Malaysian Journal of Library & Information Science, Vol.1, no.2, December 1996: 1-11

AN OVERVIEW OF ELECTRONIC PUBLISHING

T.C. Ling
Mashkuri Haji Yaacob
K.K. Phang
Faculty of Computer Science & Information Technology
University of Malaya
50603 Kuala Lumpur, Malaysia
e-mail: tchaw@fsktm.um.edu.my
mashkuri@fsktm.um.edu.my
phang@fsktm.um.edu.my

ABSTRACT

Provides a definition on electronic publishing, as well as the various components including the necessary infrastructure, the technology required, the production processes and the means of distribution. Some brief accounts of exemplary models of electronic publishing are given and discussed. The main issues and problems which may be encountered when migrating from print to electronic publishing are also addressed. These include SGML or HTML marking-up of archive. Indicate transactions data received by the Malaysian Journal of Computer Science, a prototype electronic journal.

KEYWORDS: Online publishing; Electronic publishing; Communication channels; Scholarly journal publishing; Electronically assisted publishing; Standard electronic document format; Data only format; Display based format; Structured markup format; Hypertext Markup Language (HTML)l; Stand Generalised Markup Language (SGML).

INTRODUCTION

Johann Gutenberg started a publishing revolution five hundred years ago with the printing press. Currently and going into the next century, the World Wide Web (WWW) and Internet are without doubt introducing a new era in which the same kind of impact, if not greater, would be seen on the way we store, promote and distribute (or transmit) information. With the increasing popularity of the Internet and WWW even among the Third World countries like Malaysia, the age of paperless publishing and global interaction is just not too far away. At this age of repeated scientific breakthroughs, who can still truthfully say that they are immune from databases that are electronically retrievable and the magnetic digitisation of data? As Drott (1995) contends "new communication channels have, in fact, been emerging for decades." Electronic publishing is just another new communication channel, and in five years time, it will be as common as words on paper.

Albeit a viable and acceptable standard of electronic publishing has only recently been made possible through improved

software capabilities and increased ubiquity and networking of PCs in both formal institutions and the general community, the idea of electronic publishing is not something new. It was introduced much earlier in the mid-seventies. With computer technologies progressing by leaps and bounds, it is not difficult to foresee a widespread use of electronic products everywhere, even in Malaysia. The journal publishing fraternity which have the closest link with academicians and intellectuals should be the first to take a stride into the electronic era.

DEFINITION

Although it is difficult to draw a clear line on what constitutes electronic publishing, we can be certain that computer technology will always play a part in the publishing process, so much so that we can call it an *electronically assisted publishing*. Hawkins et al (1994) have defined electronic publishing as "the use of electronic media, computers and telecommunications to deliver information to users in electronic form or from electronic sources". However, it should be emphasised that electronic publishing is concerned with the delivery of information by and between computers only, without the involvement of paper in transferring information between end users. This distinction should be made clear so as to differentiate from desktop publishing which is concerned with the preparation of information on computers, but often for final delivery on paper. Thus, electronic publication will cover electronic bulletin boards, online newspapers, books, email, electronic journals, as well as real time downloadable information services, software and even long distance conferencing. It should be added that electronic publishing also involves the use of floppy disks and CD-ROMs as an electronic mode of transfer. Owing to the very nature and coverage of the WWW and the Internet, the main form of electronic delivery which will be discussed will be the *online* mode. Thus electronic journal publishing is publishing a journal electronically and distributing it without using paper as a medium to disseminate information, and this paper will focus on this aspect.

The potential benefits of publishing electronically are so compelling that few would be foolish enough to dismiss its potential impact on the publishing industry, both at present and in the future. Electronic publication enjoys enhanced peer participation, higher interactivity, faster review, quicker navigational design, lower production costs in the long run and faster access to the information.

