United States Educational Foundation in India,
DRTC/Indian Statistical Institute,
DLIS/University of Mysore
Joint Workshop on Digital Libraries
12th – 16th March, 2001

Paper: E

Electronic Publishing: An Overview

Jagdish Arora
Head, Computer Applications Division
Central Library, IIT Dlehi

1 Introduction

The history of mankind has already witnessed three revolutions and it is on the threshold of fourth. The first revolution took place a few thousand years ago when language first emerged. The second cognitive revolution was the advent of writing, tens of thousands of years ago. The spoken words allowed oral depiction of thoughts and written language made it possible to preserve them. The third revolution took place in our own millennium with the invention of printing press in 1500. The post Gutenberg's era witnessed complete and absolute control of printing technology and printed artifacts as a media of scholarly communication over the scientific and literary world. The Internet, world-wide web and the electronic publishing, in particular, represent the fourth revolution that seems to have humbled monopoly of printed artifacts as sole media of scholarly communication. The revolution has just began, and is going under the process of adaptation. Authors, publishers, users and librarians are only just beginning to take advantage of the potentials of electronic media.

The emergence of Internet, particularly, the world-wide web (WWW) as a new media of information storage and delivery would have lasting impact on the publishing and information delivery system in the 21st century. The Internet and Web technology would not only change the way information is stored, retrieved, communicated and broadcasted, it would also change the way publishers, publishing industry and scientists functioned in the era predominately guided by the printing technology (Cox, 1997).

Like most revolutions, the players involved in it find themselves in a challenging situation, each with redefined roles to play. The electronic revolution is causing all the players, i.e. scholars, librarians, publishers and vendors to question their traditional roles in the process of publication, sale, dissemination, storage and access. The authors, guided by the doctrine "publish or perish", a legacy of Gutenberg's printing technology, continue to favour printed journals in preference to the "electronic only journals". The printed words are still seen as authoritative medium and format for publication of peer reviewed research. Tenure, promotion and the research grants depend on publishing record of applicants. However, the trends may change as the electronic versions continue to improve upon advantages that they offer over their printed counter parts. The librarians would have to discover themselves in a renewed role as facilitators of information gateways and as negotiators for subscription to products and services of electronic publishing in a consortia mode. The publisher would continue to involve in the process of refinement, packaging, presentation and marketing of intellectual property while media of communication shift from the printed artifacts to electronic publishing and access. The subscription agents may have to take up the role of aggregators or facilitators of a large number of journals published by different publishers. EBSCOhost and Blackwell's E-Journal Navigator have already redefined their role in the changing scenario. Lastly, user's preference depends on several factors including demographics. An ACS study of a group of University of California graduate Chemistry students revealed that "if they can't get

information via their computers, which unlike the library, are available 24 hours a day, they may not bother to get it".

The Internet has long been a medium of experimenting electronic publishing. Increasing numbers of publishers are using the Internet as a global way to offer their publications to the international community of scientists and technologists. The technology provides an unparalleled media for delivery of information with greater speed and economy. The electronic delivery of journals also results in elimination of paper, storage and transportation costs and the ability to handle complex data tables, moving pictures, sound, images and video clips. Further, unlike sequential design of printed papers, the web technology makes if possible for the publishers to design things that are not possible in print such as interactive hyperlinks to related resources, links to full range of multimedia, etc.

Recent developments in new information technology and products of electronic publishing has enforced shifts in ideological and cultural practices of traditional libraries. Side Bar 1 lists major shift in technology and its applications that led to appearance of multitude of products and services generated by the process of electronic publishing. Side Bar 2 lists cultural and ideological shifts consequent to major technological upheaval. The librarians, with their concern to provide increased access to the resources within the limits of their diminishing budgets have started experimenting with the products and services of electronic publishing.

The article delves on the evolution of electronic publishing, their promising features and unresolved issues. Traditional publishing processes and transition from print to electronic media are discussed at length. Elements of value additions that an e-journal receives and the cost involved in it are enumerated. The article outlines economics of electronic journals, prevailing pricing models and formats and media types used for presenting electronic documents on the web. Lastly, the article enumerates categories of players in the field of electronic publishing and models that have been proposed as alternatives to the exisiting publishing scenario. Of various products and services of electronic publishing, the libraries are more concerned with electronic journals as the most expensive, intriguing and highly-demanded items in their libraries. This article, therefore, concentrates and elaborates more on electronic journals.

2 Evolution of electronic publishing

Evolution of electronic publishing can be traced back to the 1970s when computers were first used to assist printing of abstracting and indexing services. It has since evolved along the technological ladder for the past thirty years. The databases came online first in the late 1960s and Dialog became the first commercial online service in 1972. By 1975 there were 300 publicly available online databases. Creation and remote accessibility of online bibliographic databases is considered as an important landmark in electronic publishing. Sophisticated online databases were built during the 1970s and the 1980s using state-of-the-art technology of distributed database management system linking different remote systems using data files generated in the

process of electronic phototypesetting of printed abstracting and indexing services and other primary journals. As such, online hosts like Dialog and STN were not only offering online databases but also full-text online journals for several years, although as simple ASCII or text files without graphics and pictures. By 1988, 3,893 online databases were available from 1,723 database producers and 576 online services. In 1989, there were almost 1,700 full-text sources in sixteen online systems. Availability of CD ROM in the late 1980s as a media with high storage capacity, longevity, and ease of transportation triggered production of several CD ROM information products which were earlier available through online vendors or as conventional abstracting and indexing services in printed format. Moreover, several full-text databases also started appearing in the late 1980s and the early 1990s launching the beginning of the digital era. Some of the important full-text digital collections available on CD ROM include: ADONIS, IEEE / IEE Electronic Library (IEL), ABI/INFO, UMI's International Business Database, UMI's General Reference Periodicals, Espace World, US Patents, etc.

Online Public Access Catalogues (OPAC) as well as online bibliographic databases can only be used to find bibliographic details without their contents in full-text. This limitation led to the demand for electronic publishing as a tool to develop full-text databases in digital form. The 1990s brought in a true revolution in the digital library system with the advent of WWW (World Wide Web). The WWW offered web server at the server-end and web browser at the client-end for all prevalent platforms. The availability of ready-to-use, publicly available, user-friendly graphical web browser for all prevalent platforms eliminates the need of extensive support and user's training. Standard WWW clients such as Netscape Navigator and Internet Explorer are being upgraded regularly for added functionality such as e-mail client, support for JAVA and Active X and the ability to view important document formats without having to install plug-ins for them. HTML, the *de facto* language of the web, is an extremely simple yet powerful tool for the presentation of products of electronic publishing. The Internet and associated technologies, made it possible for web-based electronic information products to include multimedia objects such as text, image, audio and video. These technologies thus brought in the graphical components to electronic publishing which was earlier missing.

Digital document imaging system, which employs computer hardware and software to scan and store images of documents in digitized formats, was evolved in the 1980s to overcome the limitation of text storage and retrieval systems which could only store textual information. The document imaging system, however, found its real application with increase in storage capacity and decrease in prices of evolved storage devices coupled with the power of Web technology to handle multimedia format.

There has thus been a steady move up the technological scale for IT applications in the libraries from online bibliographic databases and OPAC to low-end electronic publications available as ASCII files, to being organized and searchable on gophers (1992), and to being tagged and graphically viewable full-text systems on World Wide Web sites (1994).

