

Manuel Bonik and Dr. Andreas Schaale

### **Piracy and STM publishers – a view on the worst side of the Internet**

Since the introduction of ebooks the publishing industry has to deal with the problem of piracy. Until recently this problem has concerned only other industries like the software, film or music industry. The STM (Science Technology Medicine) publishers have been affected in particular. On the one hand the scientific text books editions are relatively small (compared to fiction) on the other hand there is the extreme “must have” factor: students and scientists *must* own these books for their studies and research work. There exists a larger number of illegal book sharing portals in the internet. We know (e. g. from the numbers published by the internet metrics alexa.com) that the interested students (and not only they) are making extensive use of these offers. One can assume, that the “replacement rate” (using pirated books instead of buying them) and the lost revenues are high for the STM publishers.

Some (German) statements of those publishers (nonfiction) indicate, that many trade publishing houses are not aware of that problem. It is difficult to measure the actual rate of piracy. The number of illegally distributed books can be only estimated by analyzing the well known piracy sites. As far as that kind of analysis is the prerequisite for taking down those download links (most of the file hosters are supporting the notice and takedown procedure based on the Digital Millennium Copyright Act – DMCA), we have an overview of the current situation. The DMCA is at the moment the only way to fight the illegal distribution of books via direct downloads from file hosters (not P2P networks, which actually are an insignificant problem for the STM publishers), by taking the corresponding links down.

Behind the popular piracy sites there exists some kind of infrastructure, where the DMCA doesn't work. The rules are different. Let us call the “mother of the ebook piracy sites” simply “L”. We do not want to support any kind of piracy. Therefore we will mention the real name of the site only to authorized persons.

“L” is not a very popular domain (based on comparisons of the traffic rank with other well known sites). However, “L” plays an important role in the “back office” of many illegal libraries. In a certain sense this side is a wholesale warehouse for other ebook pirates. Analogous to the antique library of Alexandria (“L” is at present about 4 times larger), “L” exchanges copies of books with many other piracy libraries, the recently closed side library.nu was one of them. “L” stores copies of books in a “safe” way – the servers are out of legal reach and do not cooperate in the DMCA process (there are 4 independent mirrors for the books). The links are “always on”. The way the cataloging (tagging of items) is organized one can assume that there are experienced librarians behind that side. Further one can say that there is no visible commercial interest (except donations) – it seems that the (known) maker of the side wants to distribute “knowledge for everybody”.

It is correspondingly difficult (or impossible) to start legal steps against this side. The infrastructure (domain, server, mirrors) are distributed over several countries and even continents. “L” has an excellent camouflage.

What does the library “Alexandria 2.0” contain (this comparison is not unjustified)? At present “L” stores and offers for download more than 800.000 STM books. This is not a small number! The books are in different languages:

Books in the main languages offered on „L“:

<b>Language</b>	<b>Number</b>	<b>Share</b>
English	530.019	64,3%
Russian	211.690	25,7%
German	27.899	3,4%
French	11.163	1,4%
Spanish	7.671	0,9%
Chinese	1.361	0,2%
Arabic	1.298	0,2%

Among the books there are current ones, as well as older books. The following table gives an overview about the current books:

Books by publication year (2007-2012):

<b>Year</b>	<b>Number</b>
2012	1.701
2011	29.462
2010	41.699
2009	47.234
2008	50.074
2007	52.701

These statistical data are incomplete, because about 30% of the books are without the information of the publication year.

The statistics of the publishers which are affected by „L“ represents the „Who is who“ of the leading STM publishers of the world.

Number of books at “L” of the top publishers (status of March and April 2012)

<b>Publisher</b>	<b>Mar 9 2012</b>	<b>Apr 18 2012</b>
Springer	46.036	48.009
Wiley	20.565	21.675
Cambridge Acad. Press	13.627	14.364
Oxford Univ. Press	10.432	10.957
Routledge	8.889	9.074
CRC	7.159	8.186
McGraw Hill	7.206	7.750
Academic Press	5.390	5.698

Palgrave M M	4.687	4.804
Elsevier	4.203	4.555
Harper	3.872	4.013
World Scientific Pub.	3.436	3.526
Osprey	3.236	3.237
Prentice Hall	3.010	3.222
O'Reilly	2.994	3.015
Humana	2.810	2.831
"For Dummies" (Wiley)	1.714	2.684
Penguin	2.473	2.628
MIT Press	2.088	2.619
Addison Wesley	2.479	2.570
Princeton Univ. Press	2.101	2.155
Birkhäuser	2.021	2.105
Apress	1.998	2.026
Random House	1.633	1.743
California Univ. Press	1.648	1.700
Taylor & Francis	1.328	1.510
SAGE	1.396	1.436
Butterworth H	1.349	1.368
Lippincott W W	1.134	1.172
Harvard Univ. Press	984	1.153
Packt	1.122	1.148
Newnes	954	967
Pearson	859	942
Simon & Schuster	792	813
MacMillan	725	790

The statistics are based on the numbers which are published by „L“ and by the data of the search engine of the side. Besides the „standard data“ (author, title, publisher, ...) „L“ also offers the md5 hash value of the files. This makes it rather easy to see the secondary distribution of those books on other piracy sides.

In general there exist more copies of different scientific text books in the internet. However, this statistics does give a unique insight view into the ebook piracy – it determines the *lower limit* of pirated books. One can estimate, that the number of 1.000.000 pirated science books on “L” alone will be reached in 2012. Those numbers will grow further. During 5 months the amount of books has grown from 699.000 up to 824.000 in March 2012. In the time period from March 9 2012 until April 18 “L” has grown by another 6%.

It should be noted here, that the numbers of books allow only a limited conclusion how strong the piracy affects certain publishers. The number of downloads (and the replacement rate) are depending on the fact, that some publishers are taking effective measures against the distribution on popular piracy sites – others don't.

For further information you may contact [m.bonik@lisheennageeha-consulting.ie](mailto:m.bonik@lisheennageeha-consulting.ie) .