One Giant Leap, One Small Step: Continuing the Migration to Electronic Journals

JOHN H. BARNES

ABSTRACT
A number of recent events, most notably the emergence of the World Wide Web, have triggered a dramatic increase in the availability of electronic journals. Today's electronic journals make use of the technologies of the present but are also based on the experiences of the past. This article discusses the steps that have been taken to reach the point where we are today in electronic publishing, including CD-ROM systems, local data loading, and pre-Web online, and outlines some remaining obstacles, including critical mass, aggregation, local collection development, integration, and archiving, that must be overcome before libraries can make the ultimate leap from paper to electronic collections.

INTRODUCTION
With apologies to Neil Armstrong for restructuring his famous lunar line, this article will discuss the forces that have led to the recent "giant leap" in electronic journal publishing, yet remind us that it is but "one small step" down the path of library's conversion to electronic collections.

For more than a decade now, much has been written about the coming transition from paper to electronic collections in the world's libraries. The technologies-of-the-time, be they computer-output microfilm, CD-ROM, or FTP, have been utilized in projects and products that seek to facilitate this transition. While a great deal has been learned from the projects and there have been some relatively successful commercial endeavors, it has only been within the last two years that the major building
blocks appear to be in place to begin the transition in earnest. A number of significant recent events are dramatically impacting the direction of scholarly electronic publishing, namely:

- The emergence of the Web has greatly reduced the entry barriers to electronic publishing by providing a ubiquitous real-time distribution channel and eliminating the need to develop and distribute proprietary access systems.
- While most electronic publishing is still about "putting paper on the screen," new dynamic data formats such as Hypertext Markup Language (HTML), Portable Document Format (PDF), Virtual Reality Modeling Language (VRML), and others are providing new cost-effective means to "liven" online journal information through color, graphics, document linking, video, and simulation.
- Web-based electronic publishing has gained acceptance among the end-user community with the introduction of numerous mainstream consumer publications, including such notables as *Time*, *U.S. News and World Report*, *The Wall Street Journal*, *Science*, and *Nature*. Nearly every major commercial publisher is offering, or will soon offer, online versions of their publications. Even the scientific, technical, and medical publishers, long the most conservative and cautious with regard to electronic publishing, are now actively converting and repackaging their information for online distribution.
- The results from the major academic experiments of the early 1990s, such as TULIP, CORE, and Red Sage, are being publicly shared, and there seems to be agreement in key areas such as the high cost of local electronic journal storage, the importance of integration with bibliographic reference systems, the need for interdocument linking, and the necessity of a permanent electronic archive.

All of these environmental factors combined indicate a notable change in the direction of publishing that will inevitably dramatically affect libraries. The result has been a pronounced acceleration on the part of both the scholarly publishing and library communities to address the difficult questions of beginning the transition to electronic collections. Yet with all the changes brought on by the Web and associated technical innovations, much remains the same in the library's mission. We are moving down a path of continuously better solutions to the same problems of collecting, accessing, and archiving information. What are the issues that still need to be addressed in order to finally make the giant leap to electronic collections?

One of Many Leaps

The discussion of the implication of electronic publishing on libraries, users, and traditional publishers is rich with topics for debate: from
the impact on the relationship between the academic community, which produces a great deal of the scholarly literature and the commercial publisher who distributes it (Tenopir, 1995), to the advantages of electronic journals over their print counterparts (Hickey, 1995, pp. 530-31)—namely customization, searchability, information linking, availability and timeliness—to the technological merits of one storage format, be it TIFF, HTML, SGML, or PDF, over another (Weibel, 1995; Kennedy, 1996; Kirstein & Montasser-Kohsari, 1996). These debates continue, although they are not the primary focus of this discussion. Instead, we will start with the assumption that the advantages of electronic journals are accepted and outweigh the disadvantages, and that traditional publishers will continue, into the foreseeable future, to be the source of most of the information content.

In discussing electronic journals, this article refers to second-generation electronic journals (Duranceau et al., 1996, p. 50), namely multimedia representations, usually of existing print publications, that are distributed online and include searching, browsing, and output capabilities. While first-generation electronic journals—i.e., simple ASCII-text files usually noncopyrighted and distributed by a listserv and electronic-only journals—are not the primary focus, many of the same issues apply for libraries.

The Waiting Game

As is always the case with a fundamental lifestyle change, be it personal or professional, there is the period of wait-and-see as the “changer” and “changee” each hopes the other will make the next move and signal the future direction. Whether it be the invention of the automobile and its impact on travel, the standardization of the VCR and the resulting changes to home entertainment, or the proliferation of the personal computer and the impact it has had on productivity and information dissemination, each major lifestyle change is preceded by a period of considerable uncertainty.

