A Prognosis for Continued Disarray in Electronic Scholarly Communication

Gregory B. Newby
University of Illinois at Urbana–Champaign

Abstract: Activities scholars undertake to be viewed as productive and tenable (publication in traditional media) are out of sync with the activities they must engage to be well informed and well connected (participation in electronic communication forums). This work examines the challenges and provides a time line for the legitimization, codification, organization, and general maturation of electronic scholarly publishing. It is anticipated that the role of relatively unstructured, uncontrolled, and informal electronic scholarly communication will be of continued importance, yet will largely remain independent of efforts to create standards and protocols for electronic books, journals, and other transformed traditional media.

Résumé: Les activités propices à faire avancer la carrière de chercheurs (à savoir la publication dans les médias traditionnels) ne sont pas les mêmes que celles leur permettant d’être bien informés et apparentés (à savoir la participation aux forums de communication électronique). Cet article examine les défis que présentent la légitimation, la codification, l’organisation et le développement général de l’édition savante électronique, et propose un calendrier pour surmonter ces défis. Il prévoit qu’une communication électronique relativement non-structurée, noncontrôlée et informelle continuera à avoir un rôle important, tout en échappant en grande partie aux efforts d’établir des standards et des protocoles pour livres et journaux électroniques ainsi que pour d’autres médias traditionnels transformés.

Since delivering this paper, Gregory B. Newby has become an assistant professor in the School of Information and Library Science at the University of North Carolina in Chapel Hill, CB #3360 Manning Hall, Chapel Hill, NC 27599-3360. E-mail: gbnewby@ils.unc.edu. URL: http://www.ils.unc.edu/gbnewby/

© 1997 Canadian Journal of Communication Corporation

213
Introduction

Scholarly electronic publishing of all types plays an extremely important role in the academic world. Access to the Internet is nearly ubiquitous for scholars in North America and Europe. The network’s role is crucial for everything from announcing conferences, distributing calls for papers, and publicizing preliminary conference programs and tables of contents to researching, pre-printing, and publishing scholarly works. Scholars frequently subscribe to electronic journals, mailing lists, or network news discussions, and make use of the World Wide Web to retrieve current literature, news, and research. The Internet is a big part of academic life.

Scholarly publishing is the primary means by which the outcome of academic work is shared (at least in modern times). Journal articles, books, conference proceedings, and the like have been the primary delivery vehicle for scholarly work. There is little doubt that the Internet will soon augment these print media as a means of delivery and is, indeed, already doing so.

Why is the transition to electronic media taking so long? Why are we not receiving our academic journals on the Web, by e-mail, or in some other electronic form instead of in print? Examples of electronic journals, conference proceedings, and books abound, yet these are in the minority (and are often of lesser quality) when compared with print publications.

There is no short answer to the question of “what is taking so long?” This paper will present parts of a longer answer and attempt to estimate when the various components of scholarly electronic publishing will come into place. It is assumed without question that scholarly publishing will, by early in the millennium, take place largely in electronic forms. Whether this is “good” or “bad” is subject to debate elsewhere—it is submitted here that such a debate is comparable to debating whether automobiles or microwave ovens are good or bad. Scholarly publishing is. In the near future, scholarly publishing will be largely in electronic form.

There are many questions left unanswered in the debate. For example, the Web is often viewed (especially by Internet neophytes) as synonymous with the Internet. In fact, they are different: The Web (short for World Wide Web) only refers to the content and the content servers that transmit data using the hypertext transport protocol (http). The Internet, on which the Web is based, includes such facilities as electronic mail, remote login, and file transfer using protocols other than http, especially the file transfer protocol (FTP). The Web as it exists today is evolving and will eventually be superseded. The nature of computing will change; new standards for data exchange and networking will be introduced; television and other media will merge with Internet media. It is very difficult to predict what scholarly publishing will look like in 20 years, but it is not nearly so difficult to look at scholarly publishing in the late 1990s.
to determine what needs to change, what is changing, and what needs to be overcome to allow change.

This section discusses the move towards electronification of scholarly publishing, examining four major categories of challenges. Later sections will introduce details on components of the four categories.

