Publishing Biomedical Journals On The World-Wide Web Using An Open Architecture Model

E.P. Shareck, MD, Robert A. Greenes, MD, PhD

Harvard-MIT Division of Health Science and Technology, and Decision Systems Group, Brigham and Women's Hospital, Boston, Massachusetts

Background. In many respects, biomedical publications are ideally suited for distribution via the World-Wide Web, but economic concerns have prevented the rapid adoption of an on-line publishing model.

Purpose. We report on our experiences with assisting biomedical journals in developing an on-line presence, issues that were encountered, and methods used to address these issues. Our approach is based on an open architecture that fosters adaptation and interconnection of biomedical resources.

Methods. We have worked with the New England Journal of Medicine (NEJM), as well as five other publishers. A set of tools and protocols was employed to develop a scalable and customizable solution for publishing journals on-line.

Results. In March, 1996, the New England Journal of Medicine published its first World-Wide Web issue. Explorations with other publishers have helped to generalize the model.

Conclusions. Economic and technical issues play a major role in developing World-Wide Web publishing solutions.

INTRODUCTION

Information technology is at the center of a revolution in the way that medical information is developed, referenced, employed, distributed, and funded [1,2]. Internet technologies, such as electronic mail capability, direct file-transfer protocol (FTP), newsgroups, listservers, menu-based Gopher, and hypertext browsers (HTTP), have enabled individuals and institutions to develop platform-independent distributed architectures for composing, editing, organizing, and deploying text, graphics, video and sound [3].

Biomedical publications are ideally suited for distribution via the World-Wide Web (WWW) due to well-developed indexing technologies, extensive cross-linking, and longevity of their content's value. However, publishers of many traditional scholarly journals have been cautious in their embrace of the WWW.

Obstacles to the adoption of new technologies are largely economic, and include: (a) concerns about diminished subscription and advertising revenues, (b) need for time and money investments in information technology and personnel, (c) absence of a proven subscriber base, and (d) issues relating to conflicts between tracking demographic data and preserving reader privacy. Some of these problems can be addressed by (a) new developments in tools and protocols for enhanced control over layout and (b) linking and efficient, secure on-line control of payment methods and billing systems.

Interest in on-line technologies has sharpened as publishers face the challenges of diminishing subscription revenues [4], declining advertising support, and the increasing costs of production and distribution [5]. This interest is further piqued by the Internet's potential for delivering archival content. Publishers have had few choices in repurposing their archives: they have been limited to issuing reprints, granting licenses to reprint services, providing their abstracts to the National Library of Medicine to produce MEDLINE indices, or licensing their full text to third parties for processing and transformation into CD-ROM products. With the exception of MEDLINE, which is also licensed to produce proprietary site-licensed MEDLINE products, these arrangements have represented a relatively modest contribution to current operations.

In this paper, we describe the development and characteristics of tools and an architecture that support on-line publication of biomedical journals as independent entities, while providing incentives for publishers to create and participate in the biomedical information market. This journal-centric approach centers on the WWW as a medium that:

- provides an alternative distribution channel for new and archival material,

0195—4210/96/$5.00 © 1996 AMIA, Inc.

- augments the value of a publication's content by linking it to that of other publications and indices,

- offers publishers the option of a gradual adoption of different aspects of the model, thus preserving traditional relationships until they can be shifted to new balance points,

- enables information providers to establish flexible relationships with subscribers and advertisers,

- establishes a market mechanism for the development of effective information pricing structures,

- expands readership for biomedical information through links to previously untapped segments of the information market, potentially resulting in lower information costs,

- makes possible the demographic analysis of consumers and the implementation of one-to-one advertiser support,

- encourages information consumers to trade information about themselves for reduced access costs,

- provides incentives for libraries and indexing services to innovate and migrate to new roles for their expertise and technologies.

This strategy employs a non-proprietary architecture which is open and modular [6]. It fosters interconnection and adaptability by enabling publications to determine their interests and develop partnerships while preserving their independence.

METHODS

We have worked with the New England Journal of Medicine (NEJM), as well as five other publishers of biomedical journals. We have developed a phased approach that has as its end-product a distributed database in which each publication constitutes an independently developed and maintained node with non-hierarchical connections to other nodes via the Internet. The first phase of the project was centered on assisting a number of learned society publications to achieve an on-line presence with few or no external connections. Since these endeavors were not expected to generate revenue in the near future, it was important that they be implemented at minimal cost, both in terms of time and resources.

Towards this end, we took care to integrate electronic publishing where possible into the normal publication work-flow. This meant that content acquisition produced word processing files, page layout files, customized archive formats and SGML output, mandating the adoption of a common intermediate format. While SGML was attractive for its contextual markup capabilities, BRS was chosen as the intermediate format which would be submitted to the manuscript processing for on-line publication, because of its recent adoption by publishers, the importance of being able to handle the conversion of archival material, and the existence of BRS archives for a number of the journals.