In the end, the fundamental change does not take place until there is enough critical mass moving in one direction creating the necessary momentum to pull the rest along. The Model T brought direction to the fledgling automobile industry, the VHS tape format established the standard necessary for VCRs to take off, and the “Wintel” personal computer provided the standardization that has led to tens of millions of home computer users.

The same market dynamic is true for electronic publishing and libraries. Primary publishers, secondary publishers, and information vendors (hereafter information providers) have, over the past decade, experimented with projects and products aimed at positioning themselves for the much proclaimed coming of the electronic collection but, for the
most part, have taken a wait-and-see attitude before completely retooling production processes. Libraries, while experimenters in many of the projects and products, have also, predominantly, taken a wait-and-see viewpoint before "retooling" their collections by making wholesale conversions away from paper. The primary reasons are the lack of standardization and the high-cost of wrong guesses for both information providers and libraries.

The recent "giant leap" that we’ve seen in electronic publishing has taken place primarily on the supply side of electronic journals—the information provider side—and has specifically addressed the issues of standardized distribution and information timeliness. The impetus of this giant leap is undoubtedly the emergence of the Web and reduction in entry costs and the complexity it is providing to the information provider. This, in turn, is providing the essential standardization and critical mass of electronic journals that the library is looking for to finally start the migration from paper.

LOOKING BACK—THE EARLY STEPS

While the Web has not solved the information provider’s dilemma of how to get the source material to an electronic format, it has addressed many of the issues of distributing the information thereafter. Just a few years ago, the information provider seeking to distribute an electronic journal was faced with constructing a solution from start to finish. This included not only the redesigning of the journal production system, but also selecting data formats, creating access software, and implementing full distribution systems. For the primary publisher, it usually meant partnering with a vendor who could provide the technical expertise to convert the print journal to electronic, develop software to access it, and provide distribution to the library. The "giant leap" we are experiencing now was preceded by many smaller steps. Most early solutions were of one of three varieties: (1) CD-ROM, (2) local loading, and (3) "old" online.

CD-ROM Systems—The First Step

The late 1980s and early 1990s saw various journal publishing and document delivery solutions based upon CD-ROM technology. While the CD-ROM-based systems offered significant storage economies and a means to distribute electronic journals, accompanying production processes significantly affected information timeliness.

While in some instances raw journal data were available from the publisher in electronic format, more frequently this was not the case. The first step in the process was usually the conversion of the journal from paper to electronic, usually through scanning and creation of bit-mapped images. Added to this was the time for indexers to create descriptive bibliographic citations and abstracts and to link this information to the image of the article.
While much of this has now been outsourced to service bureaus who specialize in these conversion services and can handle high volumes, there still remains a notable delay in timeliness brought on by the physical distribution (i.e., packaging and mailing) of the CD-ROM. In addition to the high cost and lengthy production process for CD-ROM, the information provider was also required to develop, support, and continually enhance proprietary software used to access the electronic journals. Each new journal or journal collection brought with it a new interface and different functionality for browsing, retrieving, and displaying documents. The significant development effort for the information provider, coupled with the lack of interface uniformity for libraries, placed more roadblocks in the path of migration from paper to electronic journals.

While several CD-ROM-based electronic journal products have experienced success, they have not provided the platform for the giant leap from paper. In the end, the lack of timeliness has meant that these were essentially electronic document delivery systems—a highly useful complement to paper but not a complete replacement for the paper subscription (Lancaster, 1995).

Nevertheless, CD-ROM electronic journal systems did provide an important first step. For the information provider, they uncovered the key production, distribution, and technological issues that must be overcome in providing electronic journals and enabled them to build the necessary infrastructure for future growth in this area. Those who made the leap with CD-ROM have the great benefit of the knowledge and experience and are better prepared than others who chose to stand on the sidelines.

For the libraries which integrated CD-ROM-based electronic journals into their reference services, they too gained from the experience. They have improved the storage and accessibility of their journal collection, have greatly reduced the time to deliver documents, and have introduced their users to the benefits of electronic journals. They are better prepared for the leap.

Local Loading—A Side Step

At the same time that many information providers were focused on CD-ROM solutions, some publishers experimented with the approach of providing raw electronic journal data directly to libraries for local loading. Elsevier’s TULIP Project and the CORE Project at Cornell are examples of this type of approach (a full report on the TULIP Project can be found at: http://www.elsevier.nl/locate/TULIP).

