

Learned Publishing (2002)15, 7–19

The rapid evolution of scholarly communication

Introduction

Traditional journals and libraries have been vital components of scholarly communication. They are evolving, but slowly. The reasons for this are discussed briefly in Section 1 and, in more detail, in Odlyzko.¹ The danger is that they might be rapidly losing their value, and could become irrelevant.

At first sight, there seems little cause for concern. Print journal subscriptions are declining, but gradually. One often hears of attrition in subscriptions of 3–5% per year. (For example, the American Physical Society, with high-quality and relatively inexpensive journals, has seen a steady decrease of about 3% per year.²) At those rates, it takes between 14 and 24 years to lose half the circulation. On internet time, that is almost an eternity. Preprints in most areas are still a small fraction of what gets published. Also, library usage is sometimes reported as declining, but again at modest rates. (For circulation figures for major research libraries in the US, see ARL.³) Yet these are not reasons for complacency. Why should there be any declines at all? Ours is an 'Information Age'; the number of people getting college and postgraduate education is growing rapidly, spending on R&D and implementation of new technologies is skyrocketing. Why should established journal subscriptions be dropping, and why should many of the recent specialized journals be regarded as successes if they reach a circulation of 300? Why should many research monographs be printed in runs smaller than the roughly 500 copies of the first edition of Copernicus's *De revolutionibus orbium coelestium* of 1543?

My conclusion is that the current scholarly information system is badly flawed, and that it does not provide the services that are required. This paper presents evidence that there is indeed a growing demand for

Andrew Odlyzko
AT&T Labs – Research

© ALPSP 2002

ABSTRACT: *Traditional journals, even those available electronically, are changing slowly. However, there is rapid evolution in scholarly communication. Usage is moving to electronic formats. In some areas, it appears that electronic versions of papers are being read about as often as the printed journal versions. Although there are serious difficulties in comparing figures from different media, the growth rates in usage of electronic scholarly information are sufficiently high that if they continue for a few years, there will be no doubt that print versions will be eclipsed. Further, much of the electronic information that is accessed is outside the formal scholarly publication process. There is also vigorous growth in forms of electronic communication that take advantage of the unique capabilities of the web, and which simply do not fit into the traditional journal publishing format. This paper presents some statistics on usage of print and electronic information. It also discusses some preliminary evidence about the changing patterns of usage. It appears that much of the online usage comes from new readers (esoteric research papers assigned in undergraduate classes, for example) and often from places that do not have access to print journals. Also, the reactions to even slight barriers to usage suggest that even high-quality scholarly papers are not irreplaceable. Readers are faced with a 'river of knowledge' that allows them to select among a multitude of sources, and to find near substitutes when necessary. To stay relevant, scholars, publishers and librarians will have to make even greater efforts to make their material easily accessible.*



Andrew Odlyzko

high-quality scholarly information, and that it can only be satisfied through easy availability on the web.

Tenopir *et al.*'s important study⁴ does show that electronic resources are playing an increasing role, but current usage by established scholars is dominated by traditional media. However, it is important to look at growth rates rather than absolute numbers. In an early-1999 discussion in a librarians' mailing list, somebody pointed out that, in 1998, only 20% of the astronomy articles were submitted to Ginsparg's xxx paper archive (now called the arXiv: www.arxiv.org). An immediate rejoinder from another participant was that, while this was true, the corresponding percentage was around 7% in 1995. It is growth rates that tell us what is in our future.

This paper is only a brief attempt at finding patterns in usage of online information. What we need are careful studies, such as have been carried out for print media. (An excellent and up-to-date survey of those is presented in Tenopir and King;⁵ see also a brief summary in King and Tenopir.⁶) At the moment, we do not even have much data about usage patterns online. This is especially regrettable since these patterns appear to be in the midst of substantial changes. Although the web in principle makes it possible to provide extremely detailed information about usage (and this has led to numerous privacy concerns), in practice there is little data collection and analysis, especially in scholarly publishing. Even when data are collected, they are seldom released. Thus one purpose in writing the initial draft of this paper was to stimulate further collection and dissemination of usage data. The main purpose, though, was to look for patterns, even with the scanty data that I was able to collect, to provide a starting point for further research.

Fortunately, many new studies of electronic resources have appeared very recently. Some of the notable ones will be referenced later.^{4,7-11} In general, they do support most of the theses of this paper.

Some of the early studies of electronic usage, such as that in Lenares's interesting paper,¹² concentrated on faculty at leading research institutions. Change might be ex-

pected to be slow in such places. Although they usually have the resources to be pioneers, they have little incentive for it, since they do possess good libraries. The evidence to be presented later shows that the current system neglects the needs of growing ranks of scholars who are not at such institutions. Thus it is better to concentrate on usage of information that is freely available over the internet.

Later sections discuss in detail some statistics as well as some qualitative measures of usage of online resources. Here are some tentative conclusions:

1. Usage of online scholarly material is growing rapidly, and in some cases already appears to surpass the usage one could expect to see in traditional print journals. Much of the online usage appears to come from new readers (esoteric research papers assigned in undergraduate classes, for example) and often from places that do not have access to print journals. Evidence can be found in Guthrie⁸ and Luther,¹¹ for example, and in later sections of this paper.
2. We can expect the growth of online material to accelerate, especially as the information about usage patterns becomes widely known. Until recently, scholars did not have much of an incentive for putting their works on the web, as this did not create many new readers. While we can expect that snobbery will retard this step ('I can reach the dozen top experts in my field by publishing in *Physical Review Letters*, or by sending them my preprint directly, why do I care about the great unwashed?'), the attraction of a much greater audience on the web, and the danger that anything not on the web will be neglected, are likely to become major spurs to scholars' migration of their works online. For example, a recent study¹⁰ shows that papers in computer science that are freely available online are cited much more frequently than others. (Anderson *et al.*'s paper⁷ might appear to suggest the opposite, since free online availability there was associated with lower citation frequency. However, that result is probably anomalous, in that the

*We can expect
the growth of
online material
to accelerate*

freely available online-only articles in the journal under study were apparently perceived widely, even if incorrectly, as of inferior quality.)

