

# Chapter 4: Economic costs of toll access

Andrew Odlyzko

## Introduction

Many of the early discussions of scholarly electronic publishing, including precursors to what we now call Open Access, originated among librarians, and were stimulated by the 'serials crisis'. With prices of serials escalating far faster than inflation, libraries were forced into subscription cancellations and drastic cutbacks in book acquisitions. The hope was that something like Open Access would displace expensive publishers, or at least pressure them to moderate price increases. In practice, though, the revenues of the publishing industry have continued growing, and there are many cases of continuing high profit margins.

Historical precedents suggest that this should not be surprising, and can be expected to continue. Information is becoming ever more important in the economy, so total spending on it can be expected to grow. The invention of printing lowered the costs of information dissemination, and largely eliminated the job of scribes. But it also led to growth in total spending on books.

In recent times, library spending has kept growing, although not as fast as journal prices, and in most cases not as fast as university budgets. But something else has been happening as well, namely the rise of a large communication infrastructure. Its costs are higher than those of libraries, and it enables new types of scholarly communication (as well as improved access to library resources). It already provides essentially all the resources needed for Open Access.

The goal of this note is to sketch what the monetary costs of toll access are, and how they compare to other costs in a typical university setting. This note does not discuss the opportunity cost of toll access, namely the benefits to society of making scholarly information available through Open Access. That is covered by other chapters in this work (see Terry and Kiley, Harnad, Harnad and Shadbolt, Lynch). Suffice to note that these include the benefits to society that would have accrued as a result of: more effective research due to researchers' access to and building on relevant research, rather than research their institution can afford; better information discovery tools for researchers, and new research methods (for example based on text and data mining), that become possible when research texts and data are not hidden behind access control systems; and better research management, built on reliable data about which research is being used, by whom, when, and so on.

## Toll access costs and other costs

We do not have precise statistics on scholarly publishing. Still, it is estimated that the peer-reviewed literature grows by about 2.5 million papers a year, published in approximately 25,000 serials. Of these 2.5 million papers, somewhere on the order of 15% are Open Access (Hajjem *et al*, 2005).

The total revenues of scholarly publishers are estimated at around \$10 billion per year. This is a huge sum, but the world's annual spending on research and development (R&D) amounts to about \$1,000 billion dollars per year (National Science Foundation, 2006). Since the R&D enterprise relies heavily on the published literature, it is easy to see why there would be a reluctance to make drastic changes in something that costs just 1% of the total.

On the other hand, subscription fees collected by publishers are just one part of scholarly communication costs. Some rough estimates (see Odlyzko, 1995; Odlyzko, 1997c) suggest that the unpaid labor of editors and referees costs society about as much as publisher revenues. On the other hand, internal library costs (at least in high-wage industrialised countries) are at least twice as high, while the costs of authors' time (which are not captured explicitly in any accounting) spent writing the papers are at least five times as high. Higher still are the costs (also not accounted for in any formal budgets) of the time of readers of the papers. The frequently heard claims that only a handful of people read the typical paper are disproved by studies by Don King and others. It appears that there are usually several hundred readers per paper (see for example Tenopir and King, 2000). Therefore the total time readers devote to a typical paper is substantial. Hence, in a global sense, one could justify even a very high cost of peer review if it led to higher quality publications that were easier to locate and absorb. One could even argue for increased support for the kind of extensive editing and proofreading that some publishers used to provide. However, global optimisation is not something one can hope to achieve.

In scholarly publishing, optimisation is done on a local scale in a system with complicated feedbacks and indirect money flows. This system, in which most of the costs are born by libraries, while the scholars (who are the users as well as the creators of the material) have only an indirect say in what happens, is similar to the American medical industry (see for example Odlyzko, 1997c). As a result, there are many clear inefficiencies. However, as in the medical system, these inefficiencies are not easy to remove.