The premise was that publishers would provide image and text data for their journals directly to the library. The library would be responsible for storing the journals and developing the software necessary to access
them. Under this model, each library could, therefore, control the presentation and integration of the journals into their local library systems. Additionally, it was envisioned that timeliness of the material would be improved since the aggregation and CD-ROM production/distribution processes were bypassed.

While this model simplified the process for the publisher, it did so at the expense of the library, which was then faced with the daunting task of loading, indexing, and making available this large store of data. The cost advantages brought on by one production and distribution process shared by many were lost as each institution was required to redundantly develop its own storage and access platform. Projects such as these have shown that a significant technical and logistical infrastructure is required to support such endeavors and "the number of academic libraries really ready to support digital collections is not large" (Hunter, 1996, p. 210).

Similar to CD-ROM-based solutions, these projects were significant and important in the learning process for libraries and information providers with regard to electronic journals. They pointed out the vast differences, both in effort and cost, with building an electronic versus a paper collection. They highlighted the advantages of aggregation and economies of scale that are necessary to keep costs down, and they revealed key issues such as who maintains the journal archive and how many journals represent a "critical mass."

"Old" Online—Steps in the Right Direction

"Old" online refers to the now seemingly ancient days of pre-Web electronic journal solutions. They relied on proprietary networks, dial-up services, and the early incarnations of the Internet as the access channel to the electronic journals. The early electronic-only journals were also born under this model—a result of the significant reduction in production and distribution costs for the online electronic journal versus CD-ROM or paper equivalents. OCLC's original Electronic Journals Online (EJO) service is an example of this type of service. EJO pioneered the online electronic journal in 1992 with the development of the Guidon interface and the introduction of The Online Journal of Current Clinical Trials with the American Association for the Advancement of Science.

Like their CD-ROM counterparts, these online electronic journal systems relied on proprietary interfaces, usually customized for the specific features of the journal. Unlike CD-ROM, they suffered less from issues of information timeliness because of the elimination of the physical production and distribution of a disk yet were plagued by other issues, such as a limited telecommunications infrastructure, proprietary client interfaces, and low content availability. They saw some success, more typically from the individual subscriber who was focused on a few
journals than the library researcher who wanted to search across a broad body of information. The information timeliness and journal presentation issues had been largely addressed, but there remained significant obstacles such as high startup and maintenance costs, proprietary client interfaces, and questions about archives which remained unresolved for information providers and libraries.

**Today's Electronic Journal—Bigger Steps**

As we have all witnessed, the level of electronic publishing activity has grown exponentially over the last few years. The primary driver of this accelerated activity, of course, is the emergence of the Web. From the thousands of individual online journals, newspapers, and "e-zines" to niche-targeted online professional "clubs" such as BioMedNet, MathSciNet, and ChemWeb, to publisher-specific web sites which provide access to most, if not all, of a given publisher's content, Web-based publishing is taking many forms.

It appears that the Web is the standardizing element that has been needed to stimulate the migration to an electronic environment. Its ubiquitous nature and the uniform browser access software seems to have removed one of the final barriers, namely distribution and access, in the path of the information provider looking to make the leap to electronic journals. Until the distribution and access problems were resolved, primary publishers were not likely to invest significantly in retooling their journal production processes to produce electronic material. It was more cost effective to rely on the secondary publishers or service bureaus to make the conversion from paper—thereby continuing to propagate a document delivery solution over the true electronic journal solution.

From all appearances, the Web will certainly have a profound impact on how the average consumer accesses information. The explosion of information and seemingly insatiable appetites of consumers for it seem to be testimony to that. But are the needs of the consumer, the average home computer user, the same as those of the library patron? Is the Web really the missing piece in the puzzle for libraries looking to make the leap to electronic collections? Or is it just one of several pieces missing?

**Approaching the Leap**

The emergence of the Web has undoubtedly greatly increased the volume of information available online and has introduced millions to the value of electronic information. But for the researcher, this has largely come at the expense of relevance and cohesiveness. Getting 1,000+ hits on a Web search engine is not necessarily providing a valuable service to the user. Similarly with electronic journals, having to locate the various relevant journals on the Web and then repeat searches across each one to locate the desired information is also of limited value. The Web has driven
an information explosion by improving access to it, but it has also quickly reminded us of the role the library plays in enabling the researcher to find the right information and to assure its continued existence.

