Googling for Ahab or E-Bookville revisited

S. Berner
sberner@ecognus.com

In 2001, which seems like pre-history in Internet time, I wrote a research report about e-books in which I said, *inter alia,* that

> E-books seem to have a chance of success if a number of issues are addressed by all involved sectors: the publishers, the hardware manufacturers, the legal sector and the writers.... To succeed, electronic books must be widely accessible by all segments of the population. Educational institutions should play a large role in improving society's technological literacy, while ensuring that the content remains strong and valuable. Manufacturers and content owners should put the general benefit of the society above their profit making schemes, and assist in making their products available to the public domain (Berner 2001).

Only a year later, Google launched its 'book search' project and took the world by storm. The whole idea seems totally mad: to digitize millions of books currently gathering dust in the world's largest academic libraries. The aim? To enable the millions of info-hungry Web-surfers to Google '+Ahab +whale' and find the required lines in Moby Dick. It was a dream that Google co-founders Sergey Brin and Larry Page had in 1996, 'that in a future world in which vast collections of books are digitized, people would use a "web crawler" to index the books' content and analyse the connections between them, determining any given book's relevance and usefulness by tracking the number and quality of citations from other books' (Google 2007). The project started in 2002, and in the first two years managed to hook not only the Bodleian with its one million plus manuscripts, but also a number of printing presses of world renown: Blackwell, Cambridge University Press, the University of Chicago Press, Houghton Mifflin, Hyperion, McGraw-Hill, Oxford University Press, Pearson, Penguin, Perseus, Princeton University Press, Springer, Taylor & Francis, Thomson Delmar and Warner Books. Then Harvard, the University of Michigan, the New York Public Library, Oxford and Stanford come on board with over 15 million books ready for digitalization. In 2005, Google started similar partnerships with other countries, namely, Belgium, France, Germany, Italy, the Netherlands, Spain and Switzerland.

Everyone started talking about Google Books, and the issue is still a hot topic on-line. Blogs are being dedicated to it; legal turf wars are fought over it, and it might even cause the start of the Franco-Anglaise Language War, if we are to take the French seriously (the battle has been flickering since last year). In 2006, the president of France's Bibliothque Nationale, Jean Noël Jeanneney, wrote a book in which he shared his fears of the impact of Google Books on European culture, stating that 'by the very nature of the library collections that Google proposes to put online, American and British works will dominate, leaving behind that portion of the world's hundred million books not in English' (Knoblauch 2007) So for Jeanneney, Google Books, far from being the beneficial instrument of spreading knowledge, will damage the world's cultural heritage, presumably because it is in English and thus will 'extend the dominance of American culture abroad'. So for the French, it is either 'cultural diversity' or no culture whatsoever, not if it is in English! Possibly the greatest tragi-comic

aspect of this debacle is the fact that Jeanneney's book is on Google Books ([here](#)) in, of course, English translation.

But of course Jeanneney is totally out of date. In March of this year, the Bavarian State Library announced a partnership with Google to scan more than a million public domain and out-of-print works in German as well as English, French, Italian, Latin, and Spanish. A week later, they were followed by the Cantonal and the University Library of Lausanne, and the University of Mysore announced an agreement to digitize 800000 books and other documents, both those on paper and ancient ones on palm leaves. At the same time, the Boekentoren library of the Ghent University will participate with 19th century books in the French and Dutch language, and just last month the Keio University became Google's first library partner in Japan with the announcement that they would digitize at least 120000 public domain books. Other US universities are coming on board as well: the California System (34 million volumes), Wisconsin-Madison (7,2 million volumes), the University of Texas at Austin (one million) and Cornwell (500000 books). One just wonders where the time and energy to scan all this will come from – not to mention finances. Apparently, some of the books are already poorly scanned, if we are to reason by the feedback mechanism for reporting illegible or missing pages that Google Books provides.

Not just the information nerds, but some publishers and lawyers went up in arms about it, albeit on opposite fronts. Jason Epstein preached, 'embrace the Internet or die' to publishers. The whole notion of 'fair use' got whacked on the proverbial wall and dismantled – no one is sure yet what it will become after this deconstructionist exercise, as the courts are still discussing the matter. Farhad Manjoo, the Cornwell graduate who now writes for Salon, raised a very important question. To quote him, 'if copyright law stands in the way of Google's grand aim, isn't it time we thought about changing the law? … The company … is poised to create a tool that could truly change the way we understand, and learn about, the world around us.… Can we really afford to let content owners stand in the way of Google's revolutionary idea?' (Manjoo 2005)

In October 2005, the Association of American Publishers, which represents large publishing houses, sued Google for copyright infringement and also for costing the book industry a great deal of potential revenue. If the publishers worry that the perceived infringement of copyright by Google Books will adversely affect the sales of their already out-of-print books, then one worries about their sanity. OCLC, a non-profit library research group, set out to count and catalogue the books Google would capture in its project and determined that at the five research libraries with which Google had formed deals, about 80% of the books in the stacks were published after 1923 and were still under copyright, but only a small number of these books are currently in print (Lavoie *et.al*. 2005).

