

JOURNALS PUBLISHED ON THE NET

David Pullinger

The wide variety of publishing initiatives and related projects, described in the article, are an indication of the potential for new products inherent in the Internet and the associated systems. It is only by harnessing the multimedia possibilities of these systems, matching the distribution methods to the technical capabilities of the customer and by addressing the questions of consistency and coverage, that publishers will make effective use of the opportunities in support of scholarly communication.

David Pullinger, formerly of Institute of Physics Publishing, is now Electronic Publisher with Nature, Macmillan Magazines, 4 Little Essex Street, London, WC2R 3LF

Introduction

Many individuals and organisations are using the Internet to give end users and libraries access to and delivery of information. The range and features of these services vary considerably. In this paper I want to cover some of the features of the use of Internet (or Net for short) for journals publishing and to stimulate you into seeing new uses. In particular I want to describe the way that technical solutions using the Net have implications for resourcing and skills acquisition for participants in scholarly communication, but that they also affect at what point consistency can be achieved and sufficient coverage obtained to make services useful to the user.

Servers, containing journal material which can be accessed, include both those containing 'raw' data and those providing a full searchable and readable archive. The former contain 'raw' information such as ASCII, TeX and image files which can be brought back to the user's workstation for reading and viewing. These are typically Table of Contents of serials and the text of journal articles, for example, the High Energy Physics databases. The latter use of servers is in storing information which is directly usable over the Net, for example, World Wide Web servers, the OCLC journals and the SuperJournal experiments in Europe on the new high speed networks with colour and mathematics displays. The user can log on to search and read material.

In addition to accessing information, the Net is used to deliver information. In the TULIP experiments the Net is used to deliver electronic information to libraries for their subsequent use in Campus Wide Information Systems (CWIS). In the current awareness service CoDAS, jointly run by Institute of Physics Publishing and Elsevier Science Publishers, data is delivered over the Net directly to end-user or department machines. Information is of course also delivered via e-mail, including being requested from mailservers.

Network serials publication comprises, therefore, a range of uses of the Net. It would be easy to simplify these, but creative developments will come out of seeing niches and opportunities, either in particular neglected subject area or in combinations offering new value-added services.

One might offer many classifications of the ventures into journal publishing on the Net but I have chosen one which leads to an analysis of where the costs will lie, as choices are made as to how to

use the Net. I will use as examples those with which I am most familiar, those initiated in the physics community. As a subject area, physics offers a bewildering fragmentation of experience in using computer networks, ranging from the tightly-knit community of high energy physics, each equipped and familiar with a UNIX workstation, to groups which still have little or no computer equipment or expertise. To a publisher such as Institute of Physics Publishing, the physics community offers a challenge to design a system appropriate for its customers.

Internet

The Internet is often described as the collection of networks which connect to other networks of computers. Formally the Internet applies to the network of networks connected physically and using a particular technical protocol TCP/IP. Information stores and services on Internet make use of the Internet protocol (IP) applications: e-mail; file transfer, and logging onto another computer. I will explore journals publishing and related services within the context of these three basic functions of the Internet).

E-mail based services

Historically, new 'journals' and support services to journals publishing have been conducted via e-mail. There are two good reasons for this: e-mail offers a broad distribution and added value over print on paper.

1. *E-mail offers the widest network of potential users*

The standard that defines the store-and-forward functionality of e-mail extends well beyond the Internet to other networks. Indeed in July 1994 approximately 30 more countries had networks receiving e-mail which did not handle file transfer or logging onto computers. This made e-mail the natural choice for those seeking to reach the maximum number of prospective recipients of information. Successful new magazines have identified niche subjects which could not be sustained in a geographical area, but can be with the world-wide spread of networks, despite the extremely specialised nature of the subject and low density of readership.

2. *E-mail supports two added value functions for electronic journals, immediacy and bi-directional communication.*

Immediacy is the rationale for news services being introduced separately or in conjunction with print journals. IOP publishing has a weekly newsletter distributed by e-mail for which a subscription is charged. VNU Publications have pioneered such newsletters over the past years, focussing on daily news in the computing business. Immediacy and accessibility are the reasons for widespread advertising of the contents of a journal by distributing the Tables of Content by e-mail. Immediacy is also the goal of alerting a user or customer that there is material of interest in a centralised store. By storing a personal profile of keywords, a service can do an automatic search at regular intervals and send an e-mail identifying relevant content to the user. OCLC's Guidon, for example, offers this for its electronic journals.

BYTE has successfully introduced a series of bi-directional services associated with the magazine: asking readers for comments on prospective articles, as a means of corresponding with their readers (and of knowing who they were); putting reader in touch with reader, and of offering special talk sessions over a week, say, with some famous person.

A similar use of BYTE, is the development of Psychology from the Editor's experience of the Cambridge University Press journal, *Brain and Behavioural Science*, where each article contains feedback before being 'published'.