With electronic journals, as with other information formats, the library will continue in its role of defining the collection for its constituency and assuring that that information is available, relevant, and easily accessible. The Web itself does not make this happen, but it can play a part in the solution. So what does the library need before it makes the giant leap from paper to electronic collections? A review of the findings from the various electronic journal products and projects over the past decade points out five key nonpricing areas that must be addressed: (1) critical mass, (2) aggregation, (3) local collection development, (4) integration, and (5) archives.

Critical Mass

Obvious in its statement, significant migration from paper to electronics cannot occur until there is a sufficient critical mass of journal content to warrant the effort involved in implementing the transition (Hunter, 1996, p. 210). As was discussed earlier, this is the primary barrier to any fundamental change in or out of the library and has been a leading deterrent to adoption of electronic journals up to this point. While there have been products available with several hundred titles in a subject area, they have largely been for document delivery purposes rather than true journal publishing. Projects and products in the primary electronic publishing area have rarely accumulated more than fifty titles in any subject area. Far from the necessary critical mass.

As noted, the emergence of the Web and standardized data formats are addressing this issue. In just a few years we have seen thousands of journals, newspapers, and magazines become available on the Web, and the expectation is that within a few years nearly all major journal publishers will have their information available in electronic format. The issue of critical mass will soon disappear.

Aggregation

While critical mass is a stimulator for the migration to electronic journals, it does not necessarily address a fundamental need for the library and researcher, namely a consistent and efficient means to locate information. In paper journal collections, this comes from the fact that they are all of relatively the same format, are typically available in one location, and are pointed to by various bibliographic reference tools. The same must be true, virtually, if not physically, in the electronic collection.

The Web has enabled primary publishers to more easily enter the area of electronic journals, but with this comes countless interfaces, document formats, and access procedures. The library cannot construct a viable electronic collection if it must accommodate potentially hundreds
of different variations in the access methods for its journals. Similarly, TULIP and comparable projects showed us the inefficiencies and high cost structures in a model where each library loads electronic journals locally. They clearly demonstrated the need for a centralized archive and technology sharing model. True cost savings can be realized by creating one shared online electronic journal collection, with each library free to license its own mix of journals from the pool of titles available. While electronic journal licensing agreements are unlikely to grant resource-sharing rights comparable to those available with paper collections, there is still opportunity for resource sharing in the broader sense. The resource sharing is of the infrastructure and technical expertise necessary to archive and access the information. As Jim Neal (1996) states in his article "Academic Libraries: 2000 and Beyond," "we need the virtuous library to share collections, technology, and expertise and to partner in the packaging and delivery of information" (p. 74). Aggregation of electronic journals into a common format with a consistent interface is key to making them efficiently usable by the researcher and cost-effective for the library.

Local Collection Development

As with paper journals, decisions on collection development should be made at the local library level and should be made on a title-by-title basis. The local library, be it academic, public, or corporate, is responsible for assembling a collection that matches the needs of its user community. This includes determining the individual journals to subscribe to as well as the choice between subscription and document delivery. The distribution of electronic journals on preproduced media such as CD-ROM made it unfeasible to provide this journal-level flexibility in the past. Online distribution makes it again feasible, and libraries should make it a requirement of any electronic journal service. Publishers will continue to offer content bundles, today probably still at the journal level, but increasingly this will move to the article or concept level. The new electronic journal services must facilitate all of these options and not force "pre-packaged" collections on the library.

Integration

One of the distinct benefits in moving to electronic collections, especially those available online, is the ability to directly link the tools for searching and locating information to the information itself. Just as full-text databases have flourished for their one-stop offering of locator and document, so too will the new electronic journal services flourish once they are linked to the bibliographic databases and local systems used to locate information in the library. One of the great challenges with CD-ROM-based electronic journal systems was to find a way to integrate them with the local library system to provide a single solution. Some were
successful but not without a significant commitment of local technical and support resources. The Web, and its inherent interlinking infrastructure, makes this more easily accomplished. Yet the great majority of Web-based electronic journals products thus far have been standalone services, largely “disconnected” from established bibliographic databases, such as Medline, INSPEC, or ERIC, and local library systems. The library does not experience the entire benefit of the electronic collection until all of these pieces are integrated into one service.

**Archives**

Often overlooked by those outside the library community in their rush to electronic information is the fundamental role the library plays in providing an archive for information. Most early electronic journal products have primarily focused on the benefits of access. While certainly important, they cannot come at the expense of archival rights. Just as the “I-buy-it-I-own-it” right is a given with paper collections, it must also be considered so with electronic collections. The library is looked on as the permanent record for information, with its resources available indefinitely (Neavill & Sherble, 1995); electronic information should not change this definition. Without a doubt there will be no leap to electronic journals if this basic right is not granted by the journal publishers.

