
eLib Supporting Study
commissioned by UKOLN

THE IMPACT OF ELECTRONIC JOURNALS
ON LOCAL NETWORK COMPUTING AND
PRINTING

Bill Tuck and Maureen Grieves

[FINAL REVISION]

SUBMITTED BY
INSTANT LIBRARY LIMITED
DECEMBER 1998

CONTENTS	Page
Executive Summary	i-iii
1. Scope of the Study and General Comments	
1.1 Introduction	1
1.2 Background	2
1.3 Why do we Need Print	3
2. Document Encoding Formats	
2.1 Web-based Journals versus PDF Journals	4
2.2 Encodings for CD-ROM	6
2.3 Encodings for Document Delivery	7
3. Network Access to E-journals: Survey and Test Results	
3.1 Survey of Readers	8
3.2 On-line Tests	9
3.3 Survey of Libraries and Support Services	11
3.4 Survey of E-journal Suppliers	13
4. The Impact of E-Journals on Networks and Network Printing	
4.1 Introduction	15
4.2 Handling Large PDF Files in a Shared Environment	16
4.3 Network Printing Configurations	17
4.4 Test Results	19

4.5	Types of Network Printer	20
4.6	Estimates of Scale	21
4.7	The Need for Page Charging	22
4.8	Initial Results from JSTOR	23
4.9	Overview of the Impact of E-Journals	25
5.	General and Economic Issues Affecting the Use of Electronic Journals	26
6.	Conclusions & Recommendations	
6.1	Conclusions	28
6.2	Recommendations	29
7.	References	31
8.	Appendix A: Glossary of Acronyms	33
9.	Appendix B: Technical Notes	
9.1	Equipment Configurations Used for Testing	34
9.2	Test Samples	35
9.3	E-Journal Suppliers	36

Executive Summary

The principal finding of this report is that printing from electronic journals (e-journals) in standard formats (including PDF and HTML) creates few, if any, new technical problems that are not already inherent in print services in general. For central services, the main issue is providing sufficient capacity to keep up with demand and more details of this are given below.

As the pattern of use of computers moves away from being predominantly for the processing of user-generated files to the acquisition of online information from Internet or library-held sources, the demand for printed output will grow. More of the information that is now obtained by photocopying from print sources will become available online; this includes past exam papers, reserve lists and lecture notes, as well as e-journals and Web-based documents. This is likely to create serious pressure for additional printing capacity on central services.

While this demand for additional capacity can be met by technology already available in the marketplace, it is likely to require substantial investment in upgrading systems and increasing the ratio of printers to workstations on shared services. There is a well developed market for monochrome laser printers offering very good print rates, network connections and other features that are suitable.

Investment of the order required is unlikely to be justified without the parallel implementation of accounting and page charging mechanisms. Free computer printing is already competing with paid-for photocopying; uniform charges might balance the load and rationalise the use of computer printing. The absence of page charges and the ease of printing from the electronic form encourages computer printout rather than photocopy making the likely demand even greater. The development of accounting and charging systems appears to be proceeding on an individual site basis and some co-ordination across the whole HEI sector would be desirable. This has nothing to do with e-journals in particular, but with print management strategies in general.

User support and system maintenance are likely to increase to match the growth in network printing. Although improved tools for network monitoring should enable printers to be maintained more efficiently, the increasing complexity of print demands will mean more user support is required. Almost all e-journal material is now available in either PDF or HTML format and conversion from these to PostScript2 is straightforward, although downloading and printing from PDF is likely to be less complicated and more successful. PDF is the primary format for e-journals intended for presentation in print as well as on-line, but few, if any, printers handle PDF files directly. Restricting output to PostScript level 2 can not only simplify the problems of printer selection and updating and maintaining printer drivers, but much of the higher cost will be offset by easier management.

Moving from formats such as PDF to a raster image involves a considerable change in file size. There is a trade-off, therefore, between processing power (and cost) and network congestion. This was clearly demonstrated in our tests: cheap (or old) printers mean big files and large network delays. Printing delays occur at many stages of the print cycle. Much of

the delay is in the queue, rather than in data conversion or transmission, but it also depends on whether PS1 or PS2 format is used. Most e-journals in PDF form do not encode scanned page images (JSTOR is an exception), documents in 'native PDF' are one tenth the size and print faster. To improve performance faster servers are needed, more and/or faster printers and possibly a faster network. Increasing the number of printers on the local cluster to minimise queues is one option, but given that most of the cost of such systems is in management and maintenance, it may be more cost effective to replace a single printer with a faster one.

Simple HTML text documents present few printing problems, but multi-part documents where each component must be retrieved separately can be tedious and the risk of omission is high. Colour is an increasing component of e-journals published in HTML, but PDF versions of current print journals seldom use colour. Cheaper colour printing may change this, but for network laser quality printing, colour is much slower and could create serious bottlenecks.