Standards

One major area of challenge is the relative lack of standards of all types for electronic publications. Web-based publications, electronic journals, mailing-list contents, and so forth are difficult to retrieve due to the lack of controlled vocabulary and fields, such as are found in bibliographic databases (for example, Library of Congress Subject Headings, Title/Author fields, etc.). Indexing and searching tools on the Internet—the Internet search engines—are not able to distinguish the relative scholarly value of, for example, a 12-year-old’s page of favourite television shows and a media scholar’s critique of the state of network broadcasts.

Similarly, the provisions for including basic information about a particular document (meta-information) are weak. Simply identifying the author and title is difficult to do automatically, as is getting information about the publication date and history. These characteristics are particularly evident on the Web, but are not made easier when publications are distributed by e-mail or other means. Standard general markup language (SGML) offers a method to include significant meta-information, but it is not yet widely used in public Internet forums. (In addition, the diversity of document-type definitions [DTD] makes standardized generalized markup language [SGML] problematic for standardization.)

Legitimacy

A second area of challenge for electronic publishing is perceived legitimacy for the purposes of promotion and tenure. One of the motivations behind a great portion of scholarly publishing is the need of the authors to demonstrate the quality of their ideas through acceptance of written work in peer-reviewed journals. For every field, there is a hierarchy of journals with the best reputations. Similarly, some conferences have much higher standards than others by which papers, topics, or speakers are selected for presentation. Even for those electronic publications with strict peer review and a complete editorial board, these electronic journals, conferences, and books do not have the perceived status that print publications do.
Quality

The quality of electronic scholarly publications is also a problem. Quality can include issues such as the presentation, page layout, design, and graphical quality of articles, the peer review and editorial process, or the credentials of authors whose work is published.

Perceptions

A final main area of challenge is perceptions or models that academia has of scholarly electronic publishing. Even if issues of quality, legitimacy, and standards are met, the role of electronic (versus print) publications in academic life is based on perceptions the academic community has of that role. If e-journals are not perceived to have the same value for tenure decisions as print journals, then they will not have the same value. If conferences that only have electronic proceedings, not print proceedings, are not perceived as being of as high quality with those with print, then the perception will apply.

Specific instances associated with standards, legitimacy, quality, and perceptions will be discussed in the following sections, along with the prognosis for overcoming them. Overall, we can anticipate a multi-year transition towards an increased role for electronic publishing. There are, today, hundreds of examples of electronic journals, books, conference proceedings, etc. and millions of examples of Internet resources that are useful or play some role in academic work. In the future, we can anticipate that the term "scholarly publishing" will refer to materials in electronic form, with print used for specific subsidiary purposes such as archiving or appearing opulent. However, there are still many steps to be taken to reach this future.

Informal communication

Network newsgroups, mailing lists, and Web pages are frequently used to share preliminary research results, discuss issues, and to keep in touch with other scholars. The importance of these types of forums varies somewhat in different academic disciplines, but there can be no doubt that many individual scholars are able get important benefits from informal electronic communication.

Although books may be published on the Web, and electronic journals may be distributed by e-mail, the largest current use of newsgroups, mailing lists, and Web pages is for content that is not yet ready to be published as a journal article, conference submission, or book. Such forums may be used for "skywriting" (Harnad, 1996), for pre-publication of results, and many other purposes.

Today, it is easy for scholars to distinguish between, for example, e-mail discussion lists and print journals. Few scholars would be inclined to list the network newsgroups they read on their curriculum vitae, yet most would list
every conference presentation or journal article. Although some grey areas exist, there is a fairly definite boundary between “communication” activities of scholars and their “publication” activities. (One notable grey area is that many e-journals publish materials such as short essays that might have also been suitable for distribution to public mailing lists.)

Several areas of change to informal scholarly communication are under way. The first is that archives of communication forums are frequently used as information stores. Archives of mailing lists, current newsgroup contents, and even (though less frequently) logs of Internet Relay Chat (IRC) sessions or other interactive network forums are available for search or retrieval. (IRC is an informal synchronous channel for text-based Internet communication similar in structure to Citizens’ Band radio). This does not necessarily force a change in the communication that takes place in the forums, but it does change the means by which such forums might be accessed.

