

## Open Source, Open Science, OpenCourseWare

*BJÖRK, Bo-Christer*

*Swedish School of Economics and Business Administration, Helsinki, Finland*

*<http://www.shh.fi/>*

*The Internet has in just a few years radically changed the technical foundation for how the supply chain of scientific publications and teaching materials functions. As researchers we can with just a few clicks find a significant part of all the information we need for free on the World Wide Web. As teachers we can find huge amounts of digital material which can be downloaded or linked from the web and included in presentation overheads, or hyperlinked as reading material. Yet the business and legal (copyright issues) infrastructure has hardly changed and presents a barrier to innovation and reengineering of the overall process. This paper describes some recent trends in how the Internet influences these two fields (publication of research results and production of teaching material) as well as related developments in the organisation of software development, and discusses them both from an economic and philosophical perspective.*

**Keywords:** *Internet, scientific publication, teaching material, open source*

### Introduction

During the 20<sup>th</sup> century the number of university researchers and teachers has increased tremendously. Where universities a century ago were calm and secluded havens of thought and contemplation for the privileged few, there are now millions of academics worldwide, competing in a rat race for academic advancement and external research funding. The requirements to produce results rather quickly have increased (“publish-or-perish”), and it seems that nowadays you almost have to include a business plan for commercial implementation with most applications for research grants. Keeping up with the latest developments has become a critical issue in most disciplines. The issue of knowledge management is today vital for the academic world.

The technology that can help us communicate and share information has fortunately evolved in parallel with the above-mentioned increasing demands. The fraternal university colleges of a hundred years back, where research issues were discussed over a glass of port at the faculty club, have

been replaced by virtual communities of researchers and teachers sharing common interests. These communities are supported by a range of technologies which indirectly or directly facilitate the sharing of information on an unprecedented scale. Cheap air travel enables us to meet in person with colleagues from distant countries, the telephone network, including cellular phones, makes it possible to get hold of a person rapidly, and last but not least the Internet makes it possible to send, duplicate and retrieve information in a very cost-effective way. In particular data exchange in digital form over the Internet makes it very easy to reuse, adapt and link information.

In the following paper I'm going to discuss some aspects of this Internet-enabled exchange. It has become increasingly clear in the last few years that the Internet revolution has triggered a paradigm shift, which challenges the infrastructure of many areas of society. Symptomatic for the concerns that the old establishment has, is the recent accusation from a highly placed Microsoft spokesman that the Linux movement is engaging in Un-American activities, since

it promotes the free sharing of knowledge, rather than profit-orientation and closed, proprietary solutions.

I'm going to discuss recent developments in three related fields, namely:

- Development and distribution of computer software
- The scientific publication process
- Exchange of teaching material

Information sharing in these three domains has technically been vastly facilitated by the emergence of the Internet. Nevertheless some of the practices for intellectual property and copyright that evolved during the last century, before the emergence of the Internet, have created barriers to an optimal use of these technical opportunities. The reasons are mostly the commercial interest of parties such as software vendors and commercial publishers.

For two of the above categories strong idealistic movements have developed, which challenge the current paradigms and promote the free unrestricted dissemination on the Internet as the optimal solution from the viewpoint of society as a whole. These two movements challenge the currently dominating paradigm, which advocates the commercial enterprise model as the best solution for achieving optimal production and/or an optimal supply chain. For the third type of product, academic teaching material, the discussion has hardly even started.

### **Open source**

The open source movement, as exemplified by the culture that has evolved around the operating system Linux, is based on the hypotheses that open, non-commercial sharing and co-development of computer code, leads to better technical solutions than the prevailing business paradigm which stresses copyright and where the inner workings of a software product are kept secret from customers obtaining a license (Ljungberg 2000). One of the most famous essays about the open source movement is Eric S. Raymond's (2000) Cathedral and the Bazaar, which originally was published on the web. In it Raymond

likens the organisation of Open Source projects to a bazaar, and traditional systems development efforts to cathedrals. Curiously, when Netscape announced a new version of its web-browser a couple of years ago, it released it as open source, partly influenced by Raymond's writings!

Pekka Himanen has discussed the philosophical underpinnings of the hacker culture, from which the open source movement emerged (Himanen 2001). Going a bit deeper than Raymond's in readings of philosophy and in particular work ethics, he compares the organisation of the hacker culture that has nurtured the open source development to the original academy of Athens and the organisation of mainstream software development to a medieval monastery. He discusses the main motivations that drive hackers as survival, social connections & recognition and pleasure seeking and finds that the latter two (provided the basic economic conditions for survival are catered for) dominate as motivations for how they organise their work, rather than maximisation of financial rewards.

The true long-term impact of the open source movement on software development and distribution is still largely unproven. It may be that open source as a way of organising information system development is effective only for certain specialised niche products, which have a truly global user base. Also the motivations of the hackers, as discussed by Raymond and Himanen, are somewhat reminiscent of the Hippie culture, and may not be tenable as solutions for more than a small minority groups.