Partially as a result of the complicated scholarly publishing system, but even more as a result of the extreme inertia among scholars, change is extremely slow. Just about the only way to quickly alter established habits is through mandates (see Odlyzko, 1997b; and chapters by Swan, Harnad and Sale, this volume). A sense of the glacial rates of evolution that are typical of academia is given by the report "ARL Statistics 2003-04" (Kyrillidou and Young, 2005). Graph 1 in that publication shows that, for the large research libraries of North America that are members of ARL, the following changes took place:

|                         |       |
|-------------------------|-------|
| interlibrary borrowing: | +148% |
| student population:     | +18   |
| circulation:            | -1    |
| reference transactions: | -34   |

**Table 4.1: Changes in North American university research libraries**

These are noticeable changes, but they occurred over the 13-year period from 1991 to 2004! (Interlibrary loans, which are a small part of the typical research library budget, grew the fastest, to compensate for local library collections not being able to keep up with the growth of literature. But even there, the annual growth rate was only 7.2%.)

### An institutional perspective

Let us next consider a concrete example, that of the University of Minnesota. It is a large state university with the full range of undergraduate and professional schools, and several

campuses. It has approximately 65,000 students, some part-time, most full-time. During the 2003-04 academic year, total spending came to \$2,100 million, of which \$422 million was sponsored research, funded by outside sources. ARL statistics (Kyrillidou and Young, 2005) show that in that year, library expenditures were \$31.6 million. However, that figure applies only to the Minneapolis and St. Paul campus (although that is the largest campus, by far). Also, this figure, because of the ARL methodology, does not include some significant accounts, especially employee benefits. If we include all those other items, we find total library budget for the entire University of Minnesota of \$40.8 million, of which \$13.6 million was for acquisitions of serials, books, and (the most rapidly increasing part) electronic databases. Thus we find that libraries spent two thirds of their funds on internal costs (primarily pay for their staff) and absorbed just about 2% of the total university budget. (If we had proper accounting for all expenses, we would obtain a somewhat higher figure, since the libraries are not charged for depreciation or maintenance of their buildings, nor for usage of the common communications infrastructure.) A decade earlier, the libraries' share of the total budget had been closer to 2.5% (although the nominal library budget was far lower than in 2003-04).

Libraries are just one part of scholarly communication. At the University of Minnesota, total spending on information and communication technologies (ICT) during the 2003-04 academic year came to about \$150 million, or just about 7% of the entire university budget. (This includes about \$4 million of the libraries' spending.) Most of that spending, again over two thirds, is on people, all those systems and network administrators and webmasters (the vast majority employed by departments and other small units) who are required to keep the infrastructure operating.

ICT enable a multitude of activities, from high performance computing, through telephone services and Internet access, and on to payroll and facilities for students to sign up for classes.

Although ICT spending in Minnesota has been stable at just about 7% of the total university budget for the last few years (and the 7% figure is similar to those of many other large research universities in the U.S.), it has surely grown substantially compared to two decades ago, say. (Precise historical data is not available, unfortunately, since spending is decentralised and was not tracked until recently.) And much of that growth of ICT spending has been driven by the demands of scholars for better communication, some related to teaching and departmental affairs, but much for their research. That is what has enabled the rapid growth of new types of scholarly communication that is described in Odlyzko (2002), in which researchers download papers from online archives and search using Google instead of asking librarians for help. (Note the substantial decline in reference transactions in the ARL data in Table 4.1.) With increasing volume of information available on the web, ease of access is paramount, especially since satisfactory substitutes are often available even for the most esoteric results.

The point of discussing a particular institution is that data for ICT spending for it is available. What we find is that this spending is far larger than the library budget, and over 10 times as large as the book and serial acquisition costs of the library. So costs of supporting Open Access, which even under the most conservative assumptions are far lower than the costs of the library, are a very small fraction of ICT costs. Almost the entire required infrastructure is already in place, and it requires relatively small steps to make Open Access a reality, either by mandating self-archiving, or by funding publication charges made by Open Access publishers.

## Conclusions

The explicit costs of toll access, namely the costs of journals and books, are what attracts most attention. But those costs are only a small piece of the intricate jigsaw of various expenditures that support scholarly communication in its various forms. Internal library costs are in large research institutions in industrialised countries about twice as high as the costs of serials and books. And total library costs are dwarfed by the costs of the information and communication technologies infrastructures. Many of those expenditures have multiple purposes, and all have their constituencies. Given the slow pace of change among scholars, we should not expect any major changes in funding flows. However, Open Access does not require much new funding, although the 'gold' route might require a 'level playing field' to be established in terms of public funding, as between Open Access and subscription models. The infrastructure for it is basically already in place, built for other purposes, and with enough capacity to handle the extra load that Open Access would impose.