

Predicting the Future of Scholarly Publishing[†]

by John Ewing

"I believe that the motion picture is destined to revolutionize our educational system and that in a few years it will supplant largely, if not entirely, the use of textbooks." *Thomas Edison, 1922*

"It is probably that television drama of high caliber and produced by first-rate artists will materially raise the level of dramatic taste of the nation." *David Sarnoff, 1939*

When Orville Wright flew his airplane over a small stretch of rolling grassland in 1903, the managing editor of *Scientific American*¹ predicted that thousands of planes would soon fly over every city, delivering patrons to theaters. On the eve of the First World War, two famous British aviators² argued that planes would prevent wars in the future (because they brought people together). Scientists, engineers, and futurists have *always* conjectured the consequences of technology. In the case of planes, the experts were right in recognizing that they would profoundly affect our lives in the coming century ... but they were certainly wrong in foretelling what that effect would be.

Once again the experts are predicting the future. The *digerati*³ tell us that the Internet has changed everything, that technology will revolutionize the way we do business, and that nothing will again be the same. Maybe. But the experts provide few facts to back their predictions, and they preach a digital future as an act of faith rather than a reasoned conclusion. It's hard to tell hype from reality when someone promotes technology with religious zeal.

What about scholarly publishing? Here, a *special* group of experts is predicting (and promoting) the future. The experts foretell the imminent collapse of scholarly journals and some advocate revolutionary replacements—refereed postings, e-prints, and overlays. In many countries, government agencies have embraced these predictions, providing support for alternatives—PubMed Central, the Public Library of Science, the arXiv. And experts offer miraculous solutions to previously intractable problems, describing a revolution in scholarly publishing that will provide universal free access to scholarship ... at no cost to anyone. The "free" alternatives seem to be enticing solutions to our present, very real problems.

How can we *predict* the future rather than merely *wish* for it? Good predictions are difficult without facts—facts about the past and about the present. This sounds obvious but, amazingly, experts and enthusiasts often dismiss past experience. They argue that

[†] Based on a talk given at the Conference on Electronic Information and Communication, Tsinghua University, China, August 29-31, 2002.

since everything will soon change, experience is not relevant. This kind of sophism is especially prevalent in discussions about the Internet, where experts tell us the old rules no longer apply.

But they are wrong: Making predictions without facts is mysticism, not science.

Facts

Here are some facts about the current environment in scholarly publishing.

Alternatives to journals have been widely publicized, and some of these are remarkably successful. The best known in mathematics are the arXiv (<http://www.arxiv.org>) and MPRESS (<http://mathnet.preprints.org>). The former is a repository of papers, and the latter is a distributed system with links to repositories, including the arXiv itself.

- As of mid-2002, the mathematics sections of the arXiv holds approximately 20,000 papers, with about 15,000 of those contributed by individuals and the remainder migrated from previously existing preprint servers.⁴
- MPRESS has links to about 25,000 papers (including those in the arXiv).
- Since 1998, mathematicians contributed 12,618 papers to the arXiv (through mid-2002). During this same time, Math Reviews indexed more than 280,000 journal articles.

Alternatives to journals are more popular in some fields than in others, but they have played a prominent role in discussions about electronic publishing. In 2001, the Association of Learned and Professional Society Publishers conducted a survey of scholars in many disciplines⁵. One part of the survey considered preprint servers.

- When asked whether preprint servers were important in their work, about one-third (32%) said Yes. (Among physicists, 55% answered Yes.)
- When asked whether they *used* preprint servers, 12% of the respondents said they did. (Among physicists, 32% did.)

Most scholars don't understand the scholarly literature—not its contents but rather its extent and complexity. When mathematicians think about "journals", they think about the best known and most visible—the ones they scan on the new-journals-shelf in the library. But the mathematical literature is far more complex and diverse. MR divides all journals into two classes: Those from which *every* article is either indexed or reviewed (the "cover-to-cover" journals) and those from which articles are selected for inclusion (the "others").

- In 2001, MR indexed or reviewed 51,721 journal articles⁶.
- Those articles came from 1,172 distinct journals.
- In 2001, 591 (50%) of the journals were "cover-to-cover".
- That left 581 (50%) classified as "other".

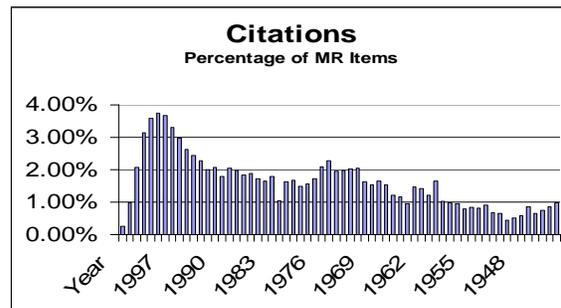
- And 30,924 (60%) of the articles were in "cover-to-cover" journals.
- Leaving 20,797 (40%) articles in the "other" journals.
- This means that 40% of the journal literature is outside the "mainstream" mathematics journals!