A second area of change to informal communication is somewhat less obvious, and has to do with gatekeeping and membership in the forums. Moderated newsgroups and mailing lists have been with us for some time, but private lists for scholars are seen less frequently. What we can anticipate is a more structured order for the ability to participate in or post to the most important informal communication forums. This stratification will be for purely pragmatic reasons: readers of the forums are frustrated when the level of discussion is limited by the frequent messages of newcomers or when commentary is more likely to come from graduate students than from well-known scholars. Private mailing lists already exist, but the model of these lists being for private discussion among eminent scholars but which may be observed by anyone interested is less frequently seen.

A final area of gradual change to informal scholarly communication is the means by which participation occurs. Currently, mailing lists have the feature of arriving in one’s personal electronic mailbox. Network newsgroups, however, must be sought out by a separate news-reading program. Electronic journals might arrive by e-mail, be posted as Web pages, or made available in other formats. We can expect some shifting in how materials are distributed as search and retrieval techniques are refined. For example, we might anticipate that query-by-profile systems will identify and deliver materials of interest from mailing lists without a subscription to the lists. Another example is the use of unified front-ends for network news, e-mail, and Web pages that we see in 1997’s Web browsers.

Informal scholarly communication is greatly facilitated by the Internet. The current generation of new scholars might find it difficult to imagine times when meetings, conferences, letters, and telephone calls were the primary method of discussing and sharing academic discussion. To the extent that “weak ties” among scholars are the truly important ones for getting their
work done, there is a great promise that continued enhancements to how we use the Internet for informal scholarly communication will prove tremendously empowering for all scholars.

The organization of information

Electronic library card-catalogues, bibliographic databases, CD-ROMs, and other systems for information retrieval rely on fields for identifying different types of information, and on controlled vocabularies for subject indexing. The tools we use today for accessing the Web, e-mail, electronic journals, etc. do not usually have these capabilities. Even when the meta-information about a particular document is present, there is no guarantee that automatic search engines or browsers will be able to access it correctly.

Standards for the communication of meta-information do exist, however. SGML may be used to tag author, title, and subject fields. Z39.50 is a bibliographic interchange standard that can allow multiple interfaces to access a database, such as a library card-catalogue. (WAIS [Wide Area Information Server], a set of standard tools for networked information retrieval developed in the early 1990s, was based on an earlier implementation of Z39.50.) Even with HTML, the META tag allows for the communication of fielded data.

The problem is not so much in the ability to include meta-information as in the lack of an ability to use it effectively. Perhaps more important is the problem of people self-authoring their own materials on the Internet (for Web pages, e-mail discussion groups, or even scholarly papers or conference proceedings) without knowledge of how to apply such meta-information.

The solution to this problem will likely come in the near term, through the tools we already use to access electronic information. New hypertext markup language (HTML) tags are introduced frequently (the current META tag may be used to communicate author information), and TITLE already exists but is used more for a running heading than an actual document title. Other fields can be introduced, and search engines will be able to offer the capability to search on these fields. This will lead to problems of training people to use such fields effectively, but this is less of a problem for the academic community than the general public. Regardless, the fact that millions of computer users have overcome the difficulty in mastering such arcane skills as HTML, uniform resource locators (URLs), and e-mail addressing gives hope that the public can learn to use features such as fields, authority lists, and query expansion and truncation effectively.

Information retrieval (IR) tools for full text exist, but they do not usually perform very well except with trained searchers. While efforts are under way to develop more sophisticated means of dealing with full text (Harman, 1994), the greatest hope for the near term is to add capabilities to search network-based publications using existing types of IR systems.
Involvement of commercial publishers

Commercial publishers (for the purposes of this section we include academic presses in this category) are in the business of creating products for sale. It has been demonstrated that the actual physical publication—the journal or book—accounts for only a portion of the costs of the publication process (see Fisher, 1996). Editing, reviewing, proofreading, publicizing, and many other activities are involved. In the case of commercial publishers, a goal is to profit from the income generated from the publications. Even in the academic press world, there is a necessity to strive to break even, if not profit.