3. The need for traditional peer review is overrated. Odlyzko¹³ discusses extensively the inadequacy of conventional peer review, and how much more useful forms were likely to evolve on the internet. (That paper was written before the ascendancy of the web.) While open review and comments on published papers have been slow to take hold, something else is going on. People are coming to my web page in large numbers looking for specific papers. While in almost all cases I do not know what brings them there, it is pretty clear that they are getting pointers to the material from a variety of sources, such as bibliographies and references on other home pages. It is a form of peer review, and it brings many readers even for papers published in obscure and un-refereed places.
4. Concerns about information overload and chaos on the internet are exaggerated. While better organization of the material would surely be desirable, people are finding their way to the serious information sources in growing numbers as is.
5. Ease of access and ease of use are paramount. Material on the web is growing, and scholars, like the commercial content producers, are engaged in a 'war for the eyeballs'. Readers will settle for inferior forms of papers if those are the ones that can be reached easily.
6. Novel forms of scholarly communication are evolving that are outside the boundaries of traditional journals.

These conclusions and predictions are supported by data in the rest of this paper. It does appear that while journals are not changing fast, scholarly communication as a whole is evolving rapidly.

1. Rates of technological change

The conventional notion of 'internet time', in which technological change is accelerated tremendously, is a myth. Rapid change does

occur occasionally, and the adoption of web browsers is frequently cited as an example. Less than 18 months after the release of the first preliminary version of the Mosaic browser, web transmissions constituted more than half of internet traffic. However, this was a singular exception. Cell phones, faxes and ATM machines took much longer to spread. Even on the internet, new systems are usually adopted much more slowly. How come IPv6 is still basically invisible? Why is HTTP1.1 spreading so slowly? How about TeX and its various dialects (which go back more than two decades)? Email itself took a while to diffuse, even at universities. The internet has changed much, but it has not made for a dramatic increase in the pace at which new technologies diffuse. A typical timescale for significant changes is still on the order of a decade. This was noted a long time ago:

A modern maxim says: 'People tend to overestimate what can be done in one year and to underestimate what can be done in five or ten years'. (Licklider,¹⁴ footnote on p. 17)

Further discussion of rates of change is available in Odlyzko,¹ which presents many examples (such as music CDs, ATM machines, credit cards and cell phones) supporting the thesis that consumer adoption of new technologies is slow. (For more evidence, see also Klopfenstein¹⁵ and the references therein.) Thus we should not be surprised if electronic scholarly communication does not turn on a dime.

The rare rapid adoptions of new technologies (aside from unusual situations such as that of the web) appear to be associated with the presence of forcing agents that can compel rapid change.¹ On the other hand, sociological changes tend to be very slow, taking a generation or two.

Aside from simply observing that, historically, new technologies have been taking of the order of a decade to be widely adopted, one can also build quantitative models that explain this time scale. Suppose we have two competing or nearly competing services, A and B. Suppose usage of A is static, while that of B increases at 50–100% per year, which in the business world

Concerns about information overload and chaos on the internet are exaggerated

definitely qualifies as spectacular growth. One can easily imagine that *B* might not be noticed until its usage reaches 1% of the established service *A*. From the moment that 1% threshold is reached, even at growth rates of 50–100 % per year, it will take between 7 and 14 years before *B* reaches parity with *A*.

Usage of electronic forms of scholarly information has typically been growing at 50–100% per year, as is shown in various tables in this paper. On the other hand, print usage has shown little change, as far as anyone can tell. Thus the simple model above tells us that a decade is about the length of time we should expect for new modes of electronic communication to become dominant, if current growth rates continue.

2. Disruptive technologies

Clayton Christensen's book has become a modern classic.¹⁶ It helps explain the failure of successful organizations, such as 'Encyclopaedia Britannica', to adopt new technologies. The example of the *Britannica*,^{13,17} is very instructive. It was and remains the most scholarly of the English-language encyclopaedias. However, it could not cope with the challenges posed first by inexpensive CD-ROM encyclopaedias, and more recently by the web.

What Christensen calls disruptive technologies tend to have three important characteristics:

1. initially underperform established products;
2. enable new applications for new customers;
3. performance improves rapidly.

Electronic publishing has these characteristics. Little material was available initially, screen resolution was poor, printers were not widely available and expensive, and so on. However, online material was easy to locate and access, and could provide novel features, such as the constant updating of the genome database. Moreover, costs, quality and availability have all been improving rapidly. (It should be noted

that print also had these characteristics when compared with hand-written manuscripts.^{18,19}) That is why direct comparisons of traditional journals or libraries with electronic collections are not directly relevant. For example, Stevens-Rayburn and Bouton's 1998 paper²⁰ is effective in demonstrating that the web at that time could not substitute for a regular library. It still can't, even in 2001. However, that is not the relevant question.