Printing from most common image formats (GIF, JPEG, PNG) can be managed by most current Web browsers or PDF readers; more exotic file formats and the special software required by some e-journals may create problems. The proliferation of different interfaces, and the possible need for dedicated equipment, is clearly undesirable and many central services on shared access workstations do not allow such special software.

The demand for print can be a response to limitations on access and network congestion and it may be better to invest in improving access than to increase print services. More access terminals/workstations and faster connections, is one part of the solution; improved subscription and access rights administration is another.

The current load on local networks specifically from e-journals is likely to be small as usage is still relatively low. This may change as the swing from print to electronic sources grows. Few sites appear to have implemented higher speed local networks for links between workstation clusters and printers and it is not clear whether these are at present needed, or that they would greatly improve performance. With shared systems, the bottleneck is with printer congestion and queues on the print server; this is not specifically an e-journal problem but simply reflects the high current load on such systems.

Local or regional networks do not seem to be a problem, but there can be serious congestion on international (and even some national) links. High volume publications can be virtually inaccessible from remote (non-UK) servers and high-use UK servers can be difficult to access at certain times. Holding e-journals locally may be an option to improve response times, but at a significantly increased cost in storage and administration. A better strategy may be to mirror non-UK e-journals (principally from US publishers) on JANET sites or on regional sub-networks. This happens to some extent already, but a more co-ordinated policy is required. JSTOR which is now mirrored at MIDAS is an important step in this direction.

The general conclusion is that printing, networks and e-journals must be seen in the context of the larger problem of the transition from a print-based information economy to an electronic one. Many of the fundamental issues underlying this change are still under debate. While e-journals do not in themselves introduce any significant new technical

requirements into the printing environment, by potentially increasing the level of demand and the scale (including file sizes) they will show up any weaknesses or inadequacies. Trouble-shooting faulty network printers imposes a considerable cost on printing, over and above the equipment and consumables cost.

Present network print services were designed to handle user-generated material, rather than high volumes of non-user source material, such as e-journals; they could prove inadequate as demand for printed output from e-journals increases. A considerable increase could create delays, with individuals trying to finish their latest research paper being frustrated by the demands of those printing out network searches. There is little sign yet of any wholesale conversion from copying to printing, but a switch of say 25% could have a major impact. JSTOR is an important project in this context, as it is attempting to apply cutting edge technology to the problem of creating digitised archives of print journals. It is unlikely to be a major resource for undergraduate study, but it serves as a good test case of what is possible on shared network facilities. Our study indicated that network/printer installations may be able to cope, but only just. If demand for this kind of resource were to expand significantly then there would be little chance of satisfying it with the systems at present installed.

It is unlikely that the growth in demand that will follow from the switch to online information sources can be managed without page charging. Without it there is no constraint on how much network-sourced material might be printed out, which risks equitable access for other applications. Difficulties on congested systems, as queuing priorities push large files to the back, can already result in a two hour wait for output. Page accounting (and possibly charging) may also become necessary as a component of copyright management and as a means of providing access to non-subscribed material on a pay-per-view basis. This is a complex issue, both at a technical and policy level, and cannot be discussed in any detail here. There are print controls within the PDF structure itself. HTML, on the other hand, has no effective means of restricting access.

THE IMPACT OF E-JOURNALS ON NETWORKS AND PRINTING

1. Scope of Study and General Comments

The following report on the impact of electronic journals (e-journals) on local network and printing environments within the HEI sector, analyses the problem in fairly general terms and reports on some initial practical findings. It takes into account information collected via testing a variety of e-journal sources and from as wide a range of suppliers and support services (including libraries) as practicable. The study tries to look at the generic aspects of the subject, but realises that that with the diverse nature of the HEI sector much of the more specific information may not be pertinent to all types of institution.

Most of the technical evaluation was carried out at University College London. We are grateful to the Department of Physics at UCL for providing access to network services and responses to our survey on e-journal usage, to the library and JSTOR project for much useful information, and to the Information Services unit for technical advice on network printing. Many other individuals or organisations provided useful comments and suggestions, including printer manufacturers, e-journal suppliers and other academic institutions.

The report is divided into six main sections. After an introduction and discussion of some general issues, we look at the different document encoding formats for e-journals and related material. Next follows an account of our investigation into the problems of accessing e-journals across the networks, to see what kinds of material are available and what problems are encountered in retrieving and printing it. Network printing on shared resources is then considered to see what impact any substantial growth in e-journal use might have. Finally, some general issues, such as the economics of e-journals are considered, along with the conclusions of the report and recommendations for action or further study.

1.1. Introduction

The motivation for this study is the very practical question of trying to assess the impact upon printing demands - particularly within libraries - of the move towards provision of scholarly journals in electronic form. What problems will this pose, both technical and administrative, for the library and support services that will be called on to supply and maintain suitable equipment for this purpose?

A related issue is the networking demands that provision of e-journals will create. Network speed and capacity crucially affects response times, which in turn determine the effective usability of online resources. The relationship between ease of access, response time and print may be complex but, broadly speaking, it is likely that poor response times and/or intermittent access to electronic sources will increase the demand for a printed output. The more responsive the system, the more seamless the connection, and the easier the access, the less will the end user feel the need for the security of a printed copy.