Solutions to the needs of publishers to profit from their work on electronic publications are forthcoming, but have not yet emerged. A variety of economic models exist (see Newby, 1996), none of which are exactly matched to the type of one-item-one-fee approach amenable to books and journals.

The forthcoming solutions involve stronger emphasis on copyright and the creation of forums for the distribution of published items on a per-use basis. Although subscriptions to book series and journals will still exist, we can anticipate a far greater role for pay-once-use-once schemes for accessing electronic publications. For example, a Web search might yield an abstract for a scholarly article. Someone seeking to read the article could provide payment, then get access to the article to read and perhaps print one copy. The publisher would thus expect to generate revenues for their products over a far longer period of time than they do currently. This is because current models for print publications involve getting a copy of a book, journal, etc., then using it in perpetuity. In the new model, the publisher would sacrifice the one-time payment for the book, but then reap profits from its perpetual use.

Many forces on the Internet are working to assure the security of network-based transactions, where information or goods are delivered immediately, based on interactive payment. Use of the Internet for commerce is already upon us, and the amount of commerce on the Internet will grow exponentially through at least the first years of the millennium. Publishers will be able to use the same mechanisms as any merchant.

A remaining problem of concern to publishers is the issue of copyright and piracy. Currently, there is little to prevent someone with a single electronic copy of, say, a journal article from distributing that article to her friends and colleagues without a charge. Publishers want to be able to insure they can get compensation for every copy, without fear of illegal duplication. Although past history with software, music, and even print publications demonstrates the difficulty of preventing piracy, every indication is that piracy will be getting far easier. For example, one impediment to my copying an entire electronic conference proceedings to my personal hard drive (and perhaps making copies for my friends) is the size of the files involved. But as the storage capacity on my home PC exceeds several gigabytes, and the ability to write
CD-ROM becomes commonplace, the size of the files involved (and even the network bandwidth needed to retrieve them) will become trivial.

Publishers must work in several areas to overcome the difficulties of avoiding piracy. First, an effort must be made for authoritative sources to be easily and cheaply obtainable. If a pirated copy is easier and cheaper to get than the original, this will create a problem for publishers. Second, publishers and others need to provide the public with better knowledge about copyright laws. Many individuals will prefer to do the "legal" thing, but today's Internet offers plenty of evidence that most people do not understand the copyright status of electronic documents. Third, publishers must make their materials non-trivial to copy. This point is in conflict with current easy standards such as hypertext markup language (HTML), but fits reasonably well with Adobe publication distribution format (PDF) files and SGML. An example from the software world is the case of Microsoft Office on the Macintosh, where files are stored in at least four different locations on the computer, making it impossible to simply copy one directory to another computer to steal the software. Finally, and most importantly, publishers should strive to give reason to end users to make use of their publications on an ongoing basis. This can be accomplished by embracing the dynamic capabilities of the electronic world: providing interactive forums for readers; updating publications on a frequent basis; being pro-active about developing publications based on interest in current publications; and so forth.

Editorial structure

Print journals and conference proceedings of the mid-1990s involve entire teams of people. Editorial boards, layout experts, graphic designers, a reviewing corps, and so forth. At the same time, most electronic journals and conference proceedings are the work of only a few people; sometimes only one person. The great empowerment that the Internet plus modern computing tools offer to authors enables such electronic publications, but at the cost of some quality from having other people, with their expertise, involved.

There are only a handful of electronic journals that have editorial quality comparable to that of print publications. Yet it is the editorial board, the editor, and the publisher that help to maintain the stature of leading print publications. In turn, this leads to increasingly high-quality submissions—which leads to an ability to attract higher profile editors, publishers, etc.

There is no quandary here, it seems: the definition of the "best" or "most important" publications is, and has been, based on the quality of the works they contain, the authors they attract, the editorial board they list, and the overall professional presentation of the publication. There is every reason to suspect this set of criteria applies regardless of whether the format of the publication is print or electronic. There is some doubt about whether publishers
are a necessary component or not, but the print world has certainly demonstrated the value that publishers can add to scholarly publications.

The mission for creating "important" scholarly publications in electronic form is fairly clear, and some publications have already taken the necessary steps. Resolution of some of the other problems mentioned here will aid in progress towards the creation of electronic publications with the same editorial quality as print publications, but (as some key electronic journals demonstrate) there is no significant technical or social barrier to their creation today.