3 Electronic publishing: products and services

Electronic publishing is manifesting itself in a wide range of products and services, although most of them emulate the traditional publishing while others are revolutionary in their design and approach. While the present trend to imitate and emulate the traditional models of scholarly communication may continue for some time, eventually the capabilities added by the new media would be used in more innovative ways. Some of the important products and services of electronic publishing are:

3.1 Electronic journals

Electronic journals (or e-journals) may be defined very broadly as any journal, magazine, e-zine, webzine, newsletter or type of electronic serial publication which is available over the Internet and can be accessed using different technologies such as WWW, Gopher, ftp, telnet, e-mail or listserv. Internet-based electronic journals started to appear in the beginning of 1990. These journals were mostly delivered as an attachment to e-mail while their back issues were mounted on anonymous ftp sites and users were required to download them from this site. The libraries and information centres made them accessible through their gopher site. 1995 witnessed the peak of Gopher technology which then dropped suddenly and dramatically by 1997. With the advent of WWW technology in 1993, electronic publishing became more than a novelty. The web as a means of delivery of electronic information has grown steadily since then.

With the advent of CD ROM technology as optical storage media in the mid-80s, several electronic journals started appearing on CD ROM. The first major development in this direction were projects experimenting with electronic equivalents of printed journals. One of the oldest examples is ADONIS where images of articles published in printed journals are distributed on CD ROM. Still older examples are full text databases run by the major host organisations. Online hosts like Dialog and STN were not only offering online databases but also full-text online journals for past several years, although as simple ASCII or text files without graphics and pictures. In 1989, there were almost 1,700 full-text sources in sixteen online systems. All of these projects involve journals and all of them are by definition electronic, but these journals were not truly electronic, they can at best be described as electronic versions of printed journals.

Number of electronic journals has grown in dramatic proportion from less than 10 in 1989 to more than 8,500 in April, 2000. The 37th edition of the Ulrich's International Periodical Directory (1999) reports that of total 1,57,173 serials listed in the Directory, 10,332 are available exclusively online or in addition to its paper counterpart and 3,451 on CD ROM. Table 1 below provides number of electronic journals available on the Internet in Science and Technology.

	1991	1992	1993	1994	1995	1996	1997	2000 (April)
E. Journals	27	36	45	181	306	1093	2459	
E. Newsletters	83	97	195	262	369	596	955	
Total	**110**	**133**	**240**	**443**	**675**	**1689**	**3414**	**8,500**

Table 1: Growth of Academic Electronic Journals and Newsletters (ARL)

3.2 Electronic books

Borchers (Borchers, J.O., 1999) defines an eBook as a portable hardware and software system that can display large quantity of readable textual information to the user and let the user navigate through this information. An eBook is digital reading material that a user can view on a desktop or notebook personal computer, or on a dedicated, portable device with a large storage capacity (1,500 to 50,000 pages) and the ability to download new titles via a network connection. While the proponents of e-books consider them as an advancement of Gutenberg era, skeptics, on the other hand, maintain that consumers will never choose to curl up with an eBook. The reader hardware is expensive, e-titles cost about the same as their print counterparts, and ink and paper are still easier to read and handle. More and more traditional book publishers, as well as those catering to the professional and business communities, are seeing the potential of digital publications and are working to ensure that they enjoy a share in the market's growth. In fact, analysts expect the market for eBook titles and other electronic documents to exceed US$2 billion in the next few years.

Electronic books is not a new concept. Project Gutenberg started digitising public-domain texts for download in 1992. But they couldn't overcome the problems of reading text on-screen, glare, bad layout, and variations in visual fidelity and richness. Nor do they offer the rich range of content and visual presentation that consumers want. Today, companies such as Adobe, MightyWords, Everybook, and Gemstar are in the business of developing technology that would hopefully help in transforming eBooks into a medium as friendly and ubiquitous as their print counterparts. Even Microsoft has joined the fray with its Reader 1.0, a proprietary eBooks reading application. Adobe CoolType is yet another recent development that dramatically improves on-screen text resolution of digital content. CoolType's cross-platform, cross-font compatibility would be a reassurance for the eBook consumers that they would finally be able to enjoy clearer, crisper type and a reading experience that is closer to the clarity of printed type. Time Warner Trade Publishing has recently launch iPublish.com, an online publishing venture to unite authors and readers in the digital marketplace. Reader-members will be able to download eBooks, or post submissions for possible publication in print and digital formats. Time Warner also has an equity stake in Bookface.com, an online bookstore of downloadable eBooks. Not to be outdone Random House's new AtRandom enterprise will make an initial 20 full-length novels, biographies and self-help books available for print-on-demand or as online digital books. Random House has also partnered with Xlibris.com in an effort to entice would-be authors to publish their works online. Original works are placed in the Xlibris online bookstore and made available through Borders.com in print or

electronic format. iUniverse.com and barnesandnoble.com are two other players in e-books and print-on-demand market.

3.3 Electronic encyclopaedia

Availability of enormous storage space on the CD ROM coupled with sophisticated search software witnessed appearance on several encyclopaedias on CD ROM. Later, web version of these encyclopaedias became available as important reference tools on the web. Encyclopaedia Britannica can be treated as an example of a formerly flourishing business that fell into trouble in just a few years by neglecting electronic media. Encyclopaedia Britannica has since collapsed, and was sold to Jacob Safra, who is investing additional funds to cover losses and revamp the business. The expensive sales force has been dismissed, and while print versions can still be purchased from bookstores, the focus is on electronic products. This collapse occurred even though Encyclopaedia Britannica had more than two centuries of tradition behind it, and was by far the most scholarly and best known of the English-language encyclopaedia.

Encyclopaedia Britannica was vulnerable largely because it had an enormously bloated cost structure. The $1,500 to $2,500 that purchasers paid for each set included a couple of hundred dollars for the printing, binding, and distribution. Most of the rest was for the sales force and general administrative overhead. The editorial content apparently amounted to well under ten percent of the total price. While Encyclopaedia Britannica was still sold at US $1,500.00 - $2,500.00, the market was flooding with $50 CD-ROM encyclopaedias. Although they did not have the same quality of content, nor nicely printed volumes, but they did have superior searchability, portability, and an irresistible price. After some abortive attempts to sell first $1,200, then $300 CD-ROMs, Encyclopaedia Britannica is now offering its CD-ROMs for $125 or even less. Web versions of several important encyclopaedias are available over the Internet (Odlyzko, 1999).

3.4 Online electronic databases

Availability of CD ROM, and more recently DVD ROM, as a media with high-storage capacity, longevity and ease of transportation, triggered production of several CD ROM-based information products including several bibliographic databases which were earlier available only through online vendors or as abstracting and indexing services in printed format. Thousands of CD ROM databases are currently available from multitude of CD ROM producers including Silver Platter that alone produces more than 250 CD ROM information products. Moreover, several full-text databases also started appearing in the late 1980s and early 1990s launching the beginning of a new digital era. Some of the important full-text digital collections available on CD ROM include: ADONIS, IEEE / IEE Electronic Library (IEL), ABI/INFO, UMI's International Business Database, UMI's General Reference Periodicals, Espace World, US Patents, etc. CD ROM networking technology is now available for providing web-based simultaneous access to CD ROM databases on the LAN (Local Area Network)

as well as on WAN (Wide Area Network). More evolved technology allow caching and integration of entire contents of CD ROMs / DVD ROMs on to a server, which, in turn, provides web-based simultaneous and faster access to the information contents of CD ROMs / DVD ROMs. Most of the online databases that were earlier available on CD ROM are also now available on the web with added functionality and features. These databases including Medline (several versions), AGRICOLA and ERIC.

3.5 Online courseware

Amongst electronic resources created exclusively for the web, imbibing all features and facilities offered by the new technology, include web-based educational tutorials called "online courseware". The online courseware are proliferating the web as a strong contender for distance education. Telecampus, Canada (*www.telecampus.edu/*) lists more than 12,000 online courseware available on the web. Online courseware are in the forefront of technological, multimedia and instructional innovation that provide a higher degree of interactivity and flexibility to the users. These courseware are comprehensive resource kits that offer the benefit of self-pace and are focused on developing practical skills that can be applied immediately. Institutions of higher learning, especially distance and continuing education departments are actively supporting and contributing to the development and implementation of computer-assisted instructions and multimedia courseware. (Chakravarti, 1999).