What is the present situation? E-journal sources of fully refereed, high quality publications now number in the thousands and on this score initiatives such as the Pilot Site Licence Initiative (PSLI) must be reckoned as highly successful [1]. But how is this material actually

being used? Who is accessing it and by what means? Are users content to read 30-page articles on their high resolution screens, or are they printing them first? For one reason or another, the many recent surveys of e-journals and their use have generally failed to address these questions and it is at least partly to make some initial attempt at finding the answers that this study is directed.

1.2 Background

There are a large number of e-journals now available. Several surveys have been carried out [2,3] and a number of extensive journal listings are available [4,5]. The range of titles is growing exponentially, whether of electronic editions of existing print journals, or new electronic-only publications. A high proportion are peer reviewed and many are available only on subscription (except for sample issues or trial services). Serious scholarly publications may now be numbered in the thousands and all major publishers (along with many minor ones) seem eager to display their wares on the Internet or World Wide Web.

In addition to this substantial body of serious publication there is a growing mass of informal and pre-publication material such as: conference proceedings, e-prints, mediated discussion groups, working papers, personal Web sites. Formats and quality may vary, but much appears valuable.

Despite this remarkable growth in available material, however, actual usage by real researchers appears to be surprisingly low at present. Much of the evidence for this is anecdotal. For example, one of the heaviest users in the UK of the Academic Press IDEAL service recently reported that even in its busiest month fewer than 400 articles had been downloaded from its 175 e-journals [6]. Similarly, as an initial part of the present study, a short survey was made of the use of online publications by members of a highly active research group in physics. Of the 50 researchers contacted by questionnaire only 10 replied, and of these only a few had attempted to access e-journals on more than one or two occasions. None would claim to be regular users of such material. Although these results should not be taken as in any way definitive, they are a reasonable indicator that there may still be a problem in making the transition from print to screen.

In our survey, the principal reasons given for this distinct lack of enthusiasm were that:

- relevant material is difficult to locate;
- access and subscribing procedures are complicated;
- the range of titles relevant to any given subject is limited;
- retrieval can be painfully slow;
- on-screen browsing is uncomfortable and inefficient.

Any specific problems that may be associated with printing from e-journals must therefore be looked at in the light of these more serious difficulties. Research culture seems still to be dominated by print on paper. The preference remains for traditional materials and where these are reasonably well provided for, via well-stocked libraries, there is little incentive to tangle with the complexities and frustrations of the electronic version. Increasing costs and diminishing library budgets may eventually take their toll, but until there is real pressure for a cultural shift, coupled with a significant improvement in ease of use on the factors listed above, the pattern of use by practising researchers may only change at the margins.

1.3 Why Do We Need Print?

There are a number of familiar arguments as to why print might be a superior medium to the electronic for scholarly publishing - as well as equally familiar counter-arguments. These have been well covered in the literature and it is not our purpose to rehearse them again here (see, for example, the article by John MacColl in *D-Lib Magazine* [7]).

Rather, the question to be answered is, in respect to the great quantity of scholarly material now accessible electronically and readable on-screen, why is there such a strong need to print it out at all? Among the most obvious reasons is the fact that access to many e-journals can be painfully slow, even with supposedly fast connections into the local segment of the Internet. This particularly applies to e-journals hosted on servers in North America, where network traffic can create serious congestion at certain times of day – leading to the familiar complaint that “the United States ceases to exist in the European afternoon” [8]. National connections are often no better, either through inadequate servers or poor network topology. In other situations, connection time may be charged for, or subject to inactivity timeouts (such as with modem links). Online browsing of Web-based e-journals becomes impossible in these conditions.

The usual solution is to download the article (to either a local server or PC) and read it there. This can be a complicated procedure, however, particularly in the case where the article includes graphical or image inserts or is made up of multiple parts, perhaps including sound and video clips. The browser may require each component to be saved separately, which can be very tedious. External links may also be lost. In addition, there can be complicated questions of copyright and licensing.

Printing from the downloaded file can be even more problematic. Except for the simplest of HTML documents, structure is almost certain to be lost. Embedded URLs will not generally be printed, graphics or image files may not appear, and any hyper-textual links will vanish.

Even in the face of such difficulties, the imperative to print seems to remain strong. It provides security - against network inadequacies or the not unusual problem of how to find the material again once lost in the Web. It provides a kind of flexibility - the usual arguments of being readable in the bath or on the train. It also provides the satisfaction of ownership - as if in some sense possession of the printed copy were a substitute for reading. Printing out may also be necessary for certain types of activity, for example students printing out for course work, or researchers printing out when collating information, as this is still not easy to do on the screen.