INTEGRATING OF ELECTRONIC PUBLISHING

Infrastructure

The infrastructure is the basic facilities and installations needed for the functioning of the whole digital publishing of materials. The publishers need to set up an in-house server or rent server space from an internet service provider, a few microcomputers for digitisation of their materials and

products, for tagging with hypertext and graphic production, a backup system, and many other auxiliary peripherals. The clients will need to have appropriate viewing equipments and adequate access to personal computers. Undoubtedly a market research needs to be carried out before starting on this venture.

Technology

An appropriate electronic viewing technology is essential so that the electronic publication can be made easily available to the user. This issue, however, is relatively simple to resolve. The technology involved is usually not expensive. The Internet browsers such as Netscape Navigator and Microsoft's Internet Explorer can be easily downloaded free of charge or with a minimum fee. Special consideration and at an additional cost, as expected, should be given to the development of suitable search and retrieval software and facilities, and an improved ordering service. The technical aspect of the software and hardware setup can be easily resolved as long as there is no financial constraints. The most appropriate type of software and hardware platform should also be determined by considering the user factor and the production and distribution requirements.

Production

It must be borne in mind that producing the original for electronic documents is no cheaper than producing the original for a paper publishing project. A lot of skilled labour will be needed and usually the cost will run somewhat higher for electronic publishing than for paper, because the industry has less experience and the tools are not mature. Yet, if we are thinking about a multimedia-enhanced editions, about ephemeral information that come and go before they can be put on paper, repeated updates and reprints of materials, there is nowhere else to look into but the digital world.

When switching to an electronic format, besides the usual layout of the document, we need to take into account the reliability of digital storage, the search ability of indexed archives, the transmit ability of electronic files, and the cost efficiency of electronic duplication. However, for a single electronic publishing format to accomplish all these objectives is almost impossible, because different needs and priorities demand trade-offs. Thus at least a dozen standard and proprietary electronic document formats have to be considered.

Basically, there are five distinct document format to choose from: data-only, display-based, structured markup, styled, and proprietary.

Data-only format is the simplest where ASCII or text document is the usual format. They can be produced by just about any word processor, spreadsheet, or database management software, and almost any kind of computer will be able to read them. These documents are very small and can be transmitted quickly, though many layout features will have to be sacrificed.

In using the display-based formats, there are more control on layout and on-screen appearance. These formats can be produced using products like Adobe Acrobat, Common Ground Software's Common Ground, or Novell's Envoy. Products such as Adobe Acrobat produces PDF file, which uses image compression and font-simulation to provide the recipient with a more or less replica of the original even if the recipient lacks the font used to create the document.

Structured markup format can now be found everywhere, because HTML (Hypertext Markup Language), a subset of SGML (Standard Generalized Markup Language), is presently the only format we can expect all Web browsers to support. Currently, just about every word processor has an option to output files in HTML format. This includes Corel WordPerfect, Lotus WordPro, and Microsoft Word. The new WordPro Web Builder even provides everything users need to create, publish and manage internet documents from within WordPro. However, for more advanced features, a web editor such as Softquad's HoTMetal Pro can be used.

Styled format is the WYSIWYG (What You See Is What You Get) format we produced using a word processor or a page-layout program. One example is the Microsoft's RTF (Rich Text Format), which is an open format with a published description. This is widely supported by almost every major word processor such as WordPerfect.

Many proprietary formats are designed for or have been adapted for electronic publishing. Adobe PageMaker, for example, can present files designed as paper documents as on-screen slide shows. There are more extensive and advanced document management and presentation systems, such as Personal Librarian from Personal Library Software and Knowledge Retrieval System from Knowledge-Set, which are particularly geared for retrieving documents from large collections on CD or local networks.

Distribution

Depending on the type of format chosen and various other factors such as the size of the publication and the niche market, the level of product interactivity, speed, and the facilities available, the choice of distribution varies. Publishers can choose between CD-ROM, client-server system (Gopher, WWW), mailing list, online database, using FTP or attaching file to e-mail. By carefully assessing the multiplicity of resources at hand and critically evaluating the needs and trend of the market, everyone should be able to make a sensible decision on the ways of production and choices of distribution.