3.6 Electronic preprints

Ginsparg preprint archive, started in 1991, has become the fundamental communication method for a growing roster of fields, starting with theoretical high-energy physics, later spreading to other areas of physics, and now also to computer science and mathematics. It is a sterling example of how technology can lead to a sudden, profound, and beneficial transformation. Yet in 1998, this archive still processed only 24,000 submissions, which is substantial, but small compared to perhaps two million papers in all science, technology, and medicine areas. (Odlyzko, 1999) The attractions of the archive are great. It transforms the mode of operation of any community of scholars that embraces it, and the transition is invariably one way, as not a single group has abandoned it. It quickly becomes the dominant mode of communication inside any group that embraces it. However, in spite of extensive publicity, it has not yet swept scholarly communication. It appears that there were special cultural factors that led to the quick adoption of the archive by Ginsparg's own community of theoretical high-energy physicists (primarily the reliance on massive mailings of preprints), and it has been a struggle for pioneers in other areas to duplicate the process. There are still many areas (especially in chemistry and medicine) where not just preprint archives, but preprints themselves, are rare, and in which prestigious journals get away with policies

that forbid any formal consideration of a paper that has been circulated in preprint form. A few examples of preprint servers in other disciplines are:

UK e-Print archive mirror	*http://xxx.soton.ac.uk/*
Open Archives Initiative	*http://www.openarchives.org/*
PuMed Central	*http://www.pumedcentral.nih.gov/*
American Mathematical Society Preprint Server	*http://www.ams.org/preprints/*
CERN Preprint Server	*http://preprints.cern.ch*
Chemical Physics Preprint Database	*http://www.chem.brown.edu/chem-ph.html*
Chemistry Preprint Server	*http://www.chemweb.com/preprint*
Economics Working Paper Archive	*http://econwpa.wustl.edu/wpawelcome.html*
SISSA Preprint Server	*http://babbage.sissa.it/*
High Energy Physics Preprint Database	*http://wwwspires.slac.stanford.edu/FIND/hep*
Nitride Semiconductor Research Preprint Server	*http://nsr.mij.mrs.org/preprints/*
Clinical Medicine and Health	*http://clinmed.netprints.org/*
Department of Energy's PrePRINT Network	*http://www.osti.gov/preprint/*
Theoretical Chemistry Preprints	*http://www.chemie.uni-regensburg.de/pub/preprint/GENINFO.html*

3.7 Print-on-demand

Print-on-Demand books are digitally printed from electronic files by high-quality laser printers, and then bound and cut. It is a process of replacing traditional paper media with digital print files. Printing becomes a demand process where the end-user determines the requirement for printed copies. Substitution of the digital file for paper media does not change the publishing process, but eliminates the requirement to distribute and stock paper. File servers are used as the publications stockroom and networks are used to distribute documents. The Print-on-Demand method is quite new and is a cost effective and efficient way to print one copy at a time. Print-on-demand services use new photocopying technology combined with streamlined binding methods and economical full-color digital printing to give 100, 250, or 500 books that look as good as if they were produced with traditional printing and binding equipment.

The process by which documents are printed in high volume hasn't changed much over the last 50 years until now. With Print-on-Demand solutions, publishers and other outlets can now print what they want, where they want it and when they want it. One of the amazing benefits of Print-on-Demand (PoD) technology is the ability to create a single document from a variety of different file formats. PoD solutions allow creating a document containing pages from virtually any number of applications. The shelf-life of information is getting shorter and shorter all the time. Because on-demand printing requires little or no set up time, document production can begin as soon as document creation is complete. And since on-demand printers, like the DDP series from Hitachi,

can be networked, documents can be printed at locations world-wide as they are needed. The technology including hardware and software are compact, relatively inexpensive, multifunctional and networkable. Soon, one will be able to walk into a bookstore and get any book printed in a short timeframe. The phrase "out of print" could soon be out of vogue. Barnes & Noble and Barnesandnoble.com are planning to use the latest technology to print books to order. This new twist to publishing will cut costs and better manage inventory. Previously more than one million books have been out of print with 90,000 titles disappearing each year. Many publishers were forced to turn down quality books with valuable editorial content because there was no market to justify the costs. This custom-printing service is the long awaited solution for persons needing a book title that is out of print because of a small press run.

Barnes & Noble, leading publishers, has brought an enormous amount of content previously unavailable to readers will now be able to reach the marketplace. More recently, NetLibrary has announced its entrance into the short run and Print-on-Demand marketplace. IBM, Xerox, Lightening and Sprout are some other active players.

Discussion forum, e-conferences, e-mails, etc. are also products of electronic publishing of informal nature. These products are not being covered here.

4 Media types and formats in electronic publishing

Products and services of electronic publishing may be produced in various formats and media types. For example electronic journals may be produced on CD ROM, delivered via e-mail or be available in full-text on the Web. They may also be made available as a mirror site for a given country or university or group of universities. Products and services of electronic publishing can also be hosted on Campus Intranet / LAN.

A defined arrangement for discrete sets of data that allow a computer and software to interpret the data is called a file format. Different file formats are used to store different media types like text, images, graphics, pictures, musical works, computer programs, databases, models and designs, video programs and compound works combining many types of information. Although almost every type of information can be represented in digital form, a few important file formats for text and images are described here. Every digital object needs to have a name or identifier which distinctly identifies its type and format. This is achieved by assigning file extensions to the digital objects. The file extensions typically denote formats, protocols and right management that are appropriate for the type of material.

4.1 Formats and encoding used for text

Text-based contents of a digital library can be stored and presented as i) simple text or ASCII (American Standard Code for Information Interchange); ii) unstructured text; and iii) structured text (SGML or HTML or XML).

4.1.1 Simple text or ASCII

Simple text or ASCII is the most commonly used encoding scheme used for facilitating exchange of data from one software to another or from one platform to another. "Full-text" of articles from many journals has been available electronically through online vendors like Dialog and STN in this format for over two decades. Typically what is stored in the text of each article, broken into paragraphs, along with bibliographic information is a simple tagged information.

Simple text or ASCII is compact, economic to capture and store, searchable, inter-operable and is malleable with other text-based services. On the other hand, the simple text or ASCII can not be used for displaying complex tables or mathematical formulae. Photographs, diagrams, graphics, special characters cannot be displayed in ASCII. ASCII format does not store text formatting information, i.e. italics, bold, font type, font size or paragraph justification information. Simple text or ASCII in many ways is inadequate to represent many journal articles because of the reasons mentioned above. Although simple text or ASCII is extremely useful for searching and selection, its inability to capture the richness of the original makes it an interim step to structured text formats.

4.1.2 Structured text format

Structured text attempt to capture the essence of documents by "marking-up" the text so that the original form could be recreated or even produce other forms. Structured text format have provision to imbed images, graphics and other multimedia formats in the text. SGML (Standard Generalized Markup Language) is one of the most important and popular structured text format. ODA (Office Document Architecture) is a similar and competing standard. SGML is an international standard (ISO, 1986) around which several related standards are built. SGML is flexible language that gave birth to HTML (Hyper-text Markup Language), *de facto* markup language of the World Wide Web, to control the display and appearance of documents. Like simple text or ASCII, structured text can be searched or manipulated. It is highly flexible and suitable both for electronic and paper production. Well-formated text increases visual presentation of textual, graphical and pictorical value of information. Structured formats can easily display complex tables and equations. Moreover, the structured text is compact in comparison to the image-based formats, even after including imbedded graphics and pictures.