**Longevity of electronic publications**

The Internet has not yet been successful as an archival location for storage of publications (with few notable exceptions; see URL http://www.archive.org). On the Web, outdated material (such as announcements for last year's conference) can lead to the appearance that the site is not maintained properly—especially when, instead of leading to this year's conference, Internet search engines lead directly to last year's conference or the sponsoring organization's home page.

Only 50% or so of mailing lists and newsgroups are archived, and the archives are seldom perpetual. Rather, archives of last year's mailing list content might be deleted to make space for this year's archives. The cost of on-line storage is the culprit here—for even as disk drives get cheaper, the demands on system administrators for new mailing lists, more Web pages, and large disk quotas force continued diligence over allocation of resources.

In academic settings, there is typically an office for archives, or an archival library that is part of the main library. Modern archivists are well aware of the limitations of storage in electronic form, and only accept items such as floppy disks or magnetic tapes with the foreknowledge that these materials will be almost completely unreadable within just a few years. In the academic library setting, there is competition among budget items to acquire books and periodicals and develop computing facilities, in addition to general upkeep, salaries, etc. It does not seem likely that many libraries will be able to develop electronic archival capability (even for their own in-house materials) without significant changes in their budget allocations.

At a typical college or university, a computing services office maintains campus-wide facilities for computing, networking, Web-page storage, etc. Even in the universities that have appointed an "information czar"—a vice-chancellor or other highly placed individual with joint responsibility for the library and the computing environment—it is unlikely for the computing services office to engage in active archival activities.

What we can expect for the next few years is a tremendous and ongoing—and permanent—loss of electronic materials. As individual faculty move on, or as old computers are retired, or policies shift, or this semester's classes
start, the old Web pages, mailing-list archives, newsgroup contents, and so forth will be removed. As a new version of an electronic book is authored, the old version will be purged. It will take years yet for the academic environment to adjust to the needs of identifying and permanently archiving electronic materials. This function seems destined for the library, yet the library is not yet ready. One important step to their readiness will begin shortly, when libraries start to acquire publications in electronic form. A few have taken steps in this direction by subscribing to and archiving mailing lists and electronic journals. The larger step will not occur until the library must pay the same large annual subscription fee for an electronic journal as it already does for a print journal, CD-ROM database, book, etc.

In the commercial world, we can forecast a brighter near-term future. Inasmuch as access to older materials is valuable, there will be database providers or other vendors who will maintain such access. Thus, we can imagine that issues of electronic journals that are commercially published will remain available. There is still cause for concern, however: we know that out-of-print books still retain their copyright (at least for 75 years or so, depending on the country). Yet obtaining legal permission to reprint these out-of-print books, perhaps for a college seminar, is difficult and costly. Can we expect the same difficulties occurring with out-of-print electronic publications, where unusually large fees are levied for access to materials?

Luckily the role of scholarly commercial publishers will still be tightly bound with the need of scholars to have their work published for the purpose of obtaining tenure. We can expect some level of responsibility, then, on the part of the publishers to maintain permanent access to such works, even if a different fee structure applies for older materials.

Libraries can be expected to play their part in maintaining permanent access to materials they acquire (at least to the extent they currently do for print materials). However, they may be limited by the copyright or licensing constraints of the publisher. For example, it is current practice for many CD-ROM database vendors to require that all old copies of the CD be returned when a revision comes out and that the library may not keep any copies after they cancel the subscription. In this case, the library is unable to retain access to materials except as provided by the vendor.

**Legitimacy of electronic publications**

As should be clear from the sections above, there are some good reasons why tenure review committees are not, largely, ready to accept electronic publications as having the same value as print publications. Apart from the editorial process and quality of the electronic publications, the main issue is simply that most current electronic publications do not have editorial boards with the same “big names” as leading journals do. Many are maintained by one or a
few junior faculty, and many more encourage the publication of student papers or do not enforce peer review.