As for books still in print, Google made it clear in Frankfurt as early as 2004 that 'for each book found, a user would see several pages of the book with the phrase or subject of the search highlighted. The page would also offer links to several online retailers, where the book could be bought. Publishers do not pay to participate in the programme; rather, Google would make money from the service by selling advertising on the search pages, and it would share those revenues with the publishing companies' (Webb 2004).

Maybe the only way publishers think a book would sell would be to maintain as much secrecy about its content as is possible. Disclosure might drop sales, and it will have nothing to do with being able to read it on-line. This whole notion is ridiculous and yet, in 2005, when Google Books became public news, the American Authors Guild sued Google on the premise that 'it's not up to Google or anyone other than the authors, the rightful owners of these copyrights, to decide whether and how their works will be copied' (Mills 2005). Take this legal individualism to its logical end and it will be the author's right to decide which

library carries his books, where they are sold, and who reads them. This is the proverbial 'shot-in-one's-foot'. As one of the un-offended writers said,

> 'the large majority of current author fear regarding digitized, accessible versions of their work is based on two primary factors: Ignorance and ego. The ignorance is the lack of understanding that for the vast majority of authors, the ability to pop up in an Internet search on a subject would be a *good* thing: It's free publicity and also acts as a taster for people who (very likely) have no idea who you are and what your writing is like. The ego is the assumption that a whole bunch of people are just *gagging* to steal one's work at the slimmest opportunity' (Scalzi 2005).

Not to be left behind, in 2006 the French joined in the legal bullfight, when the French publishing group La Martiniere sued Google for 'piracy'. It seems that La Martiniere owns interests in the US.

Fred von Lohmann, an attorney at the Electronic Freedom Foundation, said that the Google Books lawsuits were a one-sided situation, and that the publishers had no argument to prove that it harmed their sales. And while a significant number of library books are protected by copyright, they are also out of print – 70% or more by some estimates. Someone owns these books, but since they are perceived to have no commercial value (because they are no longer sold in stores), publishers do not have any incentive to promote and market them, let alone to go through the expense of scanning them and making them searchable on-line (O'Reilly 2005). So why not let Google, and why not let the knowledge-hungry world have them for free? Or at least get to know about their existence? Apparently, according to the media expert Siva Vaidhyanathan of the New York University, there is no national registry of copyright holders in the United States, as there is a national registry of patents. 'It's impossible for a company like Google, or a historian, or a documentary filmmaker, or anyone to find out who owns what. Even publishers don't know what they own. It's just impossible' (Manjoo 2005). And they are suing Google demanding that it does find the unknown and asks their permission.

The Court cases are still dragging and no one is saying much about what is happening. But it looks like a bit of a paradox is happening: first politicians pontificate about the need to provide life-long education to the masses (both 'education' and 'masses' being undefined concepts) as the 'in-thing' for the 21st century. Then it is rendered impossible by those who 'own' the knowledge as they send their legal bouncers to bash any institution that takes these pontifications seriously, behaving like 18th century luddites.

No one is even considering the benefit Google books would have on the third world. This makes one feel sorry that we are faster at sending in tanks and bombers, than we are at giving the world access to some good books.

References

Berner, S. (2001) A Review of Current Issues Affecting the Future Marketability of Electronic Books, unpublished report, p.16.

Google (2007) History of Google Book Search. [Online]. Available WWW: http://books.google.com/googlebooks/newsviews/history.html.

Knoblauch, M. (2007) Book Review. [Online]. Available WWW: http://www.amazon.com/Google-Myth-Universal-Knowledge-Europe/dp/0226395782.

Lavoie, B. *et al.* (2005) Anatomy of Aggregate Collections. [Online]. Available: http://www.dlib.org/dlib/september05/lavoie/09lavoie.html.

Manjoo, F. (2005) Throwing Google At the Book. [Online]. Available WWW: http://dir.salon.com/story/tech/feature/2005/11/09/google/index.html?pn=4.

Mills, E. (2005) Authors Guild sues Google over library project. [Online]. Available WWW: http://news.com.com/Authors+Guild+sues+Google+over+library+project/2100-1030_3-5875384.html.

O'Reilly, T. (2005) Search and Rescue. [Online]. Available WWW: http://radar.oreilly.com/archives/2005/09/ny_times_op_ed_on_authors_guil.html.

Scalzi, J. (2005) Google Books. [Online]. Available WWW: http://www.scalzi.com/whatever/003738.html.

Webb, C. (2004) Google Books It To the Finish Line. In Washington Post, 08/10/2004. [Online]. Available WWW: http://www.washingtonpost.com/wp-dyn/articles/A17498-2004Oct8.html.

About the author
Sam Berner (B.Ed., Dipl. LIS, Postgraduate Diploma in Information Management) is a principal of the company ECognus (Brisbane, Australia). She is a knowledge management consultant, assisting small to medium enterprises to benefit the most from their intellectual assets. ECognus also provides services in the area of tailored software applications and the digitization of business processes.