The mainframe was not dethroned by the PC directly. The PC could not do most of the tasks of the big machines in areas such as payroll processing. The computing power of the mainframes sold each year is still increasing, and has been increasing all along, even when IBM was going through its traumatic downsizing in the early 1990s. It is just that the PC market has been growing much faster, and the mainframe has been consigned to a small niche, and the revenues from that niche have been declining. I think this is a useful analogy to keep in mind. Traditional journals and libraries are still playing a vital role, but, 'journals are not where the interesting action is'.¹ The real issue is that, 'in this new electronic age, if it isn't online, for many purposes it might as well not exist'.²⁰ Further, even if it is online, it might not matter if it is not easy to access or is not timely.

3. Effects of barriers to use

Even small barriers to access reduce usage significantly. There are some wonderful statistics collected by Don King and his collaborators²¹ (and Fig. 9.4 on p. 202 of Lesk,²² reproduced from Griffiths and King²¹) which show that as the physical distance to a library increases, usage decreases dramatically. A recent statistical tidbit of a similar nature that I have collected is the reaction of the mathematicians at Penn State when all journal issues published before 1973 had to be sent to offsite storage because of space limitations. This move was widely disliked, even though any volume can be obtained within one day. The interesting thing is that the mathematical research community of about 200 faculty, visitors and graduate students asks

*Even small
barriers to
access reduce
usage
significantly*

for only about 850 items to be recalled from storage per year. That is just over four items per person per year. It seems likely (based on extrapolations from circulation figures for bound journals that are immediately available on shelves) that usage of this material was much higher when it was easily accessible in the library in their building.

When subscriptions to journals are cancelled, articles from those journals are obtained through interlibrary loans or document delivery services. Some libraries (Louisiana State University's perhaps most prominent among them) have consciously decided to replace journal subscriptions with document delivery, after making a calculation of how much the journals cost per article read. While I do not have comprehensive statistics, my impression is that such moves save more than preliminary computations suggest. The dirty little secret behind this phenomenon is that usage of document delivery services is lower than that of journals available right on the spot. Having to fill out a request form and wait a day or a week reduces demand.

Librarians have known for a long time that ease of use is crucial. They experienced this with card catalogues, where materials whose catalogue entries were left in the paper card catalogues were not being used. Thus the current shift towards online usage had been anticipated.

. . . there's a sense in which the journal articles prior to the inception of that electronic abstracting and indexing database may as well not exist, because they are so difficult to find. Now that we are starting to see, in libraries, full-text showing up online, I think we are very shortly going to cross a sort of critical mass boundary where those publications that are not instantly available in full-text will become kind of second-rate in a sense, not because their quality is low, but just because people will prefer the accessibility of things they can get right away. (Clifford Lynch, 1997, quoted in Stevens-Rayburn and Bouton²⁰)

Today, we have evidence than Lynch was correct. Note that *Encyclopaedia Britannica*

has been a victim of this trend. Being the best did not protect it.

The shift to online usage is exposing many of the limitations of the traditional system. Research libraries are wonderful institutions. They provide the best service that is possible with print technology. However, in today's environment, that is not enough. Most printed scholarly papers are available typically in something like 1,000 research libraries. Those libraries are accessible to a decreasing fraction of the growing population of educated people who need them. Further, even for those scholars fortunate to be at an institution with a good library, the sizes of the collections are making material harder to access. Hours of availability are limited. Also, studies have shown that even when a book that is searched for is in a given library's collection, in about 40% of the cases it cannot be found when needed (see endnote 10 to Chapter 2 of Buckland²³ for references).

The basic problem, of course, is that it is impossible in the print world to make everything easily accessible even in the best library in the world. Space constraints mean that some material will be far from the user. In practice, most libraries can store only a tiny fraction of the material that might be of interest to their patrons. While they have been careful about selecting what seemed to be most relevant, experience shows that when easy electronic access is provided to large bodies of material not normally available in the library, there is demand for it.^{11,24} That is a major factor propelling the move towards bundling of electronic journal offerings and consortium pricing.¹⁷

The easy access to online resources is leading to increasing usage, as will be discussed later, and is also documented elsewhere.^{7,8,11,25} But not all online accesses are equal. Many scholars (including myself) use Amazon.com's search page as a first choice in doing bibliographic searches for recent books, since it is more user-friendly than the electronic catalogues of the Library of Congress, say. 'Both Academic Press and the American Institute of Physics (AIP) noted that they experienced surges in usage after they introduced new platforms that simplified navigation and access.'¹¹

The shift to online usage is exposing many of the limitations of the traditional system

Ease of use has an important bearing on pricing. Odlyzko¹³ predicted that pay-per-view was probably doomed to fail in scholarly publishing, because of its deterrent effect on usage. (More evidence and arguments supporting that prediction have been developed.^{26,27}) Publishers have now (after experiments with PEAK and other pricing models) moved to this view as well. For example,

[Elsevier's] goal is to give people access to as much information as possible on a flat fee/unlimited use basis. [Elsevier's] experience has been that as soon as the usage is metered on a per-article basis, there is an inhibition on use or a concern about exceeding some budget allocation.⁹

Authors like to think of their articles as precious resources that are absolutely unique

Similarly, 'Philosophically, Academic Press is opposed to a business model in which charges increase with use because it discourages use.'¹¹

Easy access implies not only greater use, but also changing patterns of use. For example, a recent news story²⁸ discussed how the internet is altering the doctor-patient relationship. The example that opens that story is of a lady who is reluctantly told by the doctor she might have lupus, and leaves the clinic terrified of what this might be. She then proceeds to obtain information about this disease from the internet. When she returns to see a physician (a different, more pleasant one), she is well informed and prepared to question the diagnosis and possible treatment. What is remarkable about this story is that the basic approach of this patient was feasible before the arrival of the web. She could have gone to her local library, where the reference librarians would have been delighted to point her to many excellent print sources of medical information. However, few people availed themselves of such opportunities before. Now, with the easy availability of the web, we see a different story.