When, as is inevitable, the proportion and visibility of electronic scholarly publications shifts so that there is a far greater number of electronic journals, books, conference proceedings, etc. that have the same indicators of high quality and respectability as current print publications do, there will be no further need to convince tenure review committees of their worth. It appears unlikely, however, that this shift will be accompanied by a wholesale power shift away from commercial publishers and faculty with tenure.

While there is adequate room on the Internet for all types of scholarly publishing activities, there is also a continued role for commercial and academic publishers. Even as the fee system, copyright laws and expectations, and publication process evolve to encompass new electronic media, the basic role of scholarly publication as a means towards achieving tenure will remain. Indeed, even in many current academic environments where the role of tenure is changing, there still exists the need for scholars to self-legitimize through publications, in order to maintain or increase their academic status.

In 1997, there is a tremendous demand for quality control in electronic information. The level of interest in the Internet expressed by the corporations that already dominate Western media and communications makes clear that the obvious and easiest means of judging quality will be by source, not content. This is the same reason why public-access cable television is not popular, yet dreary situation comedies are—the glitter, the colour, and the snappy patter that media corporations produce cannot be matched by a single creative individual with a camcorder.

Similarly, we can expect that brilliant scholarly publications will have difficulty reaching their widest audience unless they are published by an important publisher or written by an already important author. There is still plenty of room to bypass the major players in the scholarly publishing field (whomever they turn out to be), just as independent films can win awards and independent music labels can get mass-market airplay. The 80/20 rule still applies: 80% of the material we see will come from 20% of the sources. Current television, newspaper, and radio ownership is closer to a 99/1 rule, as fewer than 20 companies control 99 percent of the mass media in the United States in 1996. See, for example, the June 3, 1996 special issue of The Nation (Miller, 1996). The democratic nature of the Internet, such that it is, combined with the specialized needs of the scholarly community, can give us hope that the ratio will be more favourable.

**Good signs**

In conclusion, the overall picture presented here is one of challenges, but also of considerable progress in meeting those challenges. Perhaps the largest
single force is the desire of scholars to participate in the electronification of scholarly publishing. It is in scholars' best interest for publications to be widely and instantly available and to avoid at least some of the delays inherent in the print publication process. From the consumer end, what scholar or student has not found it more convenient or expedient to search the Internet for publications of interest, rather than the library's card catalogue?

There is no reason why scholars cannot list electronic publications on curriculum vitae, and, provided they are Internet users, no reason why members of tenure review committees cannot take them into account. Perhaps the names of the new electronic journals will not be familiar, but the sponsoring institutions, editors, reviewers, or other authors may be.

Academic and commercial scholarly publishers have been relatively slow to move wholesale to electronic format, but almost all are interested and have some active projects. The level of maturity found in transmission protocols such as http and the extent to which expectations for royalties and subscriptions are reasonable seem to indicate that there is indeed little reason to hurry, lest the hurried lead to poor products or lost profits.

The Internet as a whole, and the means we use to communicate, store, and transmit information, is not yet in a nearly finished state. There is every reason to suspect that the desktop computer of the near future is today’s supercomputer; that today’s T-1 network connection is tomorrow’s modem; that interactive graphics and displays of tomorrow will make today’s VR games look like "pong." Even if problems of effective retrieval from full-text databases prove difficult, we will be able to engineer current means for searching to work more effectively with electronic publications.

This work has attempted to paint a realistic picture of ongoing activities and some important challenges in the move towards the electronification of scholarly publishing. It is accepted at the outset that scholarly publishing as we know it will take place largely in electronic formats. The exact timing of this change is difficult to predict, as is the timing for overcoming specific challenges discussed here. On the whole, though, there are no problems that appear intractable, and enough interest in solving them from outside the academic world (media outlets, microcomputer vendors, database providers, banks) that we can expect these problems to be solved fairly rapidly. New problems will arise, no doubt, and the road to scholarly publishing of 2010 or 2020 will be rocky. Even though the destination is unclear, the path for the upcoming few years is before us.

References


Scholarly Communication in the Next Millennium
Selected Papers from Canada's Policy Conference

Rowland Lorimer, John H. V. Gilbert, and Ruth J. Patrick, Editors

Canadian Journal of Communication
Volume 22 Issues 3/4 Special Issue