### **Free scientific publishing**

For the last few centuries the foundations of the established scientific process have been open publication of new scientific results, supplemented by subsequent critique and citing, and eventually leading to consolidation of the most relevant and scientifically proven results into a body of knowledge. This process has, in addition to the in-person defense of a thesis, been supported by the emergence of the peer reviewed scientific journal. Originally such journals were published by scientific societies as a service to

their members, but due to the rapid growth in the number of journals and papers, the publication process was during the latter half of the 20<sup>th</sup> century to a large extent taken over by commercial publishers. Due to the enormous growth in scientific literature a network of scientific libraries has evolved to help academics find and retrieve interesting items, supported by indexing services and inter-library loan procedures.

This process worked well until the mid nineties. The mix between publicly funded libraries on one hand and commercial publishers and indexing services on the other was optimal, given the technological border condition. The quick emergence of the Internet, where academics were actually forerunners as users, has radically changed the situation. At the same time there has been trend of steadily rising subscription prices ("the serials crisis") and mergers of publishers. Today one publisher controls 20 % of the global market. As a reaction a new breed of publications has emerged, published by scientists motivated not by commercial interests, but by a wish to fulfill the original aims of the free scientific publishing model, now using the Internet to achieve instant, free and global access (for recent discussions see Nature 2001, Science 2001). The author of this paper, belongs to this category of idealists, and has since 1995 acted as the editor-in-chief of one such publication (the Electronic Journal of Information Technology in Construction).

This movement has until now lacked a label (such as open source) or a clear leader (i.e. Linus Torvalds) but has during the last year or so become increasingly visible. An interesting debate ensued about the publication of the results of the world-wide research concerning the sequencing of the Human Genome. The well-known journal Nature opted to publish the information for free on the web (Nature Genome 2001). It was felt that the results of genome research is the result of a long collective endeavor by scientists world-wide and belongs in the public domain.

Recently a group of scientists in the fields of molecular biology and biochemistry, including several Nobel laureates, wrote a manifesto threatening to start

boycotting (in their roles as authors, reviewers and subscribers) any journals in their field that refuse to allow the free posting of their papers on the web at the latest six months after initial publication (SciAm 2001, PubMed 2001).

Everybody can agree that the free availability of scientific publications in full text on the web would be ideal for scientists. Results from a recent study carried out by this author and his colleague (Björk and Turk 2000) clearly indicate that scientists prefer downloading papers from the web to walking over to a library and that web material that is freely available, free-of-charge is preferred to paid for or subscription based (Figure 1).

There is, however, a serious debate about the cost of scientific publishing on the web (For opposite opinions see Odlyzko 1998, Tenopir and King 2000). It is clearly in the interest of commercial publishers to claim that web publishing is almost as expensive as ordinary paper based publishing, in order to justify the increasingly expensive subscriptions. Advocates of free publishing cite case examples of successful endeavors (cf. Walker 1998) where costs have been markedly lower. The much publicized Los Alamos preprints server for physics papers (xxx 2001), has for instance achieved costs which are only small fraction of the figures quoted by commercial parties. An important point would be to include the whole life-cycle cost of a scientific paper in the analysis, in which case the costs born by the publisher are only in the order of 3-5 % (Björk and Turk 2000). Nevertheless publishers require full copyright for their services!

It is not only the publishing itself which is becoming a battleground between commercial interests and idealistic scientists. Since the emergence of data base technology in the 1960ies a number of commercial indexing services have emerged, which libraries subscribe to. Traditionally these have relied on manual and or highly structured input of items to be included, a costly and also selective (and thus discriminatory) process. Now scientists are building automated web search engines which use the same web crawler techniques as used by popular tools like Alta Vista

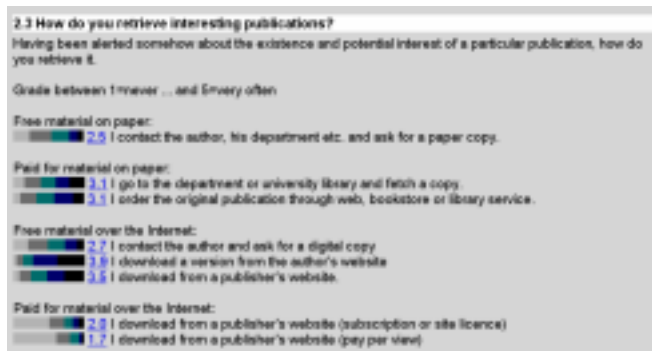


Figure 1. Some results from a recent web-survey of the reading and authoring habits of researchers in construction management and construction IT (Björk and Turk 2000). Material, which is available free-of-charge on the web, is the most popular means for accessing scientific publications.

etc, and which apply them to scientific publications published in formats such as pdf or postscript (Lawrence et al 1999, NECI 2001). If technically successful, such engines can be run at very low cost and thus be made available at no cost.