If instantaneous access to the electronic version could be guaranteed, then printing would seem unnecessary in most cases. There are, in fact, some examples where access to electronic data is more or less guaranteed and instantaneous. The most obvious is the personal CD-ROM encyclopaedia. In this case, where the material is immediately available on the desktop, there seems little need to bother with a printed version of any retrieved information. The same is not true, of course, for library-held CD-ROM material, where the need to print out the results of a search for later reference, are even more imperative.

2. *Document Encoding Formats*

A related question is, if online e-journal material is always going to be printed out on retrieval, then what format should be adopted so as to make this as convenient as possible?

2.1 *Web-based Journals Versus PDF Journals*

At present, the principal contending formats for e-journals are HTML and PDF. Other possibilities include TeX, DVI, PostScript, and even plain ASCII text. For material that must be scanned in from a printed original TIFF (Tagged Image File Format) must also be considered. Specific word processor formats, such as MSWord or RTF, have also been used (a useful summary is contained in the paper by Wusteman [9]). There is a strong tendency now, however, to absorb all these 'non hypertext' formats into a PDF structure, either by direct conversion (in the case of TeX, DVI, PostScript or MSWord) or else by wrapping the file in a PDF envelope (in the case of TIFF and other graphics formats).

This leads to a neat categorisation of e-journals into those that are essentially print-orientated (and best encoded in PDF), and those designed to be read on-screen (encoded in HTML). It is the latter that we are calling 'Web-based', since HTML is the natural language of the Web, even though access to PDF files might also be via the Web (as, for example in the Academic Press IDEAL archive).

Web-based e-journals, encoded in HTML and designed to be read on-screen, are growing in number and include virtually all current electronic-only publications. But there are many that also have a printed analogue. For example, Project Muse (based at Johns Hopkins University) encodes almost all its e-journals in HTML, but only 2 of the 40 or so are electronic-only. One of these exceptions is a mathematics journal, for which HTML is still an inadequate encoding facility and PDF must be used. The others have a conventional printed analogue, which means that the publication stream must be adapted to create two quite different forms of output. In contrast, many dual publication journals (such as those from Academic Press, and most others within the UK site licensing initiative) are offered only in hard copy print or online PDF formats (although in the case of Institute of Physics Publications, an additional format of compressed PostScript is also available).

Not all print-originated journals are in PDF, however. The CatchWord company, for example, now offers a service to publishers to supply their publications via the network as e-journals, but it has adopted a proprietary format (called RealPage) rather than PDF. This needs the client to run special reader software [10].

An early example of this approach is the service offered by OCLC, which published one of the very first electronic-only journals ("Current Clinical Trials"). This originally required each PC reader to run the specially designed "Guidon" software. Their current approach is towards the use of conventional Web browsers enhanced, where necessary, with the capabilities of the Java processing language [11].

The ProQuest Direct service from UMI offers two different interfaces, one based on Windows, which is specially configured (and proprietary), and the other based on a standard

Web browser. In both cases a similar range of options is available, although it appears that there is greater flexibility in the search facilities of the special-purpose software [12].

A further example is the JSTOR project which combines standard Web browsers with a specially designed print facility, again based on proprietary software. The special problem here is the size of the page image files that make up the document archive. It is claimed that these are not so easily handled within PDF, although this is provided as an alternative [13].

The general trend, however, appears to be towards standardisation on either HTML or PDF, and using standard web browsers or the Acrobat reader, respectively, to access and manipulate the resulting file. Whether either will persist in the long term is, of course, debatable as shown by the comments from Ginsparg, McKnight and others in articles cited below [14,15]. Nevertheless, at present it is probably true that a very large proportion of e-journal distribution falls into one or the other of these two camps and the likely move is towards greater dependence on these as de facto standards.

Both HTML and PDF are well able to handle documents with multiple parts and different types of data - so-called multimedia - though each has certain strengths and weaknesses in this respect. However, since PDF is a print-oriented format (even though the Acrobat reader is designed for on-screen display) and HTML is screen and hypertext oriented, it would seem reasonable to expect that printing from PDF files might be easier. Our tests did, in fact, show that, except in the simplest of cases, downloading a journal article in PDF and printing the file is generally less complicated and far more likely to produce successful results than trying to print from a linked collection of HTML files.

Images (or 'bit-maps') can be included in both PDF and HTML. Web-based material will generally include images as GIF or JPEG files within the HTML structure. Both formats (GIF and JPEG) were designed specifically for encoding images for rapid transmission over networks. GIF (Graphics Interchange Format, a proprietary standard developed originally by CompuServe) is particularly suited to the encoding of line drawings or cartoon-like images and special characters, while JPEG (an ISO standard developed by the Joint Photographic Experts Group) is best for colour photographic images. A new standard called PNG (Portable Network Graphics) is currently being promoted by the standards group of the Web (the so-called W3C - World Wide Web Committee) as an alternative to GIF. All standard Web browsers should be able to handle the three formats - GIF, JPEG and PNG.