4.1.2.1 Standardized General Markup Language (SGML)

SGML is application independent, non-proprietary and extremely flexible mark-up language. It was first developed in 1970 as GML (Generalized Markup Language) and evolved into both a national and International Standard (ISO, 1986). SGML is frequently referred to as a meta-language, which means that SGML is not a single language but a language that describes a family of markup languages. In other words, SGML is the framework for defining particular markup languages. SGML is an

effective solution for handling complexity of electronic publishing because of its powerful and flexible structuring capabilities, as well as for its capacity to capture and organize information about the publications ("metadata"). It provides for descriptive, as opposed to procedural markup. That is, it simply, states names to categorize parts of a document instead of specifying process to be carried out (Kardorf, 1998).

SGML uses text characters both for the text as well as for mark up that describes that text. It has no proprietary codes; instead each user (or group of users) may create whatever codes are necessary and meaningful for what is being published. A publisher can define his own set of codes for books and journal publishing. The key to self-defined codes in an SGML document is called DTD (Document type Definition). Codes sets or DTDs can be specific to a single book or journals or can span to a group of related books or journals. An SGML document consists of three distinct parts namely:

> **Declaration:** It gives fundamental information like language of document and code set being used (i.e. English/ASCII)
>
> **DTD:** Details of codes and rules restricting their use.
>
> **Instance**: The text being published, marked up with the codes described in the DTD.

SGML concerns itself with the structural features of a document while the appearance and display features are left to the ultimate presentation system to determine how those features appear on display or print. Resultantly, when documents move from system to system, or portions of one document are used in another, they don't need to be recoded. Because of its powerful and flexible structuring capabilities, as well as its capacity to capture and organize information about the publications, SGML-coded documents can be used effectively to search information contents of documents based on the structure and content of the information. Many SGML depositories are considered as "text databases". Since they enable a publisher to organize the published information in different ways for different contexts.

Contents of an SGML documents are stored separately from its format, resultantly contents on parts of contents can be rendered in different ways for different needs, platforms and display methods. SGML is often used as an archival format and for document reuse and repurposing. Richly-coded SGML documents also facilitate more complex searching than unstructured, word-processed text. For fully marked-up documents, searches can be made on bibliographic citation marked <site> or such citations can be extracted from each document to create a citation database as a secondary product.

SGML liberates documents from the cumbersome and costly process of conversion from system to system. It does not require any special hardware or software. It is possible to create a valid SGML file in any word processor or text editor although

there are number of SGML-based system available in the market. SGML preserves the document and its coding from obsolescence as well. Owing to the fact that an SGML document incorporates the key to its own codes (Declaration, DTD), it is possible to validate SGML codes by parsing the SGML file. Parsing is a process by which the document instance is checked against the declaration and the DTD to make sure all the codes in a file are legal and used properly.

4.1.2.2 Extensible Markup Language (XML)

XML is subset of the Standard Generalized Markup Language (SGML). It is designed to make it easy to interchange structured documents on the web. Like SGML, XML also deals with the structure of document and not its formatting. The Cascading Style Sheet (CSS) developed for HTML would also function for XML to take care of formatting and appearance. Unlike HTML, XML allows for the invention of new codes. XML files are not only consistent and compatible with SGML, it also simplifies SGML in many ways. For example, while SGML allows "tag minimization", enabling the omission of end tags, XML always requires explicit end tags that make it a lot easier to write tools and browsers. XML introduces the concept of a "well-formed" document, one in which the tags used are nested correctly and proper XML syntax is followed. In addition, like SGML, XML allows for "valid" documents too, which go a step beyond "well formed" status by using an explicit structure defined in a DTD. "Well-formedness" is a very appealing feature of XML, because it allows publishers to tag what they are publishing in whatever way is meaningful, without being confined to a specific set of tags (as with HTML) or needing to write a DTD.

XML document may require companion XSL (Extensible Style Language) to refomat in into RTF, LaTeX or any other format. XSL also makes it possible to offer database functionality from XML documents with no actual database needed. XML also defines how Internet Uniform Resource Locators can be used to identify component parts of XML data streams. Akin to an SGML document, XML documents can also be verified to ensure that each component of document occurs in a valid place within the interchanged data stream by defining the role of each element of text in a formal model, known as a Document Type Definition (DTD). An XML DTD allows computers to check, for example, that users do not accidentally enter a third-level heading without first having entered a second-level heading, something that cannot be checked using the Hypertext Markup Language (HTML). However, unlike in SGML, DTD is not a necessity in XML. If no DTD is available, either because all or part of it is not accessible over the Internet or because the user failed to create it, an XML system can assign a default definition for undeclared components of the markup. (Kardorf, 1998)

XML allows users to:

- bring multiple files together to form compound documents;

- identify where illustrations are to be incorporated into text files, and the format used to encode each illustration;
- provide processing control information to supporting programs, such as document validators and browsers;
- add editorial comments to a file.

Like SGML, XML does not have a predefined set of tags of the type defined for HTML that can be used to markup documents in a standardized template for producing particular types of documents. XML is formal language that can be used to pass information about the component parts of a document to another computer system. XML is flexible enough to be able to describe any logical text structure, whether it is a form, memo, letter, report, book, encyclopaedia, dictionary or database.

XML is based on the concept of documents composed of a series of entities or objects. Each entity or object can contain one or more logical elements. Each of these elements can have certain attributes (properties) that describe the way in which it is to be processed. XML provides a formal syntax for describing the relationships between the entities, elements and attributes that make up an XML document, which can be used to tell the computer how it can recognize the component parts of each document. XML differs from other markup languages in that it does not simply indicate where a change of appearance occurs, or where a new element starts. XML sets out to clearly identify the boundaries of every part of a document, whether it is a new chapter, a piece of boilerplate text, or a reference to another publication. The structure of a document can be checked if the user provides a document type definition that declares each of the permitted entities, elements and attributes, and the relationships between them.

4.1.2.3 Hypertext Markup Language (HTML)

Hypertext Markup Language (HTML) is an SGML application complete with DTD. It is designed to tell a browser how to format documents on the web. HTML is the *de facto* language of the web and is largely responsible for resurgence of interest in SGML in the past few years. Unlike SGML, HTML has a pre-defined set of codes, that are easy to learn and use and build tools for writing HTML pages. HTML codes are imbedded into the text that communicate to a web browser such as Netscape Navigator or Micorsoft Internet Explorer. Like SGML, it also uses simple text or ASCII for text as well as for the HTML codes. An HTML page can thus be built using a word processing package or a text editor. There are several HTML editors and conversion programme that act similar to a word processing package. These editors typically show the codes as they are inserted. In a What You See Is What You Get (WYSIWYG) environment, such as MS Word or other MS Windows packages, the user never see these codes. Web browsers are similar to WYSIWYG word processors because it reads the imbedded codes and then applies them to the specified text.

HTML is competent at presenting text, graphics, images in a reasonably decent layout. Web browsers readily accommodate a multitude of plug-ins that allow inclusion of audio, video, 3-D, and other specialized files. Any of these can also be included as a

link in a standard HTML page. Clicking the link loads the plug-in to view or play the file. HTML files are tiny since they are simple text files. Further, the static HTML web pages can be transformed into vibrant, dynamic and interactive web creations using ever evolving web technologies like CGI Script, Perl, Java, Javascript, ASP, DHML, XML and open database connectivity (ODBC) for incorporating interactivity on a web site.

Simplicity of HTML is also its serious limitation for books and journals. HTML does not provide enough codes to present complexities of a scientific text. It does not provide for Greek and maths characters that are important to scientific text. Moreover, HTML is all about presentation and not for structure or contents. The only contents that it describes is in Metadata codes or in its title. Furthermore, an HTML file can be derived from an SGML file any time but the reverse is not possible. The competancy of HTML at presenting text has further been enhanced with use of Cascaded Style Sheet (CSS) in its version four.

4.1.2.4 TeX

TeX is a typesetting program designed for high-quality composition of material that contains a lot of mathematical and technical expressions. Several individuals and publishing organizations have adopted TeX to generate technical books and papers. The TeX formatting language was created by Professor Donald Knuth of Stanford University, originally for preparation of his book series "The Art of Computer Programming". It has been made freely available in a generic form. However, commercial packages are also available that use TeX formatting language.