Almost all discussions about scholarly communication focus on electronic publishing. There is a recognition that the transition from paper to electronic is proceeding more slowly than first imagined, but almost no one understands how slowly.

- In 2001, only 46 (4%) of the journals covered by MR were primarily-electronic⁷.
- Only 1,272 (2.5%) of the articles were in primarily-electronic journals.
- On the other hand, approximately 34,000 (67%) of all articles had links, meaning that at least that many are available in electronic form.

Mathematicians have always known that past literature is important. Because MR recently added reference lists for articles for some journals, it is now possible to make the dependence more precise. The reference lists currently cover journals from 1998 to the present and include 336,201 citations to journal articles.

- Of all references, 53% were to articles published prior to 1990.
- More than 28% were to articles published prior to 1980.
- This is especially striking because the number of journal articles increased over time. Examining the number of papers covered annually by MR from 1950-1990, the percentage of MR items cited in recent papers varies between one and two percent for almost every year during the entire period.



Many scholars have commented about the high cost of commercial journals, but few have noted their number and size. They are gradually dominating the scholarly literature.

- In 2001, only 349(30%) of the journals were commercial.
- Yet they published 25,008 (48%) of the articles!
- Moreover, looking back ten years, we see a clear trend. In 1991, only 24% of the journals were commercial, publishing 38% of the articles.

What drives the expansion of commercial journals? While most scholars concentrate on the *costs* of journals, *revenues* are the crucial figures in understanding journal economics.

- A rough estimate⁸ suggests that the revenue from each article in commercial journals generates about \$4,000 in revenue (which may be off by a factor of 2.)

- Therefore, the 25,000 mathematics articles in commercial journals in 2001 generated about \$100 million in revenue for commercial publishers.
- An even rougher estimate suggests that for the non-commercial journals, each article generates about half as much revenue. Even for these, therefore, the total revenue was about \$50 million in 2001. (Again, this may vary by a factor of 2.)
 - And it's important to remember that mathematics is only a small fraction of all scholarly publishing. There are about 25,000 journals in science, technology, and medicine alone⁹. Just one commercial publisher, Elsevier, derives more than a billion dollars in revenue from its science journals.

These are the facts: Many scholars (although not most) promote alternatives to journals, but many fewer actually use them. Journals continue to dominate the scholarly literature in mathematics. Almost all journals are in both paper and electronic format, and almost none are electronic-only. The journal literature is highly dispersed, contained in many journals, including those that cover disciplines outside mathematics. The older literature is extremely important for current research. And finally, commercial journals are taking over an ever-larger fraction of the literature, with enormous financial incentives driving the trend.

What should we conclude from these facts? Here are two alternative predictions.

Prediction 1: The alternative models expand and pressure journals. The independents, with only scant operating margins, diminish further. The commercial journals, with deep pockets, continue to expand and add features. Commercial publishers consolidate and eventually dominate the scholarly literature.

Prediction 2: The alternative models expand and pressure journals, driving out the independent journals. The alternative models solve all their problems—financing, covering the disperse literature, archiving, etc. The commercial publishers close down their journals and walk away with their enormous profits.

Which of these predictions is correct? Many scholars hope for the second; only the first is supported by the facts.

Ecology

What's bad about promoting technology rather than predicting its consequences? We discovered the answer recently when we examined what a century of technological progress had wrought. The answer is *ecology*.

We normally think of ecology in terms our natural environment, but ecology can refer to any system and its relationship to the surrounding environment. The ecology of scholarly publishing includes many things—a system of refereeing and reviewing, the use of publications in hiring and promotion, the way in which scholars view their legacy of

research. Most experts on electronic publishing dismiss these things as unimportant; it's why so many get predictions wrong.

Should we worry about wrong predictions based on ignorance? Of course we should. If we fail to recognize that 40% of mathematical scholarship is published in multidisciplinary journals, we will design alternatives that ignore almost half the literature. If we believe that e-only journals are growing in number, when the number is shrinking, we may invest in the wrong trend. And if we ignore the fact that commercial journals take up not just more dollars but more shelf space as well, we risk sitting by like Nero while scholarly publishing is destroyed. Ignorance about the past and present of scholarly publishing is more than careless exuberance about the future; it means we can neither predict that future nor understand how to shape it. Remember—while some ecological disasters are caused by greed or malevolence, most catastrophes occur because well-intentioned people did not foresee the consequences of new technology.

The ecology of scholarly publishing is embedded in the far larger ecology of publishing, which currently has many forces driving change, and few of those forces have anything to do with scholarship or the academy. Ecological disaster for scholarly publishing would be swift and (largely) unnoticed by anyone outside academic life.