File Transfer Protocol (FTP) based services

E-mail itself, however, does not offer the user many functions. For this reason the Elsevier Science - Institute of Physics Publishing collaboration on the current awareness service CoDAS chose, in 1993, to deliver a file of pre-publication citations and abstracts to the user's machine by file transfer (FTP). There a piece of software, especially written for the service, picks up the file for use. This client software will unencrypt the file and mount it ready for the user to browse or search lists of titles and authors, examine the abstracts and mark entries for printing or request for document delivery. The second reason why CODAS chose to use FTP was

that e-mail could not handle the size of file that the service was offering. Current awareness demands a large amount of content and searchability pre-supposes sufficient material for there to be meaningful searches.

In summary, file transfer using FTP was chosen because:

- the publishing collaboration could provide added value to the user in the software over and above e-mail, and,
- the quantity of material delivered each week exceeded the capacity of many e-mail systems to handle it.

The electronic medium was chosen so as to provide as early an opportunity as possible to distribute information about articles appearing two to four weeks later in the journals.

It has not always been easy for users to set up their end of the Net in order to receive files. One piece of learning for publishers is that technical support is necessary for end-user customers. For this reason a facility to allow the user to collect files by 'anonymous FTP' from the publisher has also been introduced and further developments are planned.

Institutions as customers for electronic journals data

File transfer is also used to deliver journal articles in electronic form to libraries. TULIP, CORE, Red Sage and ELVYN are each exploring possible advantages in delivering and holding material at an organisational level for their customers, the students, lecturers and researchers. Each of these four projects was developed for quite different reasons: to explore licensing agreements; to create a complete collection; to demonstrate technical feasibility; and to prepare libraries and publishers for the electronic future.

The advantages of holding electronic journal at an organisation level include the following:

- collections of full-text electronic journal facilitate the added value of searchability, and,
- network usability is dependent only on what the organisation is willing to implement and support, rather than being determined by others and the network load.

These four projects each point out how far support issues need to be addressd in universities

in order to make effective use of the technical possibilities and how little users seem to desire to make use of such accumulating resources. The funding, offered by the US National Science Foundation to six major collaborations studying aspects of the digital library, should advance this area considerably.

Another kind of collection was made by Paul Ginsparg at the Los Alamos National Laboratories (LANL). Originally technical reports of high energy physics experiments and papers in preparation were stored in TeX. Users on an e-mail list received lists of the e-prints and requested copies of the papers by e-mail or collected them by file transfer (using anonymous FTP).

The value of this collection was in its immediacy and completeness, as almost all workers in this tightly knit community used it as a store. The need to mirror the entire database, by copying it five times around the world in order to assist access, indicates that international networking still needs further development. The LANL databases themselves have been extended in content and in pace with new applications on the Net.

Telnet based services

Most ASCII-based electronic journals have been held on a centralised store which the user could access by logging on using Telnet and then displaying the text on a computer screen. In particular the majority of e-journals collected in the CICnet store are like this.

There are two main reasons why e-journals were text only: computers were, until recently, only able to handle ASCII text, and the speed of the networks were too slow to facilitate the transfer of large colour images at the time of viewing. Both have recently changed, with GUI (Graphic User Interfaces) becoming widespread on personal computers and the development of higher speed network infrastructures (at least in some countries) funded by the government. These changes led to the opportunity to explore full text, with colour images and in-line mathematics in the UK SuperJournal Project 1993 funded by the British Library Research and Development Department.

The use of Telnet in the Internet Protocol (IP), to log onto another computer, is also the means by which CompuServe, Prodigy, Dialog and OCLC (among others) have allowed their services to be accessible over Internet.

Gopher and World Wide Web (WWW) are developed applications using Telnet and FTP as underlying processes. Indeed, new applications are designed to solve a number of previous problems:

- the user knowing where the location of relevant material is;
- moving between relevant material easily, and,
- the integration of locating, accessing, viewing and collecting relevant material.

These new applications enable information providers to move beyond ASCII e-journals and there are a number of projects putting versions of print journals onto WWW servers, for example, Loughborough University of Technology with the Institute of Physics Publishing Journal in the ELVYN project. In this case, however, it was discovered that each paper might contain as many as 150 associated graphics files - for all the Greek letters and mathematics, as well as images. New revisions of GUI browsers for World Wide Web are promised to deal with the integration of special characters and images.

To get round this particular problem and to take advantage of the interrogation of Telnet and FTP in Web browsers, the journal *Classical and Quantum Gravity* is put up on a Web browser so that those with GUI can see a good scanned image of the title and abstract (so that the mathematics is readable) and then transfer the file (which is in TeX to match the habits of that physics community) which automatically launches a TeX viewer.

Another means of logging on and downloading data in an integrated way is found in Lotus Notes. While being used for electronic publishing over the Internet, it has not yet, to my knowledge, been used for journals. Other proprietary software has been developed explicitly for journals by OCLC in their Guidon software. Using their client software, the user can log into OCLC's computer in Dublin, Ohio, US and access, search, view and download material.