Knuth developed a system of 'literate programming' to write TeX together with tools for processing the program to compile it and a "DeVice Independent" (DVI) file that can be printed as output. A DVI file contains only positioning information and pointers to fonts, text characters and rules, and must be translated to a device-specific form for printing or display. TeX input consists of a stream of mixed commands and text. Commands can be defined for many purposes, not the least important of which is to permit input to be structured in a logical manner, allowing an author to concentrate on content rather than on typographic appearance. TeX and its documents are highly portable. It has been tailored and installed on almost every platform (computer + operating system) that one can imagine, and is available as freeware, shareware and commercial implementations. Other software to form a complete and usable system usually accompanies the TeX program.

4.2 Page description language

Unlike structured text formats (like SGML or XML) that concentrate on structure of a document, Page Description Languages confine themselves to describe and define page layout and appearance of a document. PostScript and PDF are two most used Page Description Languages.

4.2.1 PostScript

PostScript is a Page Description Language optimized for printing graphics and text whether on paper, film, or CRT. PostScript is essentially a programming language that define or generate a file describing page layout and appearance of a document for an output device, be it the printer, monitor or a print file. The language is used to write PostScript printer drivers. It was introduced by Adobe in 1985 and was first used in the Apple LaserWriter. The main purpose of PostScript was to provide a convenient language in which to describe images in a device independent manner. This device independence means that the image is described without reference to any specific device features (e.g. printer resolution) so that the same description could be used on any PostScript printer (say, a LaserWriter or a Linotron) without modification. PostScript is powerful and dynamic programming language that allows tremendous range of interpretation in various applications and ability to generate visually identical pages by a number of means. However, a PostScript file must be read from beginning to end since some of the crucial instructions (like font size and font type) are imbedded in the first page. It is not possible to extract any given page from a PostScript file.

Encapsulated PostScript enables imbedding of PostScript image into a PostScript document by following Document Structuring Convention (DSC) and certain other rules. Encapsulated PostScript (EPS), use Adobe's Document Structuring Convention to describe page layout, image and its size how to move it to the correct place on the page.

4.2.2 Portable Document Format (PDF)

HTML cannot maintain all the formatting and presentation layouts of the document as it appears in original. Portable Document Format (PDF), a by-product of PostScript, is used to maintain page integrity and layout of the page. The PDF format maintains the look and presentation of the original document. However, PDF is a proprietary item that a browser cannot read on its own. PDF files require Acrobat Reader, a free reader available from Adobe's site for different platforms. A PDF file is a self-contained cross-platform document and can be viewed across multiple platforms using the appropriate reader for that platform. PDF files are page independent, i.e. a user can extract any given page for printing. Every page of a PDF document has information to display or print. PDF files preserves all of the fonts, formatting, colors, and graphics of any source document, regardless of the application and platform used to create it or the platform or device to which they are downloaded. Windows, Macintosh, UNIX, and DOS versions of Acrobat Reader are available for free from Adobe Systems, Inc. In plain language, it is a file that will look the same on the screen and in print, regardless of what kind of computer or printer someone is using and regardless of what software package was originally used to create it. Moreover, since PDF uses vector technology rather than raster, the resulting files are very compact. (Kardorf, 1998).

PDF is the *de facto* standard for electronic document distribution worldwide although it is proprietary of Adobe. PDF files are compact and can be shared, viewed, navigated, and printed exactly as intended by anyone with a free Adobe Acrobat Reader. Any document can be converted to Adobe PDF, even scanned paper, using Adobe Acrobat 4.0 software. PostScript files can be converted into PDF using PostScript interpreter called Distiller.

PDF carries little or no structural information although Acrobat does offer some navigational features. Full text in a PDF document can be indexed using Acrobat's Catalog software enabling Boolean searching on collection of PDF files. Hyperlinks can also be created within or between Acrobat documents using Acrobat Exchange. Moreover, PDF files also allow incorporation of keywords, author's name, etc.

4.3 Page image formats

The backfiles of printed journals that are not available in machine readable forms are scanned electronically as bit-map page image and saved in Taged Image File Format (TIFF-G4) a *de facto* standard for saving bitonal images with minimum loss of data. Since TIFF files are fairly large and can not be viewed on the web without a TIFF viewer, they are converted into Adobe's PDF (Portable Data Format). The PDF, in turn, can be viewed using a plug-in called Adobe's Acrobat Reader available free of cost on the Internet. Adobe's PDF format appears to have market predominance for publishers or projects that are converting traditional print to electronic format. Some publishers are also converting their computer processible files (PostScript / SGML) into PDF for retaining the appearance of their print publication as well as for the page integrity.

Further, GIF (Graphic Interchange Format) and JPEG (Joint Photographer's Expert Group) formats, compatible with the web, are used for the pictures and photographs anealled to a document in HTML / XML

5 Process of publishing and transition from print to electronic version

The process of publication of a journal whether on paper or on electronic media has certain well-defined activities which include:

- Collection of manuscript from author
- Evaluation of contents (refereeing / peer review)
- Editing the technical contents
- Improving the language, style of presentation
- Composing, proofreading, page making, designing
- Printing
- Binding
- Publicity and advertising, catalogue, etc.
- Cost estimation and pricing

- Distribution and marketing
- Feed back and updating
- Copyright and other legal aspects

The publishers are already using digital technology to produce journals. Computerized typesetting and page layout software are common place replacing letter press and hot metal typesetters. The publishers have now re-engineered their print-based production process to accommodate electronic publications. Journal articles are frequently submitted on discs or as attachment to e-mails done in one of the popular word processing software. These machine readable files are pulled into publishing software like Frame Maker, Page Maker, or other publishing packages. Publishers apply their skills in quality management, presentation and layout to this computer-processible material. The documents are converted into HTML, PostScript and PDF. HTML and PDF (sometimes PostScript or Encapsulated PostScript) are posted on the web.

Publishers are now increasingly using SGML to reap the benefit that the format offers. SGML (or XML) documents provide the benefits of a database management system without being one. Publishers code the accepted submissions in SGML in a semi-automated process using assortment of software packages available to them or using custom-made software specially designed for this purpose. The database of SGML documents are used for providing search by authors, keywords, etc. and browse the content pages of journals. Behind the web interface lies a relational database like Oracle that stores SGML documents. Search and browse results in database-generated HTML pages (HTML-on-fly), which in turn, are linked to full-text documents mostly in PDF or PostScript and abstracts in HTML. While HTML is generated instantly from SGML documents available in the database, PDFs or PostScript versions are generated as a by-product in the process of printing of the documents.

SGML is all about structure and contents. It serve as rich archive format and is used for preservation to be reused for generating additional services and products. SGML documents are also used for genenrating print version. The publishers use their publishing software tools like FrameMaker, Pagemaker, Wang System, Folio, Xy Vision, Quark, etc. to generate a print version. PostScript and / or PDF versions are created in this process, which, in turn, are incorporated in the database along with SGML documents.

A research article would cease to be a linear text, written as if it is to be read from top to bottom as publishers move away from simply replicating their printed journals in electronic formats. Technology offers ample opportunities. How each publisher manages the transition from a paper-only information system to a more complex world of multiple media will decide his or her future in a truly Darwinian Scenario, say Cox (1997).