TIFF, a standard developed in the 1980s by Aldus Corp and Microsoft, and now promoted and maintained now by the Adobe Corporation. It has its origins in the document image processing world, where scanned images of text documents are created as digital files for storage and manipulation within a desktop publishing or office automation system. The basic standard has since been extended to include a wider range of data types. It finds its role in e-journals where an archive of back issues in print-on-paper must be converted to online form, retaining the exact layout of the original. Efficient compression techniques, such as the so-called Group4 fax compression, can be incorporated within the TIFF structure, and the whole can be wrapped into a PDF file, including additional information that may be encoded as text rather than bit-map.

2.2 *Encodings for CD-ROM*

A similar classification might also be made for source material encoded on CD-ROM. Some publications, such as the Encyclopaedia Britannica, for example, are encoded in HTML and designed to be read using a standard Web browser. Others will be in PDF - Electronic Books, for example has produced a series of CDs of standard literary and scientific works encoded in PDF.

Alternative formats for CD-ROM data do exist and are used extensively. Statistical or bibliographic datasets, for example, will generally adopt a highly structured database format designed for rapid searching and retrieval. In the case of encyclopaedic or free text material, however, there appears to be a move away from proprietary encodings towards adoption of standard access software based on a Web browser (Netscape or Microsoft Internet Explorer) or the Acrobat PDF reader.

The ADONIS system, for example, which supplies page image encodings of a large number of biomedical journals on CD-ROM (by scanning the printed originals), has adopted a form of compressed TIFF which enables a very high packing density of some 10K pages per disk. This corresponds to around 100 journal issues per disk, from the 850 titles provided. Because the encoding is proprietary, the disks can be read only on a special purpose workstation.

A similar technology is used to store the patents documents supplied on CD-ROM by the European Patents Office, although in this case the page images are in standard compressed TIFF format (using Group 4 fax compression) and are accessible without proprietary equipment.

TIFF is an open standard, full specifications are made freely available and some scope exists for influencing the future directions of the standard as it seeks to include a wider range of image types. The TIFF encoding scheme also includes compression facilities (such as the Group 4 fax specification for text, or JPEG for colour images - both of which are ISO standards). To meet the needs of the printing (and pre-press) industry an extended version of TIFF 6.0, called TIFF/IT (where IT here stands for 'Imaging Technology'), has been developed and has now been accepted as an ISO standard (ISO 12639).

Because TIFF is so rich and variable it is difficult to ensure that the recipient will be able to read or print it. For this reason it may be better to encapsulate TIFF files into a PDF envelope before distribution. The overhead is relatively small in terms of additional processing (and the file sizes will be only marginally larger) but the advantages could be very great. This approach has been adopted by the PPT project, in which early issues of the Transactions of the Institute of British Geographers (TIBG) are scanned digitally, with the resultant TIFF file being compressed and then converted to PDF [16]. Later electronic issues need not be scanned first before being converted to PDF. This leads to a unified database with all documents in the same format. Even pre-print archives, such as that at CERN, are moving towards adopting PDF as a uniform way of encapsulating all kinds of input material, from scanned images in TIFF to word-processed TeX files or PostScript-encoded articles. Only one reader is then needed to access all information in the archive – or, rather, two, since HTML will still be needed for the Web interface.

2.3 *Encodings for Document Delivery*

Until recently, all document delivery services were based on the photocopier. Libraries (such as the British Library Document Supply Centre) would make a photocopy of an article on request and deliver it by post (urgent services used fax for delivery). By far the largest proportion of document delivery is still photocopy/post based.

Recently there have been moves to shift the service towards digital scanning and electronic delivery via the Internet (either using FTP or email). The Ariel system devised by RLG was the first to become widespread and is now used in a number of document delivery services, including LAMDA within the UK HEI sector.

Ariel is a software suite that enables a PC, scanner and printer to be assembled into a document transmitter/receiver able to link to any other Ariel unit via an Internet connection. Documents are scanned as TIFF images and transmitted, along with some header information (the so-called GEDI header) giving bibliographic and request details, to a similar Ariel system at the requesting site. The Internet File Transfer Protocol (FTP) is used to transfer the data, although more recent versions (including the JEDDS and EDDIS projects) will use email. The principal disadvantage here is that transmission is between Ariel workstations, rather than to the end-user's terminal.

Several document suppliers are believed to offer delivery to Ariel workstations, but do not themselves use this workstation configuration for document input as more efficient systems can take advantage of possible economies of scale at the input end.

Dependence upon Ariel at the receiving end could be most easily avoided by again using PDF to encapsulate the TIFF image of the scanned document. This could also include a simple text file for the header information. The advantage is that any standard PDF reader, such as Acrobat, should then be able to read/print the document. There are a number of current projects exploring this as a standard for document delivery, including VirLib in Belgium [17]. Users typically want to be able to request and receive documents directly from the PC on their desks. However, this is unlikely to be an Ariel-configured machine and printing documents received as raw multi-page TIFF files can present considerable difficulty to the average user. Although tools to handle this format do exist and are freely available, there is no guarantee they will work on every TIFF file. Also, unlike Acrobat, they are not a standard part of every PC's toolbox.