The arguments about effects of barriers to access and of lowering such barriers suggest that scholarly communication will undergo substantial changes. We should expect to see greater use of online material. We should also see much greater use of it by people outside the narrow disciplinary areas that

produce it. Much of this use will come from outside the traditional academic and research institutions, but a considerable fraction is likely to come from other departments within an institution. Further, the increasing volume of material, as well as the decreasing role of traditional peer review, are likely to lead to greater demand for survey and handbook material. With lower barriers to interactions and access to specialized literature, we should also see more interdisciplinary work.

4. Scholarly information as a commodity

Authors like to think of their articles as precious resources that are absolutely unique and for which no substitutes can be found. Yet a more accurate picture is that any one article is just one item in a river of knowledge, and that this river is growing. Substitutes exist for almost everything. Some people interested in Fermat's Last Theorem will want, for historical or other reasons, to see Andrew Wiles's original paper.²⁹ Many others will be happy with a reference to where and when that paper was published, and others will be satisfied with various popular accounts of the proof. Even those interested in the technical details will often be satisfied with (and often be better served by) other presentations, such as that in the Darmon, Diamond and Taylor account of the proof.³⁰ Thinking about a river of knowledge instead of a collection of unique and irreplaceable nuggets helps explain why scholars manage to function even with a badly flawed information system. Even though in 40% of the cases a desired book cannot be retrieved, usually some other book covering the same topic can be found. Spending on libraries by research universities is correlated most strongly with the total budgets, and very weakly with the quality. Harvard spends about \$70 million per year on its libraries, compared with \$25 million spent by Princeton. Yet would anyone claim that a Harvard education or scholarly output is almost three times as good as that of Princeton?

The internet is reducing the costs of production and distribution of information.

Table 1. Library of Congress electronic resource usage statistics (For each month, shows total volume of material sent out that month, in gigabytes, and the number of requests.)

Month	GB	Requests (millions)
Feb. 1995	14.0	1.1
Feb. 1996	31.2	3.9
Feb. 1997	109.4	15.1
Feb. 1998	282.0	36.0
Feb. 1999	535.0	48.6
Feb. 2000	741.1	61.3
Feb. 2001	1202.6	86.7

As a result, there is a flood of material. Much is of low quality, but a substantial fraction is very good. The question is, are scholars using it? Before looking at that question, let us consider usage of print material.

5. Usage of print journals

We are fortunate to have an excellent recent survey of usage of print journals in the book by Carol Tenopir and Don King.⁵ (A summary is presented in King and Tenopir.⁶) It shows that a typical technical paper is read (defined as not necessarily reading it carefully, but going beyond just glancing at the title and abstract) between 500 and 1500 times. These readings average about one hour in length, and in about half the cases represent the reader's first encounter with an article.

The estimate of 500–1500 readings per article is a much higher number than some earlier studies had come up with. It is based on careful studies, though. Those studies have biases that may raise the reading estimates above the true value. For example, they are based on self-reporting by technical professionals, who may overestimate their readings. (People usually report eating less chocolate and more salad than they actually consume.) Further, those figures include articles in technical journals with large circulations (such as *Science*, *Nature* and *IEEE Spectrum*) that are not typical of library holdings. If one considers library usage studies, such as those that have been carried out at the University of Wisconsin

Table 2. AT&T Labs – Research external web server statistics (Excludes most crawler activity. Number of hosts for Jan. 1997 is an estimate.)

Month	Requests	Hosts
Jan. 1997	542,644	17,886
Jan. 1998	754,477	35,943
Jan. 1999	1,204,664	67,191
Jan. 2000	1,843,319	100,077
Jan. 2001	4,190,362	178,923

in Madison (www.wisc.edu/wendt/journals/costben.html), one comes up with somewhat lower estimates for the number of readings per paper. Still, the basic conclusion that a typical technical paper is read several hundred times appears valid.

The studies reported in Stevens-Rayburn and Bouton²⁰ also show that, in the print world, articles are usually read mostly in the first half a year after publication. Afterwards, usage drops off rapidly.

6. Growth in usage of electronic information

The internet is growing rapidly. Typical growth rates, whether of bytes of traffic on backbones, or of hosts, are of the order of 100% per year.^{31,32} When one looks at usage of scholarly information online, typical growth rates are in the 50–100% range. For example, Table 1 shows the utilization of the online resources of the Library of Congress. Growth was about 100% per year for four years, and then, in 1999, it slowed down to 38%. It then increased to 62% in 2000. (These growth rates are for bytes transmitted.) Table 2 shows downloads from the AT&T Labs – Research website, www.research.att.com/, which contains a variety of papers, software, data and other technical information. The growth rate there has been around 50% per year for several years, but between 2000 and 2001, it jumped to over 120%.

It is hard to measure online activity accurately. The earliest and still widely used measure is that of 'hits', or requests for a file. Unfortunately, with the growth of complicated pages, that measure is harder to evaluate. When possible, I prefer to look at full article downloads. (That will be the

It is hard to measure online activity accurately

Table 3. Visits to Leslie Lamport's Temporal Logic of Actions web page (approximate counts)

Year	Visits	Hosts
1996	18,800	5,300
1997	19,000	5,600
1998	18,400	5,300
1999	31,100	8,000
2000	33,500	8,000

measure discussed in Sections 8 and 9 below.) Finally, as a conservative measure, one can look at the number of hosts (unique IP addresses) that requested information from a server. Even then, there are considerable uncertainties. The same person may send requests from several hosts. On the other hand, common employment of proxies and caches means that many people may hide behind a single host address, and a single download may lead to multiple users obtaining copies (as happens when papers are forwarded via email as well).