6 New media: bells, whistle and promising features

An electronic journal need not simply be an electronic mimicry of a paper journal. The available technology can provide dynamism to an electronic publication hitherto impossible in print publication. Some of the features that electronic journals can provide using the available technology include (Wilkinson, 1998):

- Linking citations and references to bibliographic databases or to full-text articles
- Links to graphics / photographs, video or audio clippings not included in the paper
- Links to corrections or to later articles that cite the paper
- Access to more detailed data or to multimedia information provided by the author
- Links to external databases like protein sequence or nucleic acid database
- Links to reader's comments or discussion forums related to the paper
- "Dual publishing" in more than one electronic journal, for example, a Chemistry article of interest to biologists could appear both in a chemistry and a biology journal
- A "living article" where the user could log in at any time and see an experiment on an ongoing basis showing data collected that day
- Embedded software programs allowing users to mirror the author's work by manipulating data or running simulations based on their own inputs

7 Economics of E-publishing

The opportunities that the technology offers has posed a challenge to the publishers to harvest raw materials from the digital ocean and fashion valuable information products. Universities and libraries, on the other hand, are forced to evaluate the possible ways of restructuring their budget to deploy their limited resources to their best advantage (Getz, 1997).

The opportunity to distribute electronically has implications for the publisher's costs and revenues. Electronic distribution of digital documents can definitely reduce the cost. However, there are several activities that are common to both paper and electronic publishing. Publisher's overhead cost is common for both paper and for support activities like editorial work, sales, etc. The major difference between the two costs are production costs for printing on paper or CD ROM production or alternatively setting-up and maintaining the digital collection on a server. However, maintaining high production value, editorial quality and presentation involve high cost that is common both to print as well as electronic version and are not related to distribution. The cost of printing and distribution of publications is modest of the overall cost on journal publishing. In essence, publications with high production values and quality will have high fixed costs. The fixed costs get distributed widely amongst multiple numbers of copies printed, with larger circulation of a journal.

Resultantly, popular publications have lower cost per copy while publications targeted to specialists in a given field have higher cost. Odlyzko (1999) observed that journal subscription costs are only one part of the scholarly information system. Internal operating costs of research libraries are at least twice as high as their acquisition budgets. Odlyzko further elaborated that for every article that brings $4000 in revenues to publishers, libraries in aggregate spend at least $8,000 on ordering, cataloguing, shelving, and checking out material as well as on reference help. It may be pointed out that Odlyzko's estimates are derived from the Association of Research Libraries (ARL) statistics that did not include cost associated with construction and maintenance of library building. If these figures are added to the Odlyzko's figures, we will have a much higher estimate of non-subscription cost.

Most publishers argues that the costs cannot be reduced much even in electronic delivery of information since most of the cost is the first copy cost of preparing the manuscript for publication. The argument is, however, refuted not only by widely differing costs amongst publishers but also by the fact that there is a rapid growth of electronic-only journals run by scholars themselves. Electronic publishing offers the possibility of bringing down not only the cost of publishing but also the library cost. For a comparison Paul Ginsparg's Preprint Server that processes about 20,000 papers per year would have costed US $ 40 - $ 80 million to publish in conventional print journals. However, it costs maximum of $ 1,00,000 per year or about US$ five per paper including half the time of a systems administrator, plus depreciation and maintenance on the hardware. Factors that would help cut-down the cost of journal publishing as well as that of library are:

i) Advances in technology, which makes it possible for scholars to handle tasks such as typesetting and layout design that used to require trained manpower and a large infrastructure. Most scholars are already typesetting their own papers and editors and referees are increasingly processing electronic submissions, even for journals that appear exclusively in print.

ii) The other factor is the peculiarity of the scholarly journal systems, i.e. scholars function as referees and serve on editorial boards without charging for it. It may be pointed out that monetary cost of the time that scholars put into the journal business as editors and referees (without charging for it) is about as large as the total revenue that publishers derive from sales of the journals.

The facts mentioned above reveal the irony that academic institutions end-up buying information from the publisher that, in the first place, was generated and later peer reviewed by their own scholars. Ginsparg's Preprint Server has sprung up as a reaction to the increasing cost of journals. Similar initiatives can be taken up in other disciplines to either completely alter the existing model of scholarly publishing or put forth a working alternate model. Scholars would continue to review the work of their peer on gratis in the alternate scenario. *Chicago Journal of Theoretical Computer Science* is one of the peer-reviewed electronic journals under development at MIT Press, which provides promising new model for applying cost-sharing and risk

reduction in academic publishing. Models based on electronic publishing would also cut down the costs of technical processing associated with print journals, i.e. check-in, check-out, repair, missing issues, claims, ordering, etc. which, in turn, would allow employing lesser number of staff in the library. The libraries may deliberately prune their expenditure on print subscription so as to commit additional resources for digital materials. Increased commitment to electronic resources would make it possible to prune technical services costs and staff employed for it.

7.1 Pricing model

One of the major issues that the publishers are concerned with is to save their economic interest in the process of providing electronic access to their printed publications. The publishers make a significant investment in the process of production of a journal which involves activities like peer-review, administration, editing, layout design, production, subscription management and distribution. Most activities that are performed for publishing a journal are common to both electronic and paper media, except for production and distribution where the cost involved is relatively low. Tenopir and King (1997) in a study concluded that the costs of electronic journals cannot be substantially lower than their printed versions.

Journals are made available through the web at varying price models. Some of the prevelant pricing models are:

i) Electronic subscription is linked to the print subscription

The electronic subscription to journals in most of the cases is linked to their printed counterparts, i.e. it may be offered free with print subscription (e.g. publications of American Society for Physics and AIChE) or priced at a fixed % over the print subscriptions (e.g. IEEE's ASPP package).

ii) Electronic subscription with campus licenses

Electronic publisher facilitate campus wide unlimited access to subscribed journals on payment of a fixed amount of platform fee. Example: Elsevier Science (ScienceDirect)

iii) Electronic subscriptions are bundled

Several electronic publishers offer access to the entire range of their electronic journals and other publications bundled into one. For example IEEE / IEE Electronic Library (IEL) and ACM Digital Library offer access to their entire site on subscription. Access to individual journals or a subset is not permisible. Similarly, Academic Press offer all journals available on their site (Academic's Project IDEAL) for 10% more than the print subscription to library consortia.

vi) ***Pay-per-look***

Publishers and aggregators have started experimenting with models wherein a user can search a database online for a modest usage fee, identify articles on interest, and then call up such articles in full-text on a per-look basis.

v) ***Electronic only***

A few publishers and aggregators have started offering only electronic version of their journals providing a modest discount for those who forego print version.

vi) ***Consortium licensing***

Consortia provide union strength to negotiate with electronic publishers for the best possible price and rights. Most publishers already have well-defined policies and offers for libraries subscribing as consortia . The consortia licensing is widely used the world over by the libraries. It is slowly picking up in India also.

vii) ***National licensing***

National licenses can also be negotiated with electronic publishers for core collections. Singapore, Taiwan and UK have arranged national licenses for some of the important full-text resources.

8 E-journals: trends and unresolved issues

8.1 Fragmenting E-journals into E-articles

The organizational problem makes it necessary to treat a journal as an entity. The individual articles, not even an issue of a journal, is really a fundamental unit. It is the citation and links between articles that define the process of scholarly communication. Printed journals, being fixed entity, constrain the flow and format of information, on the contrary, digital technology and networks facilitate it. The publishing is already shifting from a focus on issue of a journal as the fundamental unit to focus on individual articles. Favouring early release of scientific contents, major scientific publishers discussed the issue of releasing published material on an article by article basis, not bound to any specific weekly or monthly issue. The American Chemical Society posts individual journal articles as soon as they make it through the review, editing, and author proofing process, a format ACS calls "As soon as publishable (ASAP)". The production process has been modified so that instead of thinking of journals as batches of articles that get bundled up and put into specific issues that come out once a week or once a month, the ACS now treat each article as it is finished (Wilkinson, 1998). Similalry, Springer Verlag also puts articles that are ready on the web as "Online First".

Form and content may fragment even further as publishers move away from simple replication of print to online. Kirez (1998) suggests that the scientific material be prepared in modular form, with elements that can be understood on a stand-alone basis. Some of the modules could potentially be used by more than one author. The description of a common experimental technique, for example, can be written up in a module that subsequent researchers could refer via a link without having to duplicate the write-up themselves.