### OpenCourseWare

For university teachers the web is creating exciting new opportunities. Traditionally teachers have used the blackboard or hand-written overheads, using textbooks and other material as background input information. Students have had to go to the library to retrieve any reference material cited. Successful teachers have written textbooks for their fields and achieved a reputation and modest royalties by allowing some commercial publisher to publish them. Nowadays the technical possibilities are immensely better. Many of us use powerpoints or software demos in class using data projectors. The posting of teaching material on the web, as a supplement to classroom interaction, is becoming commonplace. Textbooks can be made available on the web, thus enabling the insertion of hyperlinks to the terabytes of material available out there for free. We often lend Powerpoint material and send it over the web to colleagues who adapt it to their own use. The key issue is that it is very easy to duplicate material, adapt it to local needs and to hyperlink it. In such a world it becomes increasingly difficult to uphold strict copyright, unless we create "fences" around our work, where only

selected students equipped with passwords or working inside the university website are allowed access.

In April of this year MIT webcast a press release (MIT 2001) which may have raised a few eyebrows. The essence of the press release is that over the next few years almost all the teaching material which is used in MIT courses will be made freely available on the Web for reuse by colleagues all around the world. The president of MIT, Charles M. Vest describes the project in the following way:

"OpenCourseWare looks counter-intuitive in a market driven world. It goes against the grain of current material values. But it is really consistent with what I believe is the best about MIT. It is innovative. It expresses our belief in the way education can be advanced – by constantly widening access to information and by inspiring others to participate." By the name chosen for this project MIT may in fact have coined the label of a new movement, which would apply many of the principles from open source and from the movement for free scientific publishing to the dissemination of academic teaching material. Their initiative comes at a very timely moment, when the emergence of fully virtual universities is imminent.

### Conclusions

The Internet has changed the world, and certainly the Academic World. The new technology offered by the web is superbly aligned with some of the basic

principles on which the advancement of science rests. In the middle ages a scientist wishing to expose a new finding to criticism and possible adoption by others nailed his thesis to a particular wall (a tradition still symbolically upheld for Ph.D. theses in some universities) in order for anybody who so wished to read it and criticize it. Nowadays making research publications freely available on the web fulfills the same function, but in a global academic community. Nevertheless the business process used for scientific and academic teaching publishing still functions according to the model shaped by the printing technology of the mid 20<sup>th</sup> century.

Increasingly scientists are, however, challenging the establishment by creating alternative free publication channels and by allowing anybody access to their teaching material. These developments are parallel with the emergence of the open source movement, which also is driven by similar motivations and as well as also being enabled by the Internet.

## References

- Björk, B.-C., Turk, Z.: 2000, How do researchers find and retrieve scientific publications – a case study of the impact of the Internet on the construction IT and construction management research communities, *Electronic Journal of Information Technology in Construction*, Vol 5 (2000) / 5. , <http://itcon.org/2000/5/>
- Himanen, P.: 2001, *The Hacker Ethic and the Spirit of the Information Age*, Random House, New York
- Lawrence, S.; Giles, C.; Bollacker, K.: 1999, Digital Libraries and Autonomous Citation Indexing, *IEEE Computer*, Vol. 32, Nr. 6, pp. 67-71, <http://www.neci.nec.com/~lawrence/papers/aci-computer98/aci-computer99.html>
- Ljungberg, L.: 2000, Open Source Movements as a Model for Organising, *European Journal of Information Systems*, 2000/9, pp. 208-216
- MIT: 2001, MIT to make nearly all course materials available free on the World Wide Web, MIT News, 4.4.2001, Massachusetts Institute of Technology, <http://web.mit.edu/newsoffice/nr/2001/ocw.html>
- Nature: 2001, Nature Webdebates, Future e-access to the primary literature, Nature, <http://www.nature.com/nature/debates/e-access/index.html>
- Nature Genome: 2001, Nature Genome Information Centre, <http://www.nature.com/genomics/>
- NECI: 2001, NECI Scientific Literature Digital Library, <http://citeseer.nj.nec.com/>
- Odlyzko, A.:1998, The Economics of Electronic Journals, *Journal of Electronic Publishing*, Vol. 4/1, <http://www.press.umich.edu/jep/04-01/odlyzko.html>
- PubMed: 2001, PubMed Central, <http://www.pubmedcentral.nih.gov/>
- Raymond, E.: 2001, *The Cathedral and the Bazaar*, O'Reilly, Sebastopol, California, <http://tuxedo.org/~esr/writings/cathedral-bazaar/>
- SciAm: 2001, Publish Free or Perish, *Scientific American*, 23.4.2001, <http://www.scientificamerican.com/explorations/2001/042301publish>
- Science: 2001, Science Discussion – A Public Archive for the Primary Scientific Literature, *Science Magazine*, <http://www.sciencemag.org/feature/data/hottopics/plsdebate.shtml>
- Tenopir, C. and King, D.: 2000, Towards Electronic Journals, Realities for Scientists, librarians, and Publishers, *Special Libraries Association*, Washington D.C.
- Walker, T. J, 1998 , Free Internet Access to Traditional Journals *American Scientist*, Vol 86, No. 5, Sept-Oct 1998, pp. 463-471 <http://www.sigmaxi.org/amsci/articles/98articles/walker.html>
- xxx:2001, The physics e-Print archive Los Alamos National Laboratory <http://xxx.lanl.gov/>