While it is necessary to have the same archival rights with electronic journals as with paper, it is not necessary to replicate the electronic archive as has been done with paper and microfilm. Just as the publisher must reevaluate their licensing approach with regard to archival rights, the library must reevaluate the means by which it provides the permanent archive. The library must continue in its role of defining the collection for its constituency and assuring that that information is available and easily accessible (Duranceau et al., 1996), but the redundant physical storage of the collection is no longer necessary.

In evaluating this decision of where the electronic archive is housed, the library must make tradeoffs among permanency, trust, and cost. Neavill and Sherble (1995) provide a comprehensive analysis of the costs and efforts involved in establishing a local electronic archive for the individual library. They conclude that while “local archiving appears to be the most reliable way for libraries in today’s transitional environment to ensure that their users will have adequate and continuing access to files...no library should establish an electronic archive without careful evaluation of the implications for the library and the inherent problems involved” (p. 15). Duranceau et al. (1996), in their discussion of their experiences with electronic journals at MIT, state similarly:

> we will, in effect, be moving away from a repository model in the direction of a gateway model, until and unless we can participate in some kind of national archiving project....We acknowledge that this
shift represents a major change in our thinking about our role as a research library; it forces us to meet patron needs more and more through remote access, rather than through onsite holdings. Given our financial and staff constraints, we see no other viable option over the short term. (p. 55)

At the other end of the archive control spectrum is the option of relying on the publisher of the journal to serve as the permanent electronic archive. Even putting aside the aggregation issues raised earlier, there remain key concerns with this approach. As Neavill and Sheble (1995) state: "[M]any publishers of electronic journals maintain archives of backfiles at network sites, but there is little assurance that these files will be available permanently. Publishers may discontinue individual titles or go out of business altogether" (p. 14). MacEwan and Geffner (1996) came to similar conclusions in their discussion of the electronic journals collection of the Committee on Institutional Cooperation and stated: "Currently, it is unclear whether publishers will preserve and provide continuing access to electronic materials they produce. This is a critical problem for research libraries" (p. 7).

There also exists the issue of technical expertise. The publisher’s traditional expertise lies in the area of managing the process of creating and distributing printed information. The technical infrastructure necessary to support worldwide access to an electronic archive is quite different. Many publishers, most notably the smaller professional organizations and scholarly societies, will not have the ability to create and support such a service, yet there will still exist the need for the electronic archive. Publisher as archivist, while cost-effective, may jeopardize the ability of a library to assure its patrons of perpetual access to its information.

A third option, and the one likely holding the most promise, is a cooperative effort within the library community to create a centralized archive of electronic journals. As was discussed earlier under the section on “Aggregation,” an inherent benefit of online electronic information is that there can be one copy shared by many. This may or may not reduce the cost of licensing the journal (that, in itself, is a rich topic for debate), but it certainly significantly reduces the cost associated with providing access to it. As has been done successfully in areas such as cataloging, the library community should seek to cooperate in the establishment of a shared archive for electronic journals (Neavill & Sheble, 1995). By sharing the cost of data storage, access, and format migration, libraries will be able to make the move to electronic journals thereby adding value and reducing cost. This has both short-term and long-term consequences as Neal (1996) notes when he states that “the use of dedicated or shared collection storage facilities and the expansion of digital network delivery directly to faculty and students also minimizes the need for new building
construction" (p. 76). In addition to direct cost saving, this centralized archive could assure authenticity, consistency, and permanency for each journal and would provide a single source that could be migrated as new data formats and delivery mechanisms evolve. The resolution of this issue alone will have the most significant impact on when the library makes the giant leap to electronic collections.

CONCLUSION

The explosion in information services spawned by the emergence of the Web has created an exciting time for libraries and the entire information community. This is no more evident than in the area of electronic journal publishing. Not only will most journals be available electronically in the coming years, but they will be available in ways much more dynamic than today's "print on the screen" model. Whether it be imbedded software applets, video snippets, chemical modeling, or mechanical simulations, the journal of the near future will be a rich information experience. Yet for all the technological possibilities, the basic needs of the researcher and the role the library plays in meeting these needs remains strikingly consistent. For the giant leap to take place, rather than more small steps, requires that these needs be recognized and that publishers and information vendors address them in the products they offer.

REFERENCES