Many of the alternatives to journals may temporarily solve the problem of costs and speed of publication. For those who believe scholarly journals are merely a way for publishers to sell research back to the scholars who created it, this may seem like a fine solution. But people with publishing experience do *not* subscribe to this reductionist view. Journals are not just a way to distribute words on pieces of paper or screens; they are complicated institutions, involving authors, editors, libraries, researchers, publishers, professional societies, and administrators. Each has a role to play, and each has interests represented by the institution.

The institution of journals exists because scholarly publishing is not meant only for today's scholars but for future scholars as well — for our children and our children's children. Scholarly communication is more than sending papers to one's colleagues. Validation? Archiving? Financial incentives? These are all about sustaining scholarship for the future, not about exchanging papers in the present. Who will watch over collections when enthusiastic volunteers move on? Who will pay the costs of ever-changing servers and software to keep papers accessible? Who will provide the huge sums for archiving — not only saving the bits but updating the format of millions of papers? Surely we should not rely on government agencies, which have an increasingly short-term view in all their activities.

Many of the experts on electronic publishing assure us that these questions have easy answers. But we need to remember the lessons of the past: Predicting the consequences of technology is an uncertain business. Can we solve the problem of archiving in the future? To wave our hands with the assurance that technology will find solutions is like waving our hands for nuclear waste or carbon dioxide or fluorocarbons. We need to

worry about the future because no one else will worry about something as fragile as scholarship.

Conclusions

What should we conclude? Should we steadfastly maintain the status quo? Do we avoid technology altogether? Of course not. *We should experiment; we should try out new things; we should tinker with technology and find better ways to communicate.*

But in carrying out our experiments, we need to be cautious and we need to be humble. We should remember that in the past smart people were unable to predict the effects of technology. There is no reason to believe that today's smart people are any better at predicting than yesterday's. Trying out anything that comes to mind without understanding the effect on the entire system of scholarly communication may be exciting, but surely it is not wise.

We also need to be forward-looking. The *essence* of scholarship is what we leave for future generations, not what we produce for today's. Scholarly communication is not about us—it's about the future of our discipline. Many enthusiasts who promote new projects ignore this principle. Make changes now, they argue and worry about whether they are sustainable later. But if scholars themselves don't worry about their future, who will?

What about the experts? Treat them with skepticism. More information is better? Maybe. But nearly everyone is experiencing information overload today; perhaps the quality of information is more important than the quantity. Faster is always better? Maybe. But the bottleneck on the Internet is the person receiving the information, who often is not able to process what is already provided. The Internet will solve the problems of scholarly communication? Maybe. But scholarship and the Internet are different in an essential way: The nature of scholarship is long-term; that of the Internet is transitory. Finally, be especially skeptical of the experts who demand that you are either with them or against them. Subscribe to their vision of the future or be branded a Luddite. This is a false dichotomy—resist it.

Responsible caution is not the same as mindless obstinacy. It is possible to promote electronic publishing without promoting the dissolution of institutions that have served us well. It is possible to cultivate and shape those institutions without ripping out their roots. It is possible to have a revolution without renaming the months.

If we have learned anything in the past century, it is that even the most useful technology can destroy those things we value most.

¹ Waldemar Kaempfert, 1913

² Claude Graham-White and Harry Harper, 1914

³ A term used in *Digital Mythologies*, Thomas Valovic, Rutgers University Press, New Brunswick, 2000.

⁴ The number of papers contributed by individuals can be determined by counting submissions for each year at <http://arxiv.org/archive/math>. The total number of papers, including those migrated from other preprint servers, can be determined using the total number of papers given at <http://front.math.ucdavis.edu/>.

⁵ *The ALPSP research study on authors' and readers' views of electronic research communication*, Alma Swan & Sheridan Brown, Key Perspectives Ltd, ISBN 090734123-3. The survey was sent to approximately 14,000 authors of scientific papers across many fields. The response rate was about 9%.

⁶ For present purposes, books, proceedings, and all items other than journal articles are not counted.

⁷ The term "primarily-electronic" is not precise, but indicates journals that are either electronic only or that have a subsidiary paper copy added to the electronic version, which is viewed as primary.

⁸ Competition and cooperation: Libraries and publishers in the transition to electronic scholarly journals (see §2), A. M. Odlyzko. *Journal of Electronic Publishing* 4(4) (June 1999), www.press.umich.edu/jep/, in the online collection *The Transition from Paper: Where are we Going and how will we get there?*, R. S. Berry and A. S. Moffatt, eds., American Academy of Arts & Sciences, www.amacad.org/publications/trans.htm, and in *J. Scholarly Publishing* 30(4) (July 1999), pp. 163-185.

⁹ This figure is often quoted by the Association for Research Libraries, although it is hard to determine its precise source.