With the adoption of BRS as an intermediate, a prefilter was developed in order to process nonstandard marks and non-BRS files. Both the prefilter and the manuscript processing tool were to be configurable, requiring a development environment with strong user interface features, and, because they were to be developed in the setting of an active publication with deadlines and changing requirements, the environment would also need to be interpreted and extendible. For these reasons and the availability of easily written external commands and functions, Apple Computer's HyperTalk was used for coding the prototype.

Manuscript processing was automated and initially composed of seven phases:

(1) extraction of article-specific and issue-specific values, such as dates, page numbers, titles and author names;
(2) assembly of derived values for inter-issue, inter-article and intra-article navigation elements;
(3) identification of reference citations and formation of links to and from each article's reference section;
(4) division of each article into sections to enhance server performance and tracking of document usage;
(5) insertion of values into text and graphic templates featuring HTML extensions for placement of elements;
(6) generation of text, graphic, and table of contents files;
(7) formatting of article-specific CGI script calls.

After these initial phases were implemented, the feature set was extended:

- Short file names and logical path names were added to accommodate Macintosh, Windows, and UNIX file servers. The naming convention adopted for text file was strictly numerical, for graphics files was graphic type and number, and for path names was *year/ volume/ issue/ starting page number*.

- Large graphic files were to be subdivided into panels to enhance server performance. Key word variables were added to enable the tool to detect when multipart graphic elements were present.

- Raw text extracts of author, article, and graphics data were developed for the purpose of developing site-specific indices.

- The requirements for insertion of type-specific graphic elements and type-specific generation of full-text and abstract-only versions (for subscribers and non-subscribers, respectively) mandated the development of automated type-specific template alteration.

At this point, the strategy of a universal article template was nearly abandoned, but the complexities for the user in specifying public and private versions up to fifteen article types and the need to distinguish between the first and subsequent pages of an article, suggested that the dynamic alteration of a universal template would be preferable.

The collaboration with Detmer et al. [7] demonstrated the potential for automatically formatted retrieval of author-based bibliographies and article abstracts via calls to WebMedline, a WWW-based MEDLINE engine.

With the stabilization of the manuscript processing tool's feature set and its settings, the program was transferred to the publication offices for file processing. This process also included the manual extraction, labeling and placement of graphic source files for use by both the subscriber and non-subscriber versions of the publication.

RESULTS

As of this paper's submission, the processing tool has been in use by the NEJM since the beginning of 1996 and was employed in establishing their on-line service beginning March 21, 1996 (http://www.nejm.org). Other publications are still in the evaluation phase and have not been publicly deployed.

The initial deployment by NEJM was a public version which featured full content (text, images, tables, and references) of all article types (except original articles, brief reports, review articles, and special articles). The on-line issues were published simultaneously with the paper versions.

The option to link to external resources, specifically to retrieve author and abstract data from a MEDLINE database, was not enabled by NEJM in the initial deployment. As of this writing, several MEDLINE providers are considering providing a WWW API, but most of the publishers we are working with appear to be inclined towards development of site-specific indexing and retrieval engines in addition to MEDLINE.

In working with the different publishers, we have found that it was necessary to be involved not only with their editorial offices but also their business offices. They helped to identify the following commerce-related issues to extend the scope of the model and the software:

- document delivery capability

- subscriber management (registration and authentication)

- partially enabled public version generation

Document access is currently being tracked to detect patterns in reader utilization.

DISCUSSION

New means for the distribution of literature hold the most visible changes in store for authors, readers, publishers, libraries, health care organizations, academic institutions, and auxiliary entities. Relationships that have been stable for years, in some cases centuries, are undergoing re-evaluation, with involved parties reconsidering their positions in light of new possibilities and new requirements for development, formatting, maintenance, identification, and access [8].

Publishers of paper-based biomedical journals, in particular, will be affected, since they are at the heart of editing, review, and distribution of medical knowledge. Some feel that digital and paper publication are not mutually exclusive [9] and are probably complementary [10]. These changes bring opportunities for new functionality [11], new economic models and alternative information formats for journals and their readers [12].

A number of models for the future of scientific publishing have been proposed, ranging from self-publication [13] with public peer review, to centralized repositories [14]. These models raise the possibility that biomedical journals and their publishers will be rendered unrecognizable or nonexistent [15,16]. Self-publication entails problems with dependable location, indexing and retrieval, as well as quality of review [17]. The centralized model also has implications for the survival of established peer-review processes, and the ability of publications to serve as a forum for communities of medical practitioners and disease-specific research and educational activities.

While there are serious problems for institutional consumers, individual subscribers, and publishers with the current publishing model, there are many features of the current publishing model that should be preserved and can be extended with the adoption of a flexible non-proprietary architecture:

- automatically formatted calls to CGI scripts that can be extended to emerging distributed technologies
- micropayment transfer protocols that integrate financial, demographic, and authentication information
- a shared transaction model that rewards the source of a link as well as its target,
- the ability of intermediate content processing to implement targeted promotional material under the control of both the consumer and content provider,
- the ability of users to set price points for information provided by publications and the indexing or abstracting services that point to them.