THE 'HOW TO' OF ELECTRONIC PUBLISHIING

To publish electronically, of which our main concern will be through the Internet, a publisher has two choices to make: by either becoming a content provider only, i.e. by relying on their commissioning and editorial expertise and their copyright capital and leaving distribution and deli-

very to others, or taking up too the distribution and delivery through the WWW themselves. Although the latter choice is much more compelling as the Internet provides a canvas for a new market and heightened competitiveness with the use of a new trade model, many are still undecided as to whether they should compete in this new arena or not, while many others are eager and anxious but lack the level of investment in the technology to create and run their own web sites. However everyone is urged to take a step into this new digital world.

A practical move in migrating to publishing electronically is to get into the digital process. This can be done by SGML encoding and tagging their copy-right material with a content label. The speed of the development of interface technology has made this a feasible and practicable approach, but an added cost and additional time delay in doing it is inevitable, as experts and specialists have to be employed to carry out the encoding and tagging process using SGML or HTML. And usually quite a large work force will be needed during the transitional period as all documents and material need to be tagged or marked up if they are to be made available to a host of different machines (Mac, IBM, UNIX, etc.) each running on their own operating systems (Mac OS, Sparc station, Windows NT, etc.). Meanwhile, the advantage of taking all this trouble is that an intermediate, standard format for archiving digital material enables publishers to design and repackage content to respond to new interfaces, which will ensure a speedy move to the web when both the technology and market are mature.

A structural change within the publishing organization is also a necessity besides a move to the digital process. However, many have taken the easy way out by setting up a small electronic publishing advisory team who act as a service providing department to the members in the main publishing business on how to create an electronic version of the original printed version. This without doubt, causes little changes and disturbance to the existing organizational structure, but it fails to meet the actual needs of electronic production and more problems will arise as the issues over the design ownership, responsibility and accounting procedures remain unresolved. In this respect, the Oxford University Press can act as a successful model. Of the four divisions, two, namely the Arts division and Reference division lend themselves to the CD-ROM considering their predominant lexicographic needs, while the science and journals division are responsible for the WWW site, electronic journals and the security of the gateway. The English Language Teaching (ELT) division has an electronic publishing department which acts collaboratively with other department within the Division, creating electronic materials linked to existing print titles, but remaining a profit centre within its own right. From this model it can be seen that the electronic publishing team needs to take part directly and more closely in the electronic publishing process so the issues arising due to publishing the printed original electronically can be dealt with more effectively. This model also

minimises the disturbance to the existing general organizational structure by diversifying the responsibility for electronic publishing (Weedon, 1996).

One consoling point concerning the shift to electronic publishing or Internet publishing is that the number of staff needed in order to maintain the publication seems to be quite low. A quick survey of some prominent examples, reveals the current situation: Penguin USA has five staff out of an employment of one thousand working on their electronic publications and Reed of Australia has four. Taking the first step towards electronic publishing could be difficult, but from the above examples it is obvious that to keep the electronic publishing going is a much easier job than publishing on paper. Addison-Wesley has one full-time editor on electronic production compared to five in paper publications. They operate in a niche market of computer books and find that their complementary electronic deli-very method is synergistic with what they are selling (Weedon, 1996). Besides the above specific models, other models can be thought of with multimedia capabilities, a global audience and an interactive interface of the Internet and WWW in view.

SGML or HTML

Standard Generalised Markup Language (SGML) is a standardised markup language devised to ensure that any software and hardware combination will be able to read the information tagged using SGML. HTML is actually a subset of SGML, created with only a portion of the features offered by SGML, but with relative simplicity so that it can be used on the World Wide Web without much hassle. However, in migrating to the digital process as mentioned above, we need to make a choice between these two markup languages.