In addition to the uncertainties in interpreting the activity seen at a server, it is hard to compare data from different servers. Logs are set to record different things, and some web pages are much more complicated than others that have the same or equivalent content. Thus comparing different measures of online activity is of necessity like comparing apples, oranges, pears, bananas and onions. Some of the difficulties of such comparisons can be avoided by concentrating on rates of growth. If online information access is growing much faster than usage of print material, it will eventually dominate.

Some measures of electronic information usage are showing signs of decreasing growth, or even stability. For example, Table 3 shows utilization of Leslie Lamport's page devoted to material about a logic for specifying and reasoning about concurrent and reactive systems (www.research.digital.com/SRC/tla/). Usage had been pretty stable in 1996–1998. When I corresponded with him about this in 1999, he thought usage had reached a steady state, with the entire community interested in this esoteric technical subject already accessing the page as much as they would ever need to do. However, the final count for 1999 showed a substantial

increase. The next few sections discuss data about several online information sources that are freely available on the internet.

7. Electronic journals and other organized databases

Some reports are already available on the dramatic increase in usage of scholarly information that is easily available. Traditionally, theses and dissertations have been practically invisible, and were used primarily within the institution where they were written, and even there, they were not accessed frequently. Free access to digital versions is now leading to an upsurge in usage.³³

In the remainder of this section, as a first approximation, I will equate a full article download with a reading as measured by Don King and his collaborators.

The entire American Mathematical Society e-math system was running at about 1.2 million 'hits' per month in early 1999. The Ginsparg archive (arXiv) at Los Alamos was getting about 2 million hits per month. The netlib system of Jack Dongarra and Eric Grosse was at about 2.5 million hits per month.

By the end of 1999, usage of JSTOR was several million a month, whether one counts hits or full article downloads, and was growing at over 100% per year.⁸

The Brazilian SciELO (Scientific Electronic Library Online: www.scielo.br/scielo/scielo-an.htm) project started out in early 1998. It appears to be still going through the initial period of explosive growth, with the number of pages transmitted growing from 4,943 in Jan. 1999 to 63,695 a year later. (67,143 hosts requested pages in 1999, so it was not just a small group of users who were involved.) It is too early to tell about how fast it will continue to grow, but it seems worth listing this project to show that even the less industrialized countries are participating in making literature freely available.

Paul Ginsparg's arXiv had about 100,000 papers in early 1999, and was running at a rate of about 7 million full article downloads per year. Thus on average each article was downloaded about 70 times per year. Further, these download statistics were just for the main Los Alamos server. If we

Some measures of electronic information usage are showing signs of decreasing growth

assume that the more than a dozen mirrors collectively see as much activity as the main server, then we get a download rate of about 140 times per year per article. This is misleading, though, since it mixes old and new papers, which have different utilization patterns. If we look at download activity for arXiv articles as a function of time, we find (extrapolating very freely from data kindly supplied by Paul Ginsparg) that on average an article gets downloaded around 150 times within one year of its submission, and then 20–30 times a year in subsequent years. (In particular, even articles submitted around 1991 get downloaded that often. This is different from the pattern observed by King and other for printed journal articles. Those are read primarily in the six months after publication, and then the frequency with which they are accessed decreases.) Since this again covers just the main server, we probably should again multiply these numbers by two to get total activity. If we do that, we get into the range of readings per article that established journals experience.

The *Electronic Journal of Combinatorics* had published about 200 articles by early 1999, and had about 30,000 full article downloads from its main site each year. That is an average of 150 downloads per article. Multiplying that by two to account for the many mirror sites again gets us to about 300 downloads per article per year. (Data about distribution of downloads with time are not available.)

The general impression from the statistics quoted above is that articles in electronic archives and electronic journals may not yet be read as frequently as printed journal articles, but are getting close. On the other hand, some online sources appear to be used much more frequently than they would be in print.

8. *First Monday*

Additional evidence that online access changes scholars' reading patterns is provided by *First Monday*, 'the peer-reviewed journal of the internet' (<http://firstmonday.org>). Issues are made freely available on the first Monday of each month. *First Monday* started publication in May 1996. There are

about 3,600 subscribers to the email notification service.

First Monday has provided me with access to the logs of their US web server from Jan. 1999 to Feb. 2000. (The data for Jan. 1999 are incomplete, since the main server was then in transition from Denmark to the US.) This is not sufficient for a careful statistical study, but some interesting patterns can be discerned in the data.

Over this period, the number of full paper downloads has grown from a range of 50,000–60,000 per month in early 1999, to between 110,000 and 120,000 per month in early 2000. Distinct hosts requesting articles have increased from 12–15,000 to over 20,000 each month. Thus the growth rate has been close to the 100% that we have seen occurs frequently on the internet. Since there are only 3,600 subscribers, this suggests many others learn of the material through word of mouth, email or other methods.