Some of these features will necessitate the development of information brokering services, but all will contribute to a competitive, price-sensitive information market.

CONCLUSION

Biomedical publishers are embarking on a course that will eventually lead to a major role for on-line publishing, but only if sustainable business models can be developed. The central mission of all the parties connected with scholarly publication is the development and dispersal of knowledge. It is important to keep in mind that complex relationships among readers, authors, publishers, libraries, retailers and associated service providers have evolved to support this goal by pursuing their individual interests.

To succeed in the pursuit of this mission and their interests, authors, publishers, intermediaries, and information consumers need to focus on adapting their core competencies while achieving new points of balance in their relationships. Towards this end we see the primary tasks of biomedical publications as peer review, editing, and preparation of manuscripts. We also feel that it is important that they assume the additional responsibility for the maintenance and development of archival material.

As far as biomedical publications are concerned, the current distribution model rewards the latest information only, despite the fact that most of a publication's value lies in the accumulated body of knowledge that its archives represent. With the advent of the WWW, we believe that publishers will have a vital interest in developing and maintaining their archives and that intermediaries, like libraries and associated service providers, will have an interest in consistently applying established technologies to those data and developing new ways of employing it.

With time new relationships will be forged with other publications, indexing and retrieval services, and libraries to form a network of publications, search services, and information retrieval specialists. These relationships can result in sustained quality and improved currency of information, reduced costs, editorial independence, and accessibility [18].

Even with full cooperation between all parties, there are a significant technical challenges including the development of standards for:

(1) presentation of scholarly data in digital formats

(2) routing of requests for information

(3) authentication and demographic characterization of readers

(4) financial transactions to support the system

ACKNOWLEDGMENTS

This work was supported in part by grant LM07092 and contract NO1-LM-4-3512 from the National Library of Medicine and contract N66001-95-D-6007 from the Advanced Research Project Agency of the Department of Defense.

REFERENCES

1. Hawkins DT, et al. Forces Shaping the Electronic Publishing Industry of the 1990s. *Electronic Networking: Research, Applications & Policy* 1992;2:38-60.
2. Denning PJ, Rous B. The ACM electronic publishing plan. *Communications of the ACM* 1995; 38 (4): 97-103 (http://www.acm.org/pubs/ epub_plan.txt).
3. Lacroix EM, Backus JE, Lyon BJ. Service providers and users discover the Internet. *Bulletin of the Medical Library Association* 1994;82(4):412-8.
4. Friend F. 1997 Subscription Price Projections. *Newsletter On Serials Pricing Issues* 1996. Number 154 (http://sunsite.unc.edu/reference/ prices/1996/PRIC154.HTML#1 54).
5. Odlyzko AM. Tragic loss or good riddance? The impending demise of traditional scholarly journals. *Notices of the AMS* 1994 (gopher.cecm. sfu.ca/00/Resources/Epub/Other_studies/Journal _ demise_94_Complete).
6. Shareck EP, Greenes RA. Tools for an Open Model of Electronic Journal Publication. *Amercian Medical Informatics Association Spring Congress, 1996* (accepted for publication).
7. Detmer WM, Shortliffe EH. A model of clinical query management that supports integration of biomedical information over the World Wide Web. *Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care*, New Orleans, LA, 1995;898-902.
8. Samuelson, P. Copyright and Digital Libraries. *Commun. ACM* 1995;38:15-21.
9. Fox EA, Akscyn RM, Furuta RK, Leggett JJ. Digital Libraries: Introduction. *Commun. ACM* 1995;38:23-28.

10 Gellert GA. The death of biomedical journals. Electronic journals supplement their paper cousins [letter]. *BMJ*, 1995 Aug 19,311:507 (http://www.tecc.co.uk/bmj/archive/6991ed2.html).

11. Taubes G. Science Journals Go Wired. *Science* 1996; 271: 764-766 (http://science-mag.aaas.org/science/scripts/display/full/271/5250/764.html).
12. Hunter K. The Changing Business of Scholarly Publishing. *Journal of Library Administration* 1993;19:23-38.
13. Harnad S. Publicly retrievable FTP archives for esoteric science and scholarship: a subversive proposal 1994 (http://cogsci.soton.ac.uk/ ~harnad/intpub.html).
14. Quinn F. A Role For Libraries In Electronic Publication. *EJournal* 1994: 4 (2) (http://poe.acc.virginia.edu/~pm9k/libsci/quinn.html).
15. LaPorte RE, Marler E, Akazawa S, Sauer F, Gamboa C, Shenton C, et al The death of biomedical journals. *BMJ* 1995;310:1387-90.
16. Hitchcock S, Carr L, Hall W. A survey of STM online journals 1990-95: the calm before the storm (http://journals.ecs.soton.ac.uk/survey/ survey.html).
17. Kassirer JP, Angell M. The Internet and the Journal. *N Engl J Med* 1995; 332:1709-10.
18. Braude RM, Florance, Frisse M, Fuller S. The organization of the digital library. *Academic Medicine* 1995;70(4):286-91.