TIFF files can be very large. In compressed form, text will normally run to between 50K and 150K per page, depending on the print density (and assuming a scan resolution of 300 dpi). Images or background colour could increase this substantially. Converted to PostScript (Level 1), each page image explodes to a fixed size of around 2Mbytes, a not untypical document of 10 pages then amounts to more than 20Mbytes of data (see the Example 1 below). Files of this size can create serious problems of congestion on local networks and have a very negative impact on printing services [18]. Considering that a large research university might receive over 100 articles per day through Inter Library Loan requests, the printing problem could be significant. If the average print time was 30 minutes (as indicated by some of our experiments), then around 50 hours would be needed! To be practical it will

be necessary to reduce print times to under 5 minutes per article, or 30 seconds per page, and although this print rate seems fairly modest, not all machines will be able to meet it. Level 2 PostScript is more efficient in its coding of TIFF images as it allows them to be transmitted to the printer in compressed form. The encapsulation of TIFF images into a PDF structure retains the compression feature – which allows such files to be transmitted economically before being converted to PostScript or other format for printing.

3. Network Access to E-Journals : Survey and Test Results

In order to try to understand the impact of greater availability of e-journals, brief surveys were carried out among the three main groups concerned - readers, library services and suppliers. Because of the limited time and resources available for this study it was not possible to do much more than interview a rather restricted sample from each group and collect information from relevant mail-lists. The intention was to identify any major problems that might exist and to suggest ways in which the eLib programme in general, and the site licensing initiative in particular, might try to resolve these issues. It is hoped that the results given here will provide a basis for further discussions, particularly with service providers, including libraries, document suppliers and technical support departments.

3.1 Survey of Readers

As mentioned earlier, our attempt to discover who was using e-journals (and, perhaps more importantly, who was not) was somewhat inconclusive. Librarians in general indicated that demand was difficult to assess accurately, but appeared to be relatively light at this stage. A short questionnaire distributed among a group of 50 research scientists elicited a poor response and few claims to any consistent use of e-journals. This was disappointing as well as somewhat surprising, since the number of e-journals available in this area is relatively large and the approach has been quite heavily promoted by specialist publishing groups, such as Institute of Physics Publishing and the Institution of Electrical Engineers. This clearly needs further discussion both with suppliers and libraries, as well as potential end-users.

Much heavier use appears to be concentrated in narrow subject specialisations, where the provision of e-journal source material may be filling a gap in the publishing market. One example is musicology, where a proliferation of technologically adventurous, high quality, electronic-only publications has shifted the balance away from conventional print. It appears that in subject areas where few printed publications exist, either for reasons of cost (for example, for colour images) or the technical difficulty of carrying essential information in this medium (for example, sound clips), the adoption of e-journal technology can meet with greater enthusiasm than in areas which have long been adequately catered for by conventional print publications.

Pre-print archives are another area where strong sectional interests have been able to establish a major new service with an apparently high level of demand. The Physics archives at Los Alamos and CERN are examples that may well migrate to other subject areas.

What problems readers face in accessing and printing from e-journals is more difficult to assess. Our best guide here may be the questions presented to support personnel in libraries and computer services. On the other hand, one gets the impression that the most active use of e-journals is on a do-it-yourself basis, with technical support provided by the publisher or supplier through the use of FAQ lists and email response pages.

3.2 Online Tests

In this general spirit of do-it-yourself, the present study examined a wide range of e-journal offerings (along with pre-prints and CD-ROM material) to see what difficulties they posed in retrieval and printing. A selection of articles was made that represented most of the variables likely to be present in current e-journals: PDF and HTML, scanned page images, background 'watermarks', colour image illustrations, mathematical symbols. In addition, a range of archive sites was chosen that might be representative of real access conditions: local 'London universities' MAN network, UK JANET, commercial UK Internet, US Internet, using both modem and high-speed network connections.

3.2.1 Test Samples:

The full details of the e-journal documents that were used as test samples are given in Section 9. They covered most of the document formats likely to be encountered in e-journals and served as a test base for a comparative study of retrieval and printing times at different sites and with different equipment.

3.2.2 Summary of Test Results

This test set of journal articles was accessed and printed from a variety of locations and equipment. More details of the test results, particularly those relevant to networks and network printing are included in Section 4.4.

The least satisfactory was a combination of telephone line access (using a 14K bps modem), 486 PC and PostScript level 1 printer. Downloading times on this equipment were, not surprisingly, rather high and several of the test articles failed to print. Nevertheless, even on this configuration a high proportion could be successfully captured and printed.

By contrast, probably the most satisfactory overall was a Pentium PC with 56K modem and colour inkjet printer (a combination costing well under £1000). This was easier to use and gave better print output than several library-based or other publicly accessible machines - largely because downloading times are very time-of-day sensitive, and shared printers are not always in the best condition (presumably through over-use or poor maintenance) or easy to access for long print jobs.