In a typical month, the largest number of downloads is to articles from that month's issue. In subsequent months, accesses to that issue drop in a pattern similar to that found by Don King in his studies of print journals. Half a year later, downloads are usually down to a quarter or even a sixth of the first month's rate. At that stage, though, the story changes. Whereas for print journals, usage continues to decrease with time, for *First Monday* it appears to increase. For example, there were 9,064 full article downloads from all the 1997 issues in Feb. 1999, and 19,378 in Feb. 2000. Thus accesses to the 1997 issues kept pace with the general growth of usage. Of the articles that were most frequently downloaded in 1999, 6 of the top 10 were published in previous years! This supports the thesis that easy online access leads to much wider usage of older materials.

9. My personal web page

Table 2 shows the statistics of the AT&T Labs – Research external web server, www.research.att.com. My personal web page, www.research.att.com/~amo, has also seen very rapid growth in usage. However, it is hard to discuss it meaningfully in a short

articles in electronic archives and electronic journals may not yet be read as frequently as printed journal articles

space, since most of the growth came from new papers in new areas. (The most frequently accessed papers on my home page are those on data networks. Then come papers on electronic publishing and electronic commerce. Those are followed by papers on cryptography, and the esoteric mathematics papers are last in frequency of access.) Instead, I will discuss some impressions from the usage patterns that I observe.

*People are
guided to web
pages by a
variety of cues*

During Jan. 2000, there were 10,360 'hits' from 1,808 hosts on my home page, excluding .gif files, and hits from obvious crawlers. Most of these 1,808 hosts only looked at various index files. If we exclude those, as well as the ones that downloaded only my CV or only abstracts of papers, we are left with 656 hosts that downloaded 1,198 full copies of articles. Of those 656 hosts, 494 downloaded just a single paper. Many of those 494 requested a specific URL for an article (as opposed to looking at the home page for pointers) and then disappeared. Thus on average the people who visited my home page seemed to know what they were looking for, got it and moved on.

Visitors to my web page were remarkably quiet in the face of some obvious faults. Many of the papers posted on that page, especially old ones, are incomplete, in that they are early versions, and usually do not have figures that are present in the printed versions. Still, that occasions few complaints. As one example, about a year ago, a posting to a number theory mailing list resulted in 152 downloads of a paper in the space of less than two weeks. However, only one person complained about the lack of figures in the web version, even though they are very helpful in visualizing the behaviour shown in the paper.

Another anecdotal piece of evidence of what happens on the web: Several times I have encountered people who told me that they were really glad to meet me, as they had read my papers in one area or another, and benefited from them. Moreover, conversation showed that they indeed were familiar with the papers in question. However, they also told me that they had lost the URL, and would I please remind them where my home page was? Now it is pretty

easy to find my home page on the web (my name is not a particularly common one), yet they obviously did not find it necessary to bother doing it. This, as well as the situation in the paragraph above, suggests a world of plenty. People are guided to web pages by a variety of cues, get whatever they can from those pages, and move on to other things. It is not a world of a few precious treasures that have no substitutes.

The importance of making material easily available was demonstrated in a very graphic form when I made PDF versions of my technical papers available in April 1998. There was an immediate jump in the rate of downloads. (Prior to that, mathematical papers were available only in Postscript and TeX formats, the ones on electronic publishing and related topics in Postscript and straight text.) Most PC owners do not have easy access to tools for reading Postscript papers, and were apparently bypassing the available material that required extra effort from them for reading. This is similar to observations of Academic Press and the American Institute of Physics¹¹ that better interfaces lead to higher usage.

The temporal pattern of article usage on my web page shows the behaviour that was already noted for arXiv and for *First Monday*. (As a matter of chronology, it was the observation about access patterns to my papers that led me to investigate the question in other online databases several years ago.) After an initial period, frequency of access does not vary with age of article, and stays pretty constant with time (after discounting for general growth in usage).

There is more evidence that easy online access leads to changes in usage patterns. For example, downloads from my home page go to a variety of sources all over the world. Some are leading to email correspondence from exotic places like Pakistan, the Philippines or Mexico. This is not surprising in itself, since those countries do have technically educated populations that are growing. What is interesting is that this correspondence predominantly refers to my papers that had been downloaded electronically (and sometimes requests copies of older papers that are not available in digital form, and which the requesters had learned

about from my home page). This does suggest strongly that easy availability is stimulating interest from a much wider audience. This conclusion is also supported by similar observations concerning correspondence with people in industrialized countries. Many come from outside the universities or large research institutions that have good libraries. They would be unlikely to read my papers in print. The referrer field on requests shows in a small fraction of cases where the requester found the URL. In many cases, such requests come from reading lists in college or graduate courses.

As a final note, there are often spikes in usage when one of my papers is mentioned in some newsletter or discussion group. For example, Bruce Schneier publishes *CRYPTO-GRAM*, a monthly email newsletter on cryptography and computer security. It has a circulation of about 20,000. In early Aug. 1999, it mentioned a recent preprint of mine (which I had not advertised much, and which has since appeared in a regular print journal). Over the next two weeks over a thousand copies were downloaded. I am convinced that this is a higher figure than the number of times the printed version will be read.

The *CRYPTO-GRAM* example as well as those of other visits to my home page suggest that informal versions of peer review are in operation. A recommendation from someone, or a reference in a paper that the reader trusts all serve to validate even unpublished preprints. Scholars pursue a variety of cues in selecting what material to access.

10. New forms of scholarly communication

A popular destination on the AT&T Labs – Research web server is my colleague Neil Sloane's 'On-Line Encyclopedia of Integer Sequences', accessible from his home page, at www.research.att.com/~njas/. In Jan. 2000, it attracted more than 6% of all the hits to the AT&T Labs – Research site (see Table 4). This 'encyclopedia' is a novel combination of a database, software and now also a new online journal. The integer sequence project enables people to find out

Table 4. Requests to Neil Sloane's sequence server (Hosts for 1997 estimated.)