8.2 How full-text is full-text?

Contrary to popular believe, journals described as "full-text" are not available cover-to-cover. Very few articles are provided online with their tables, charts, graphs, illustrations or photographs intact. At times such items are noted or described. In general, full-text means that the entire text of an article is available online.

As far as "full-text" of entire publication is concerned, it is rare that entire publication is made available cover-to-cover. Although the coverage policies differ widely from one publisher to another, rarely do the online periodicals reproduce advertisements, book reviews, notices, corrections, announcements, meeting calenders and letters to the editors that appear in the original printed versions.

8.3 Reliability and accessibility of data

The access to resources on Internet can be painfully slow and unreliable, especially from India due to bandwidth restrictions of physical media and data traffic on the net. Moreover, providing Internet access to e-journals in the Library would entail provision of sufficient number of Internet-enabled PCs for the users. The users would also require hard disc space and printing facilities for printing and storing the material they accessed. Moreover, a user may have to go through a number of screens to reach desired article from the home page of a publisher and each page may take some time to download.

The journals in web environment support inclusion of audio and video to documents. The addition of audiovisuals would add to the network burden where the delivery of fulltext and graphics over the net can be very slow at peak times.

8.4 Intellectual property and copyright

Protection and ownership of intellectual property in the age of electronic information are especially confusing in light of traditional copyright laws. Discussions are taking place at various platforms to review the existing copyright laws in the light of electronic information. Since the images are electronically forwarded around the Internet, it becomes very difficult to control and define what can and cannot be done.

A Conference on Fair Use (CONFU) in January 1996 working party comprising both publishers and librarians began the process of developing practiceable guidelines for

fair use of electronic information. The first discussions concerned the scanning and storage, reproduction and distribution of materials in an electronic preservation system. The working party failed to agree on any guidelines but the dialogue is still alive and is expected to result in some guidance to both libraries and academics on what is permissible without prior permission (Cox, 1997).

8.5 Electronic archiving and backfile availability

A subscriber to an e-journal gets access not only to the current issues but also to the backfiles and the whole system that enable him to use the journal in an enhanced way. Subscription to e-journals, therefore, entail dependency of libraries on continuing availability of archives of e-journals on the Internet. Considering the cost involved in keeping the backfiles online, which are more likely to be referred less frequently in comparison to the current issues, the publisher may discontinue their online availability on the web. Further more, on canceling electronic subscription, a subscriber is denied archival access as well which is quite contrary to print subscription wherein a subscriber retains the old volumes of journals on cancellation of current print subscription.

This brings up the thorny issue of responsibility for sustained and perpetual availability and access to the archives for a long period of time. An issue to ponder over is whether some of the new "electronic only" journals and their archives will still be around years from now, or will they disappear irretrievably (Machovea, 1997).

The OCLC has made commitment to provide perpetual archiving for titles offered by them. Although no definite indications are available, but UMI, Silver Platter and other commercial organizations may step-in to cater to the market for archival copies of e-journals.

9 Electronic journals: current market

The current electronic publishing market consists of traditional players offering electronic versions of their print journals as well as several new enterprises offering new products and services that are "born digital". The market also has several subscription agents in their new role as aggregators. These players include:

9.1 Publishers

Most well-known commercial publishers of traditional journals such as Elsevier Science, Kluwer Academic Press, Academic Press, Springer Verlag, Wiley InterScience and scholarly societies such as SIAM, ACM, IEEE / IEE are making their publications available online through their web sites. Several universities host specialized collections. Several universities are members to the Networked Digital Library of Theses and Dissertations (NDLTD) initiative and host doctoral dissertations submitted to their respective universities.

9.2 Aggregators

Third party aggregators provide access to numerous journals from a variety of publishers. Aggregators include organizations like JSTOR that offer extensive backfiles for more than 100 academic journals and OCLC Electronic Collection Online which offer full-text access to more than two thousand titles via their First Search Service. Other aggregators like Lexis-Nexis, Bell and Howell (UMI) and Web of Science (ISI) offer searchable indexes with links to full-text journals on publisher's site. EBSCOHost, IAC Trac SearchBank and Blackwell's Electronic Journal Navigator (EJN) provide common search interface for the journals aggregated by them from an assortment of publishers. Growing number of susbcription agents are working with publishers to provide aggregated services to packages of titles or to full-text databases.

9.3 Non-commercial portals for electronic journals

The portal sites or gateways redirect a user to the holders of the original digital material. A gateway may provide its own indexing and search services and it may combine original resources from a number of different providers. The portal sites or the gateways restrict their operation to providing linkages to independent third party sources. Home pages of all the major education and research institutions, especially in the developed world, provide an organized and structured guide to electronic resources available on the Internet. Some of the major portal sites or gateways that provide access to electronic journals and other resources on the Internet are as follows:

WWW Virtual Library	*http://www.edoc.com/*
Internet Public Library	*http://www.ipl.org/*
Michigan Electronic Library	*http://mel.lib.mi.us/*
Penn Electronic Library	*http://www.library.upenn.edu/resources/*
BUBL Information Service	*http://bubl.ac.uk/*
Argus Clearing House	*http://www.clearinghouse.net/*
Internet Index	*http://sunsite.berkeley.edu/InternetIndex/*

10 Alternative models

The information scientists have been toying with the idea of replacing the existing print-based scholarly communication system with a system that revolve around the users rather than the authors and publishers. Salient features of some of the important models offered as an alternate to the current model are given here.

J.D. Bernal, in 1950, conceived that the scientific literature would be organized so that its elements could be directed exclusively towards those with particular interest in them. The conception never saw fruitation because technology was inadequate and the users indifferent.

Paul H. Ginsparg , a Physics Researcher at the Los Alamos National Laboratory in New Mexico, started an "e-print Server" in 1991 to distribute the pre-prints electronically. The idea was received well because researchers in high energy physics were acustomed to making pre-prints of every new paper and shipping them out to about 400 people before any thing got published. But the success of Ginsparg's Server has not been replicated in fields such as chemistry where preprints are uncommon. Some of the leading journals would not consider a paper for publication if preprint have been widely distributed.

Smith (1997) suggested a web-based "deconstructed journal" containing links to relevant items of interest to its subscribers. Some of the important difference between the existing paper-based model and Smith's New Scientific Journal (NSJ) are as follows:

- Unlike in the existing model, the copyrights for the linked material would not be with the operator of the service
- The operators of this service may, or may not, arrange the quality control in the publishing process
- The role of the refereeing, organized by the publisher in the conventional model, would be played by independent organisations who would validate or give their "stamps of approval" for publication of the item
- Initial layout design of an item to be posted on the NSJ can be done by the author himself
- After a researcher has finished his article, he would put it up on a visible server and inform the "evaluators"
- The revised article approved by the evaluator can go to relevant NSJ

Varian (1997) proposed his model for electronic publishing of scientific research where in authors would submit their papers consisting of three parts, i.e. an abstract, a five page summary and a detailed full-length paper. Summary of the submitted papers would be read by members of editorial board (2 – 3) and a rating of 1 to 5 will be assigned to all articles. The editorial score of 1 – 5 will be attached to the paper. Author would be notified the score his / her paper has received with an option to withdraw it if so desired. Once author’s consent is received, the paper is posted on the server with editorial score marked on it. Subscibers of the service would be allowed to browse and search all parts of submission, i.e. abstracts, summary or full-length paper. Subscribers may also rate the paper and mark their scores anonymously.

Varian proposed that the system with scores available on the paper can be used creatively, for example, all papers above a given threshold are notified to all subscribers or subscribers may request notification for articles in a given subject area that receives a score beyond a given threshold.