Work group arrangements, where a limited number of users share an online access point and a locally networked printer, were also generally satisfactory (provided the printer was of adequate specification - PostScript Level 2 laser or modern inkjet). The problem here is the need to share print facilities, where time-consuming jobs can be anti-social. Also, while network access is generally better for material held 'locally' (such as on the London universities network) this is not always the case for remote (for example US) archives,

where time of access can be significant and unsocial hours (for example weekends, or early morning) can give a much better response. The variation in response time can be quite marked, in requesting the same material from the same server but at different times.

For printing, the serious problems are of two kinds. The first is generally due to use of obsolete equipment - hardware or software. This can make printing difficult or impossible. For example, the PDF encapsulated scanned document (Example 1) expanded to 24Mbytes when converted to PostScript (level 1), which overwhelmed the printer (presumably through inadequate buffer or processing space). The 'watermarked' page of Example 5 did eventually print, but required 90 minutes of processing (compared to the still unacceptable 5 minutes for a modern personal inkjet printer – indicating the undesirability of using fancy embossed background cover pages for e-journals!). Early versions of Acrobat also occasionally had difficulty opening the PDF document.

The second general problem is due to a mismatch between the screen layout of multi-part HTML documents and their equivalent on the printed page. Images are frequently detached from their textual descriptors, or may be split across pages. Some browsers (including Microsoft Internet Explorer 4) require images (and other inserts, such as sound or video clips) to be saved separately on downloading, which can be extremely tedious.

Text quality was found in almost all cases to be very satisfactory - even on supposedly obsolete PostScript printers (provided the cartridge was reasonably new). Image quality depended, on the type of printer and its condition, but very good images were obtained on even quite modest equipment. The problem of handling page image files on early PostScript printers is a tricky one since in all other respects they may perform very well, making it difficult to justify their replacement.

The increasing use of colour in e-journals is a factor that may create problems for libraries in the future. While cheap personal printers are able to handle colour quite effectively, the capital and running costs for high volume colour printers could be prohibitive for many institutions.

Unlike printers or other hardware, there is little justification for not replacing out-of-date software for browsers and PDF readers. These are freely obtainable and easily installed. The current standard is Acrobat 3.0, which is well supported by Adobe and has plug-in capabilities for compatibility with other industry standards such as NetScape and Internet Explorer browsers.

Even with good equipment, image printing can be very time-consuming. This is less of a problem on a personal printer than on shared equipment, where jobs may be queued and competition for resources can be intense. Priority scheduling is clearly desirable, but is not easy to implement. Ideally, such scheduling would be built into an automated charging system for printed output, but this is itself a complicated issue for which satisfactory solutions are still being sought.

3.3 *Survey of Libraries and Support Services*

A number of libraries were questioned on their policies and problems with respect to printing from e-journals and other networked resources. Several are now adopting a policy of removing PCs from the library itself, on the principle that the Information Services unit (or Computer Services) is better placed to manage the actual equipment, usually as a set of distributed clusters of workstations, each cluster with its local print facilities. The library is then able to concentrate on the problems of managing subscription licensing and access rights.

Print demand in general is very high, but the proportion that may be attributed to e-journals is at present probably quite small. Unfortunately, it is difficult to establish any realistic measure of the level of demand for e-journals from the reader's end, as no record is kept of user access. Suppliers are believed to have figures, but are reluctant to reveal them.

The demand for access to shared workstations, as registered by utilisation rates, is very high. Rationing access is likely to increase the demand for printed output from online resources (since it cuts down on screen reading times). On the other hand, such rationing is also likely to decrease the time available for access to online resources, since this type of use must itself compete with other kinds of application (word processing, data analysis, etc.). Only with unrestricted and unlimited access to workstations and network connections is there likely to be a significant rise in utilisation of e-journals.

Research staff (and students) are usually in the position of having relatively unlimited access. The evidence from our informal survey indicates, however, that utilisation of e-journals even by this group remains relatively low, suggesting that there may still be significant inefficiencies in the access system. Time is the limited resource - online access must compete with other types of activity for the limited time available to researchers. If it is quicker to go to the library shelves, they will do so.

The greatest source of such inefficiency is probably the difficulty of finding and gaining access to online material. Although the PSLI includes some 400 journals in its sources, this is still only a small percentage of the total number of e-journals available (which, in turn, is but a fraction of all scholarly journals). Furthermore, most parallel publications are only available on subscription, which creates another barrier to easy access.

Access problems are still the main issue. For the library this means finding ways of managing the subscription rights to a burgeoning array of suppliers, each requiring some kind of registration of all the institution's users so that they may be allowed access to the supplier's document server. Subscription management, user registration, access rights and authentication are the major issues.

What the users notice, however, is the poor response times to document servers and the lengthy queues at the printer. To improve response times means getting more bandwidth onto the network and better network topologies, including the organisation of mirroring and caching. The latter are technical and engineering issues that we have barely begun to address [19]. But they also carry implications for the library and its suppliers – subscriptions for e-journals involve real money, and there needs to be a balancing level of quality

assurance on standards of service. If a publisher or supplier fails to install adequate processing power on the server or network bandwidth, then the quality of service will decline. To enforce quality of service agreements will require that response times be monitored, which itself is not a trivial technical or administrative problem.