Month	Requests	Hosts
Jan. 1997	6,646	550
Jan. 1998	33,508	2,294
Jan. 1999	58,655	3,996
Jan. 2000	135,843	7,851
Jan. 2001	222,795	11,105

what the next element is in a sequence such as

0, 1, 8, 78, 944, 13800, 237432, . . .

This might seem like recreational mathematics, but it is very serious, as many research papers acknowledge the assistance of Sloane's database (or, in earlier times, his books on this subject). It serves to tie mathematicians, computer scientists, physicists, chemists and engineers together, and stimulate further research. (For an account of the project, see Sloane's recent paper.³⁴) It represents a novel form of communication that could not be captured in print form.

Another popular site that is also a locus of mathematical activity is Steve Finch's 'Favorite Mathematical Constants' page (www.mathsoft.com/asolve/constant/constant.html). It is also showing rapid growth in usage (although one that is harder to quantify, since monitoring software was changed less than a year ago, so comparisons are harder to make). Just as with Sloane's integer sequence page, it is becoming a form of 'portal' to mathematics, one that does not fit easily into traditional publications models.

11. Conclusions and predictions

Many discussions of the future of scholarly publishing have been dominated by economic considerations. Digitization has often been seen as a solution to the 'library crisis', which forces libraries to cut down on subscriptions. So far there has been little effect in this area, as pricing trends have not changed much.¹⁷ In the long run it has been clear that print would eventually become irrelevant, aside from any economic pressures, as it is simply too inflexible. Gutenberg's invention imprisoned scholarly

Digitization has often been seen as a solution to the 'library crisis'

*Ease of access
is likely to
promote the
natural
evolution of
scholarly work*

publishing in a straitjacket that will eventually be discarded. However, the inertia of the scholarly publishing system is enormous, and so traditional journals have not changed much. They are in the process of migrating to the web, but operate just as they did in print. However, we are beginning to see the sprouting seeds of new ventures that will lead to new modes of operations. Still, it will be a while before they become a sizable fraction of the total scholarly publishing enterprise.

The large majority of scholarly publications are likely not to change much for several decades. However, there will be growing pressure to make them easily available. In particular, scholars are likely to press ever harder for free circulation and archiving of preprints. The realization will spread that anything not easily available on the web will be almost invisible. Whether they like it or not, scholars are engaged in a 'war for the eyeballs' just as much as commercial outfits, and ease of access will be seen as vital.

Ease of access is likely to promote the natural evolution of scholarly work. There will be more interdisciplinary research, and more survey publications. Some of these trends are beginning to appear in the data discussed in this paper, and we are likely to get more confirmations in the next few years.

Acknowledgements

I thank Steve Finch, Paul Ginsparg, Jim Gray, Eric Grosse, Kevin Guthrie, Stevan Harnad, Steve Heller, Patrick Ion, Don King, Kevin Kiyon, Greg Kuperberg, Leslie Lamport, Steve Lawrence, Carol Montgomery, Gary Mullen, Ann Okerson, Kimberly Parker, Robby Robson, Carol Tenopir, Ed Valauskas, Hal Varian, Tom Walker, Herb Wilf, for comments, corrections and for providing helpful information.

References

1. Odlyzko, A.M. The slow evolution of electronic publishing. In A.J. Meadows and F. Rowland (eds), *Electronic Publishing – New Models and Opportunities*, ICC Press, 1997, pp. 4–18 (available at www.dtc.umn.edu/~odlyzko).
2. Lustig, H. Electronic publishing: economic issues in a time of transition. In A. Heck (ed.), *Electronic Publishing for Physics and Astronomy*. Kluwer, 1997.
3. Association of Research Libraries, Statistics and Measurement Program: www.arl.org/stats/index.html
4. Tenopir, C., King, D.W., Hoffman, R., McSween, E., Ryland, C. and Smith, E. Scientists' use of journals: differences (and similarities) between print and electronic. In M. E. Williams (ed.), *Proceedings of the 22nd National Online Meeting*. Information Today, 2000.
5. Tenopir, C. and King, D.W. *Towards Electronic Journals: Realities for Scientists, Librarians and Publishers*. Special Libraries Association, 2000.
6. King, D.W. and Tenopir, C. Scholarly journal and digital database pricing: threat or opportunity? In J. MacKie-Mason and W. Lougee (eds), *Bits and Bucks: Economics and Usage of Digital Collections*. MIT Press, 2001, to appear (available at www.si.umich.edu/PEAK-2000/).
7. Anderson, K., Sack, J., Krauss, L. and O'Keefe, L. Publishing online-only peer reviewed biomedical literature: three years of citation, author perception, and usage experience. *Journal of Electronic Publishing* 2001:6(3) Mar. (available at www.press.umich.edu/jep/06-03/anderson.html).
8. Guthrie, K.M. Revitalizing older published literature: preliminary lessons from the use of JSTOR. In J. MacKie-Mason and W. Lougee (eds), *Bits and Bucks: Economics and Usage of Digital Collections*. MIT Press, 2001, to appear (available at www.si.umich.edu/PEAK-2000/).
9. Hunter, K. PEAK and Elsevier Science. In J. MacKie-Mason and W. Lougee (eds), *Bits and Bucks: Economics and Usage of Digital Collections*. MIT Press, 2001, to appear (available at www.si.umich.edu/PEAK-2000/).
10. Lawrence, S. Online or invisible? 2001, to be published (available online at www.neci.nec.com/~lawrence/papers/online-nature01/).
11. Luther, J. White paper on electronic journal usage statistics. *Journal of Electronic Publishing* 2001:6(3), Mar. (available at www.press.umich.edu/jep/06-03/luther.html).
12. Lenares, D. Faculty use of electronic journals at research institutions. In *Proceedings of the ACRL Conference, Apr. 1999* (available at www.ala.org/acrl/lenares.pdf).
13. Odlyzko, A.M. Tragic loss or good riddance? The impending demise of traditional scholarly journals. *International Journal of Human-Computer Studies* (formerly *International Journal of Man-Machine Studies*) 1995:42, 71–112. Also in the electronic *J. Univ. Comp. Sci.*, pilot issue, 1994, <http://hyperg.iicm.tu-graz.ac.at> (available at www.dtc.umn.edu/~odlyzko).
14. Licklider, J.C.R. *Libraries of the Future*. MIT Press, 1965.
15. Klopfenstein, B. Problems and potential of forecasting the adoption of new media. In J.L. Salvaggio and J. Bryant (eds), *Media Use in the Information Age: Emerging Patterns of Adoption and Consumer Use*. Erlbaum, 1989, pp. 21–41.
16. Christensen, C.M. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business School Press, 1997.
17. Odlyzko, A.M. Competition and cooperation: Libraries and publishers in the transition to electronic scholarly journals. *Journal of Electronic Publishing* 1999:4(4), Jun. (www.press.umich.edu/jep/); *Journal of Scholarly Publishing* 1999:30(4), Jul. 163–85 (also available at www.dtc.umn.edu/~odlyzko).
18. O'Donnell, J.J. The pragmatics of the new: Trithemius, McLuhan, Cassiodorus. In G. Nunberg (ed.), *The Future of the Book*, University of California Press, 1996