Most of the browsers operating in the Web are HTML interpreters. With the introduction of the latest HTML 3.2, a document marked-up with HTML will be even more feature-bound. But SGML can offer many other functions that will not be good enough or cannot be achieved by using HTML. If we are dealing with a huge file such as a journal complete with a search tool and perhaps needing to be updated frequently, we are better off tagging it using SGML. After all, changing from SGML to HTML is much easier than the reverse. If the data were to be stored for ten years in digital form, tagging it using SGML will be less likely to need revision.

The problem now is that SGML browser is hardly available in the WWW. If an SGML document were to be made available to the client, its browser will have to be provided or the user will have to be directed to a certain site to download the software. Nevertheless, where electronic publishing is concerned, SGML seems to be the better choice, given the availability of SGML experts.

PROBLEMS TO BE OVERCOME

With the plethora of alluring features that can be quite easily offered through the

Internet, it seems that the move from print to screen has become almost inevitable. In fact, with the employment of the state-of-the-art Java and Shockwave, anything – images, sound, video – that can be digitised can be transmitted in the virtual world with comparative ease. No matter how attractive the offer may be, there are still caveats that need to be taken into account before anyone can take a jump onto the bandwagon of electronic publishing. Many of these have actually been mentioned elsewhere as seen appropriate, but the following are four main areas pertaining to the financial, pricing, online subscription and copyright aspects which must be taken into serious consideration.

Financial

Although it can be argued that the production, duplication, distribution, and storage costs of electronic publishing are convincingly much cheaper compared to that of paper publishing, all these cost-benefit cases do not cover the whole story of electronic publishing. The notion that increased volumes of print create economies of scale and thereby reducing the cost per unit for a given publication is very much true in the print industry, but this applies even more accurately to electronic publishing. In electronic publishing, the distribution cost hardly rises with the increase in the size of the item distributed. A source of solace for those who are going into electronic publishing, but there are other things to be kept in mind. We need to recognize that a publisher can only enjoy all these benefits on the conditions that they already have a team of electronic publishing professionals who are well equipped with advanced automated markup software and the necessary high-end hardware for efficient output of data and material electronically. And these are where the money will have to come first.

It should be realized that it costs just as much, and usually more, to create, capture, enhance and format data for electronic publication as it does for print-based publishing. Not forgetting to add that in order to make the data available for all platforms, it needs to be tagged using SGML or HTML (if Internet publishing is the idea). Furthermore, if full multimedia capability is intended, the inclusion of Java script and the use of Shockwave would seem unavoidable, which will usually double the work of print-based publishing. Other than these, the issues of cost of computers, salaries and other indirect costs will have to be addressed.

In a changing industry such as electronic publishing, there is a continuing research and development and equipment over-head. For those who want to survive in the bitter competition of the electronic world, they need to be able to assess the viability of various software and hardware platforms, create prototypes and try out all the competitor's products, and need ever more sophisticated machines to do so. As the volumes of data grow and the customers become more sophisticated in their requirements,

then so will the aspects of digital products like the searching techniques and customer profiling be updated.

Pricing

Compared to paper publishing, electronic publishing today lacks a generally accepted pricing pattern. This is due to a nagging problem for the publishers of electronic products: the digital products have the inherent tendency to compromise the revenue from their paper counterpart that are running parallel, for the reasons that have been discussed above. This is common in any transitional phase where both paper and electronic products are being produced, adding a significant amount to the cost base. Many will try to recoup this loss in revenue from the newer electronic product by putting a premium on the price of the printed journal for online access, or adding 5% to the subscription price of the online version, without any consideration on the market expectation on pricing. It seems that market expectation are that digital technology are cheap and efficient, thus cheaper digital products are sure to follow, offering better access to a wider range of data for the same or lower cost. Adding to the complication is the fact that there is the existence of free information on the Internet. When we come to this, it is a balancing act which everyone will have to learn through experience and by considering the various influencing factors around them, such as the market response.