11 E-Journals and electronic publishing: myths and realities

Electronic journals are most talked about aspect of Internet. Some of the views widely held about the e-journals are far from the truth. Woodward *et al* (1997) and Brunelle (2000) provide insights into myths and realities pertaining to e-journals and electronic publishing respectively:

- **Electronic journals are quick and convenient to access**: Network delays due to bandwidth restrictions are common place. Moreover, to reach a particular full-text article by browsing through table of contents, one has to go through several screens.
- **Users know the publishers of journals**: Almost all big banners in the publishing industry have launched their web-sites to provide online access to the electronic versions of their publications. Typically publisher's web site provide list of journals that are being made available in electronic format. However, the users generally do not know the publisher of journals, rather they would like to access all journals published in their field of study.
- **Readers want "Page Integrity"**: Most commercial publishers are using PDF format to display the full-text of their journal primarily to preserve the page integrity of their printed publications. Though, the PDF format preserve the page integrity, but in bargain, several benefits, that an electronic document offers, are lost.
- **Electronic journals will bypass libraries and make them redundant**: The librarians and their skills would be very much in demand in the changing scenario. The libraries and librarians would still be in demand for developing common interface for numerous electronic services that are available, for managing and negotiating subscription-based electronic journals, for searching, evaluating and providing digital collections available on the Internet through web pages.
- **Electronic journals will save money for the library**: There is little or no evidence that publishers are likely to reduce their prices when journal becomes electronic. The prices of e-versions of journals are either tagged to their printed counter parts or at an additional 10-15% on top of the print subscriptions price for e-versions. Only a small number of publishers are offering e-journals as separate, stand-alone subscription. A few publishers have started offering discounts for subscribing electronic versions only.
- **Storage and dissemination of e-journals is inexpensive or free:** Substantial cost is involved for holding the full text of long runs of many voluminous journals. Moreover, hardware, network, software and upgradation of infrastructure involve high cost.
- **Electronic journals will save paper**: It is frequently argued that the move from printed journal to e-journal will save large amount of papers. On the contrary, if a user does find a paper that interest him he would invariably take a print out.

- **Electronic journals will save publishers money**: Most of the activities like peer review, administration, editing and layout designs, etc. involved in publishing a journal either in print or in e-version, are common. The production and distribution activities that are specific to the print media cost relatively low for most journals. Tenapir and King (1997) in a study conclude that cost of e-journals publishing are not much less than traditional paper:
- **E-Journals would make subscription agents redundant**: The Libraries generally subscribes the journal in print through subscription agents, so as to avoid dealing with hundreds of publishers. Although the libraries may start subscribing through "consortiums of libraries", but subscriptions agents would make their presence felt in a new roles of aggregators and archivists.
- **Only recent issues of journals are required**: Most of the publishers provide access to the current issues with backfiles typically varying between 2 and 5 years from publishers to publishers. In reality, use of backfiles of journals continue practically for several years.
- **All scholarly journals will be electronic within a few years**: Hundreds of thousands of journals are published in a huge variety of subjects; only a relatively small percentage of these have electronic counterparts.
- **Electronic journals can be accessed free on the Internet**: Most of the publishers launch electronic versions of their print journals on free trial basis. After the free-trial period is over, the access, specially to full-text of articles, is provided for a fee. Contents pages in most of the priced services, can, however, be viewed free.
- **It is already electronic because it is computer typesetting:** Typesetting is not the same as electronic publishing because they are tagged based on physical page elements and are created from another format, not “authored’ in the typesetting language. Graphics and tables are often separate from the documents. Publishers outsource and do not own or control the typesetting process or technology.
- **All full-text formats, especially PDFs, are ubiquitous and readily available**: There are economic constraints on publishers in providing one or another format, as well as significant production issues. PDF formats can be faster but limit functionality. PDF represent print exactly, are easy to print, are not truly searchable, make it difficult to do internal linking, and do not lend themselves to any of the potential interactive features of electronic publishing. They are a convenient way to quickly provide print on-line. Text (SGML) formats are expensive and slow, but are fully functional. Text (XML) is the path to true electronic publishing. It is easier and less expensive to create than SGML. Stylesheets allow for good print formats. XML is fully searchable, easily linkable, and fully supports the promise of electronic publishing. Not all browsers support it, however.
- **Information is free, or should be**: Providing and achieving access to information is not free.

- **Eventually licensing, formats, pricing, citation linking methods, and authentication routines well be the same for all e-journals**: Standardization is the enemy of innovation.

12 Conclusion

The ongoing shift towards electronic publishing and access is expected to continue inspite of the fact that printed media is still preferred for ease of reading and portability and because of the fact that authors still consider it as authoritative medium and format for the publication of peer reviewed research. Most of the electronic journals would continue to be e-versions of their printed counter parts. The scenario may, however, change in time to come depending upon the value additions done by the electronic version of journals. With formation of publisher's alliances like Crossref, which would provide search and linking facility at article level, indexing and abstracting services are likely to loose its luster. Publishers will have to stop treating e-journals simply as electronic mimics of printed versions. The hypertext and hyper media linkages have greater applications in the literature of science and technology. The idea of fragmented articles forwarded by Kircz may have great implications in time to come.

The electronic publishing promises equitable access to electronic resources at a much wider plane though it still has to achieve economy, authority and authenticity in addition to advantage of speed and value addition that it already possess. The publishers would gradually eliminate print editions that would reduce their cost by 20 – 30%. Likewise, the libraries would also save in terms of number of staff that they have to deploy to manage print collections, as well as the space that the printed version occupy. The future of e-journals will depends on the relative advantages that they could offer to those who publish, use or manage it.

13 References

1. BORCHERS (J. O.). Electronic books: definition, genres, and interaction design patterns. ACM CHI 1999, Electronic Book Workshop. *http://www.fxpal.com/chi99deb/*
2. CHAKRAVARTI (A.K.) and ARORA (Jagdish). Internet-based online interactive courseware in IT: effective delivery of IT education in a networked society. IN NACLIN 99: National Convention on Library and Information Networking. New Delhi, DELNET, 1999, pp. 26-37.
3. COX (John E.). Publishers, publishing and the Internet: how journal publishing will survive and prosper in the electronic age. *Electronic Library*, 15(2), 1997, pp. 125-131.
4. Directory of electronic journals. IN Newsletter and Academic Discussion Lists, Washington, D.C.: Association of Research Libraries, 1997.
5. GETZ (M.). An economic perspective of E-publishing in academia. *Journal of Electronic Publishing*, 3(1), 1997. p. 28. *http://www.press.umich.edu/jep/archive/getz.html*

6. KARDORF (B.). SGML and PDF: why we need both. *Journal of Electronic Publishing*, 3(4), 1998, p. 14. *http://www.press.umich.edu/jep/03-04/kardorf.html*
7. KIRCZ (Joost G.). As quoted by Wilkinson, Sophie L.
8. MACHOVEA (George C.). Electronic journal market overview - 1997. *Serial Review*, 23(2), 1997, pp. 31-44.
9. ODLYZKO (A.). Competition and cooperation: ibraries and publishers in the transition to electronic scholarly journals. *Journal of Electronic Publishing*, 4(4), 1999. *http://www.press.umich.edu/jep/04-04/odlyzko0404.html*
10. SMITH (J.W.T.). The deconstructed journals. IN: Electronic Publishing'97: New Models and Opportunities; proceedings of the ICCC/IFIP Conference on Electronic Publishing. University of Kent at Canterbury, UK, ICCC Press, 1997, pp.73-84.
11. TENOPIR (C.) and KING (D.W.). Trends in scientific scholarly journal publishing in the United States. *Journal of Scholarly Publishing*, 28(3), 1997, pp. 135-770.
12. WILKINSON (Sophie L.). Electronic publishing takes journals into a new realm. *Chemical and Engineering News*, May 18, 1998, pp. 11-18.
13. WOODWARD (H.) [et al.]. Electronic journals: Myths and realities. *OCLC System and Services*, 13(4), , 1997, pp. 144-151.