- (available at <http://ccat.sas.upenn.edu/jod/sanmarino.html>).
19. Trithemius, J. In *Praise of Scribes: De Laude Scriptorum*, ed. with introduction by K. Arnold, trans. R. Behrend. Coronado Press, 1974. Original manuscript circulated in 1492, first printed in 1494.
 20. Stevens-Rayburn, S. and Bouton, E.N. "If it's not on the Web, it doesn't exist at all": electronic information resources – myth and reality (available at www.eso.org/genfac/libraries/lisa3/stevens-rayburns.html).
 21. J.-M. Griffiths and D.W. King. *Special Libraries: Increasing the Information Edge*. Special Libraries Association, 1993.
 22. Lesk, M. *Practical Digital Libraries*. Morgan Kaufmann, 1997.
 23. Buckland, M.K. *Redesigning Library Services: A Manifesto*. American Library Association, 1992 (available at <http://sunsite.berkeley.edu/Literature/Library/Redesigning>).
 24. Bensman, S.J. and Wilder, S.J. Scientific and technical serials holdings optimization in an inefficient market: a LSU serials redesign project exercise. *Library Resources and Technical Services* 42(3) (available at www.lib.lsu.edu/collserv/lrts/index.html).
 25. Gazzale, R. and MacKie-Mason, J.K. PEAK: system design, user costs and electronic usage of journals. In J. MacKie-Mason and W. Lougee (eds), *Bits and Bucks: Economics and Usage of Digital Collections*. MIT Press, 2001, to appear (available at www.si.umich.edu/PEAK-2000/).
 26. Fishburn, P.C., Odlyzko, A.M. and Siders, R.C. Fixed fee versus unit pricing for information goods: Competition, equilibria, and price wars. *First Monday* 1997:2(7), Jul. <http://firstmonday.org/>. Also in B. Kahin and H. Varian (eds), *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*. MIT Press, 2000 (available at www.research.att.com/~amo).
 27. Odlyzko, A.M. Internet pricing and the history of communications. *Computer Networks* 2001: 36, 493–517 (also available at www.dtc.umn.edu/~odlyzko).
 28. Kolata, G. Web research transforms visit to the doctor. *New York Times* 2000, 6 Mar., A1 and A6.
 29. Wiles, A. Modular elliptic curves and Fermat's last theorem. *Annals of Mathematics* (2) 1995:141, 443–551.
 30. Darmon, H., Diamond, F. and Taylor, R. Fermat's last theorem. In *Elliptic Curves, Modular Forms and Fermat's Last Theorem* (Hong Kong, 1993), pp. 2–140. International Press, 1997.
 31. Coffman, K.G. and Odlyzko, A.M. The size and growth rate of the Internet. *First Monday* 1998:2(10), Oct. (<http://firstmonday.org/> also available at www.dtc.umn.edu/~odlyzko).
 32. Odlyzko, A.M. Internet growth: myth and reality, use and abuse. *iMP: Information Impacts Magazine* 2000 (www.cisp.org/imp/november2000/odlyzko/1100odlyzko.htm also available at www.dtc.umn.edu/~odlyzko).
 33. McMillan, G., Fox, E.A. and Eaton, J.L. The evolving genre of electronic theses and dissertations (available at <http://scholar.lib.vt.edu/theses/presentations/Hawaii/ETDgenreALL.pdf>).
 34. Sloane, N.J.A. My favorite integer sequences (available at <http://www.research.att.com/~njas>).

Andrew Odlyzko

AT&T Labs – Research

Current affiliation: University of Minnesota

Digital Technology Center

1200 Washington Avenue South

Minneapolis, MN 55415

USA

Email: odlyzko@umn.edu

Website: www.dtc.umn.edu/~odlyzko