Online Subscription

Subscription is never a problem for a paper product, using the same old mailing system as the mechanism for dispatch. With the electronic version, because the publisher can be more selective in what sorts or fields of information to offer to their clients, things become more complicated. A client can usually select only the specific articles or information they need with the availability of full text direct search tools (whereas paper journals, in the interests of balanced issues tend to include something of interest to the whole constituency, with the result that few readers will read the whole of an issue's contents, relevant theme issues aside). However, this poses another problem as anything that can be seen on the computer screen can be easily downloaded to the hard disk. As a result the actual article cannot be provided for viewing. What can be done is to set up a teaser site where only partial information of a journal is given, and this will usually be the abstract of each article.

The Malaysian Journal of Computer Science

The online version of the *Malaysian Journal of Computer Science* (*MJCS*) published by the Faculty of Computer Science, University of Malaya, [http: //mjcs.fsktm.um.edu.my] provides a good example for this model. At the *MJCS* homepage, a brief description of the journal is given: its objectives, scope, plus the manner of subscription and pricing. This will give any potential client a very good idea of what it is all about. In addition, it is complete with a database of the abstracts of all the papers published by *MJCS*. The database has been

designed to give the user quick and easy access to the required information using a keyword search engine. Appendix A and B show some statistics concerning the transaction patterns of the online version of the *Malaysian Journal of Computer Science*.

Copyright

Owing to the freedom we enjoy on the WWW and its lack of structure, most publishers will have great difficulty in assuring the authors that they will get due recognition and citation when publishing electronically online. It needs to be asserted that copyright is an obstacle in the path of publishers who are after mass-market publications. Although the problem will be alleviated by publishing a journal (perhaps a quarterly) in a proprietary CD-ROM, we still cannot ensure that this will be free from unauthorized duplication as the technology to duplicate a CD-ROM is becoming cheaper and more commonplace everyday.

CONCLUSION

From the above discussion, it is clear that many aspects of the electronic publishing are not mature enough for everyone to usher in the digital age into their publishing house. One thing, however, is for certain. Whenever there is an easier and more efficient way to present the highly valued information to the ever advancing information-craving market, the method will be adopted. It is just a matter of time.

Publishers will have to adapt and learn fast and heave themselves into constant change to keep up with the pace of IT development. If a publisher can act quick enough to recognize and exploit the opportunities made possible by the technology, he will be the one who can thrive well into the next century where the computer will be an indispensable tool for all walks of life. All publishers, especially the journal publishers who actually occupy a space in the circle of academia, should bear in mind that it will be much less intimidating to join the race at the very beginning before the competition gets tough, as many late starters will find that they are already too far behind to ever catch up with the others.

REFERENCES

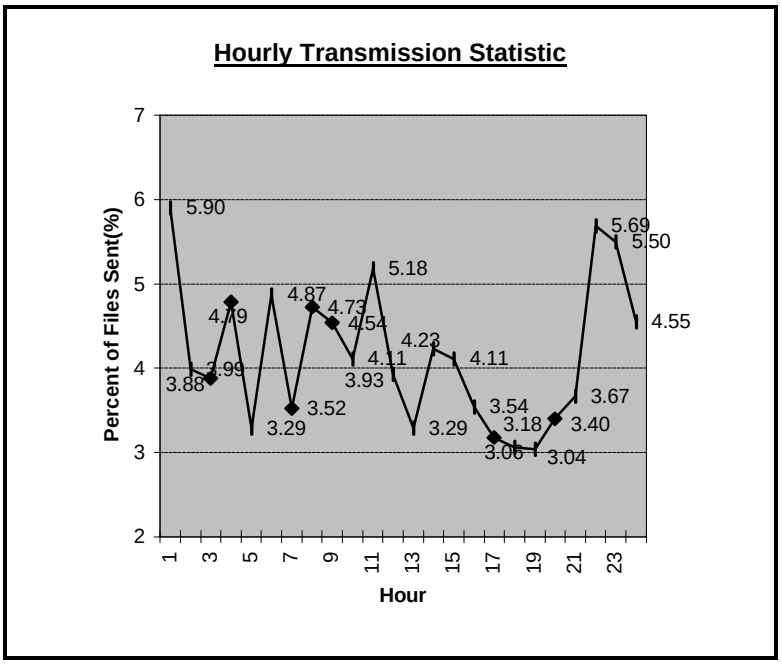
Drott, D.C.M. 1995. Reexamining the role of conference papers in scholarly communication. *Journal of the American Society for Information Science*, 46 no. 4: 299-305.