The most widespread use by publishers is to have a server on Internet which informs users,

who access it, of their catalogue contents, latest product releases and teasers to entice the users to buy print and electronic products - "Up-to-date catalogues on-line". In some cases the servers offer on-line purchase and associated services to authors - files containing author instructions, style files and information, again combining logging on with file transfer.

It seems to me that the real benefits of publishing on the Net arrive with the creative development of services that mix and match technical capability with content, so as to provide the most effective service to the user. This may indeed mean that to meet the different technical capabilities of the users, publishers will need to make their products available in a wide variety and combination of the ways described, rather than just choosing one of them. Another argument for so doing is that for some years the market will contain a large mixture of technical facilities. Look forward, imagination will be required within current limitations to make use of multimedia (not serial monomedia as most interfaces offer) and the dynamic use of data.

Location of processing and storage

The choice of service from these large number of possibilities not only affects the potential size and type of market, but also the cost. Where will the storage of all the information be done? With e-mail based services, much can be put out to the end-user or organisation (depending on the e-mail system they are using). With Telnet centralised services, one can offer a much better graphical interface, but there are higher storage requirements, line and access capabilities to support and personnel to administer the systems.

In Figure 1, I have tried to indicate the options of processing and storage, in order to encourage thinking about the requirements and costs for the publisher and the customer in providing and buying services. It is divided into three, for centralised, organisational (Campus Wide Information System) and individual end-users. Filled in are a few examples in research publishing. Further enhancements of this can be made with deals that operate, say, at the State level in the US (as CARL UnCover, for example, does in its Table of Content and document delivery system) or national level.

Stored / Processed	Centralised site	CWIS	End-user
	Centralised site		
Centralised site	most e-journals OCLC journals (ELO) WWW journals, etc (SuperJournal Project)		
CWIS		TULIP ELVYN Red Sage	
End-user	Lotus Notes publishing Los Alamos D/B (TeX) Hypertexts (zipped)	e-mail newsletters	CoDAS

articles for coding into SGML. This group of English-speaking European publishers, under the name OASIS, are introducing consistency, with the aim of achieving coverage for the user later in the cycle of scholarly communication. The questions concerning consistency and coverage lie at the root of making an effective use of the technical opportunities.

Figure 1 : Location of processing and storage

Consistency and coverage

It will be clear that there are many opportunities for new products to serve the research community and meet their diverse and multitudinous requirements. Two crucial questions are the following. At what point is consistency to be introduced and at what point is the coverage required to be obtained?

There are different views on this. Clearly in the past the main coverage has been obtained through library stock and the use of abstracting and indexing services. Going towards an 'on-demand' rather than 'holdings' policy, and the integration of A+I services and document delivery, moves that coverage from organisational level to national or international level. The TULIP and CORE projects are experimenting with moving consistency to the organisation. Small sub-disciplines such as High Energy Physics have moved towards consistency and complete coverage at the pre-publication stage by the use of TeX, but it is lost thereafter. This is an area that needs a collaborative approach among publishers and between publishers, librarians and technical service staff, including software suppliers, computer centres and network services.

For this reason it is pleasing to see publishers meeting to agree some standards for the fields that should be included in the electronic bibliographic header and abstracts of journal

To address this area, we have to look at the function of the information and then discover the coverage and consistency required to give the end-user (researchers, teachers, students) an effective service. A first pass at doing this is presented in Figure 2.

Figure 2 : What is the function of the information?

Function	Implications
alerting	pre-publication, speed, easy-to-use, subject-focussed
keeping up-to-date	cover <u>enough of specialist field</u> to know what is going on
searching	should cover <u>all of field</u> , cope with changing terminology and keywords, with choice of extension beyond field, especially for interdisciplinary work
browsing	give access directly to <u>article content</u> , including images, tables and mathematics
reading	provide mathematics and images clearly, allow different styles of reading
citation	cover all of publishing, not just one subject area
acquiring print	mechanisms for getting hard-copy that makes print useable for anticipated purpose

Ending

So far I have said very little about the content of journal articles. I have not mentioned how the introduction of electronic communication will lead to changes, just as the introduction of cost-effective colour printing in biology lead to a smaller length of text. I have not mentioned the experiments underway with video back-up for journal material by the American Astrophysical Society, nor the Journal of Protein Science's plans to place molecule rotating programs on the Net. Nor have I mentioned other key areas of the content., such as the mechanisms to maintain quality control in refereeing, which remain traditional, despite a variety of proposals for electronic alternatives dating back to the 1970's. Finally, in discussing coverage I have mentioned

the huge problems found in the technical limitations of the networks themselves, not least due to their nationally oriented development presenting poor international links for broadband services.

What I hope I have achieved is to have pointed to the way that technical possibilities offer great variety in how services are developed and yet, to be effective for the user must reconsider how to achieve consistency and coverage. I find this an exciting and stimulating time to be working in journals publishing and look forward to harnessing all the skills, expertise and imagination of the current players in a spirit of collaboration to provide an effective means of supporting scholarly communication.