypot'," Technology, *TorrentFreak*, last edited in June 2011 and archived on January 14, 2014, http://torrentfreak.com/u-s-p2p-lawsuit-shows-signs-of-a-pirate-honeypot-110601/. (up)

49. Benkler *The Wealth of Networks*, 60. (up)

50. On the free and public library movement in England and the United States see Thomas Greenwood, *Public Libraries: a History of the Movement and a Manual for the Organization and Management of Rate Supported Libraries* (Simpkin, Marshall, Hamilton, Kent, 1890); Elizabeth Akers Allen and James Phinney Baxter, *Dedicatory Exercises of the Baxter Building* (Auburn, Me: Lakeside Press, 1889). To read more about the history of free and public library movements in Russia see Mary Stuart, "The Evolution of Librarianship in Russia: the Librarians of the Imperial Public Library, 1808-1868," *The Library Quarterly* 64.1 (January 1994): 1–29; Mary Stuart, "Creating a National Library for the Workers' State: the Public Library in Petrograd and the Rumiantsev Library Under Bolshevik Rule," *The Slavonic and East European Review* 72.2 (April 1994): 233–58; Mary Stuart "The Ennobling Illusion: the Public Library Movement in Late Imperial Russia," *The Slavonic and East European Review* 76.3 (July 1998): 401–40. (up)

51. Michael H. Harris, *History of Libraries of the Western World*, (London: Scarecrow Press, 1999), 136. (up)

52. http://s*.d*.ru/comments/508985/. (up)

53. RR/forum/viewtopic.php?t=1590026. (up)

54. Aaron Halfaker et al."The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline," *American Behavioral Scientist*, December 2012. (up)