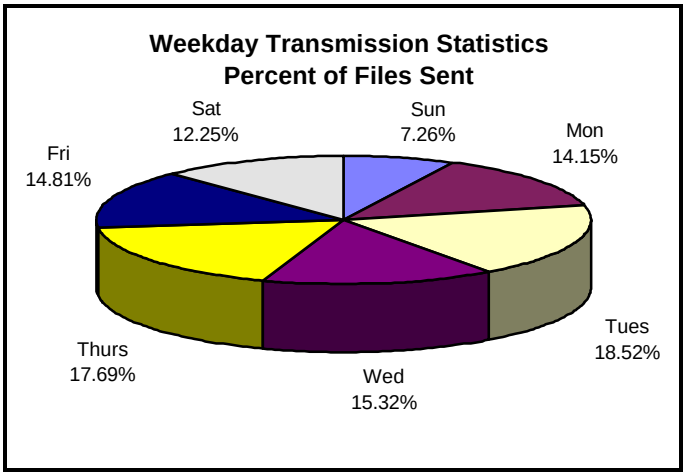
Hawkins, D., F. Smith, B. Dietlrin, E. Joseph and R. Rindfuss. 1994. Forces Shaping the Electronic Publishing Industry, In: *The Electronic Publishing Business and Its Market*, Blunden & Blunden, PIRA.

Weedon, A. 1996. The Electronic book trade and the Internet: BILETA '96 Conference Proceedings, 3. *Journal of Information, Law and Technology*. <http://elj.Warwick.ac.uk/elj/jilt/bileta/1996/3weedon/>

APPENDIX A

Daily Transactions of MJCS	Total
Average Files Transmitted Daily	243
Average Bytes Transmitted Daily	565548





APPENDIX B

Percentage of Files Transferred by Client Domain

Domain Name	Files Sent %	Domain Name	Files Sent %
Adv. Proj. Research Agency	0.06	Malaysia	14.02
Argentina	0.22	Mexico	0.34
Australia	2.53	Morocco	0.03
Austria	0.10	Netherlands	0.61
Bahrain	0.01	Network	6.23
Barbados	0.01	New Zealand	0.06
Belgium	0.08	Norway	0.18
Brazil	0.69	Peru	0.03
Canada	1.32	Philippines	0.03
Chile	0.28	Poland	0.18
Costa Rica	0.07	Portugal	0.07
Croatia (Hrvatska)	0.04	Russian Federation	0.01
Czech Republic	0.07	Singapore	0.83
Denmark	0.22	South Africa	0.05
Ecuador	0.03	Spain	0.31
Egypt	0.08	Sweden	0.74
Finland	0.81	Switzerland	0.14
France	0.78	Taiwan, Province of China	0.18
Germany	0.87	Thailand	0.22
Greece	0.13	Turkey	0.07
Hong Kong	0.08	United Arab Emirates	0.04
Iceland	0.02	United Kingdom	5.26
India	0.07	United States of America	0.33
Indonesia	0.10	US Commercial	16.28
Ireland	0.16	US Education	11.40
Israel	0.08	US Government	0.17
Italy	0.39	US Military	0.13
Japan	0.80	US Organisations	0.13
Korea, Republic of	0.93	Venezuela	0.04
Latvia	0.03	Unresolved	30.87