Printing, by comparison, is a much more tractable issue. Assuming standardisation on a limited range of document file formats can be achieved (to include PDF and HTML, say, but not necessarily raw TIFF or more exotic image types), all that remains is the need for a charging mechanism. The capital cost of printers have declined considerably, but running costs (paper, ink and maintenance) are still relatively high and it is inconceivable that a print service can continue to be run on a no charge basis, given the potential level of demand.

Many institutions charge for laser printing and/or ink-jet printing, in some cases combined with free dot matrix printing. Free dot-matrix services can be heavily used, but there are significant problems in control and quality, particularly with full-text services and downloading to floppy disc for printing elsewhere is encouraged. Charging, may recover some costs, but incurs others such as additional software/hardware and/or additional staff time, a variety of charging methods are used. The need for charging is also discussed in Section 4.7.

Card systems such as EMOS, which may be hard-wired into the printers. Additional software such as Qview and Qmeter is required to manage queuing when using in a networked environment, with users directing print requests to a print server attached to a laser printer controlled by a card system. Where networked facilities are not offered, material may be downloaded to disk and then used on a dedicated PC attached to a card controlled printer, or a card controlled printer may be attached to each PC, but this is an expensive option. If fitted with the same charging units the same cards can be used as for photocopiers and microform reader/printers; being programmed to charge at different rates according to the type of equipment used. The complications of lots of machines using the same laser printer via a card system has led to this being discontinued in at least one institution. Even where there is a charging system readers may be encouraged to download if possible to save overloading the system.

Accounts. There are various methods of handling accounts, but a print account does imply a universal ID system within the institution, which is not always the case, and may deter casual users. Some accounts are linked to computer username, the account must be in credit before printing can be done, software such as the BILL system is required to manage the accounts. Other accounts consist of a print budget system linked to ID cards. New print management systems linked to computer accounts, which must be in credit, are being considered or under development. For example print jobs are sent to a central depository and PC/printer stations are used by swiping their ID card to give access to their own queue of print jobs and print out any required jobs. The decisions on ID cards and computer user names essential for such systems may be outside the control of the library and some accounts systems are operated by other support services such as computing services.

Direct Payment. There are a variety of methods used. Users may send their print job to a print server, which is controlled by Qmeter and then have to pay staff who release the job. Although this uses staff time it does save on the costs of installing card systems.

Charging for paper, or expecting users to provide their own paper. At least one institution regarded this as a disaster, but others indicate that although it causes printer jams students do not seem to mind losing sheets of their paper to other users. Where users provide their own paper this can lead to more wear and tear with printers opened up every few minutes, and even more paper jam problems. Both methods can incur heavy staff costs in return for little or no revenue. The only advantages seem to be the system is simple to run, no investment in card systems or other software is required and it may encourage users to download and use their home printers.

3.4 *Survey of E-journal Suppliers*

The range of e-journal suppliers is almost as wide as the range of e-journals themselves, from large commercial publishing houses, to learned societies and university presses, down to small groups of a few enthusiasts with one or two titles. The entry cost to Web publishing is now so low that there is little constraint to any interest group creating its own journal.

The survey of suppliers was necessarily restricted and the list below is in no way comprehensive. They were chosen to cover a reasonably wide range of the different kinds of service available. The survey was not intended as an evaluation as such, but rather to see what different approaches there might be to similar problems and how relatively successful these might be.

The suppliers examined were: Academic Press - IDEAL; Institute of Physics; UMI ProQuest Direct; Johns Hopkins University – Project MUSE; BIDS - JournalsOnline; JSTORE; PTT project; CatchWord; CERN Pre-prints.

Brief details of the various suppliers are listed in Section 9, giving the following details: Name of publisher or service; Type of service offered; Number of journals and major subject areas; Document formats used; Location of servers.

How well do these services operate? Is access so easy and the response so fast that users will feel little need to print out the e-journals but will happily read articles online? And does printing itself cause any problems, either for the user or the library support staff?

These are the kinds of question to which we would like to find answers. But what criteria is it reasonable to adopt as a measure of quality of service? It is generally accepted that to compete with print, an online service should give a 'page flip rate' of under one second [20]. We are clearly quite some way off this at present. In any case, many e-journal services are based on the assumption that the article will be downloaded first before being read on-screen via a PDF reader. At the other end of the scale the alternative delivery service for scholarly publishing involves a visit to the library, physical searches through paper journals and an indeterminate wait at the photocopier. Against this even the slowest of e-journal services may seem speedy. The following are therefore no more than a series of random comments regarding the relative merits of the strategies adopted by the suppliers listed above.

One obvious problem area is how to deal with back-runs of traditional print journals, or

material that is only available in print format. JSTOR, PPT and CERN Pre-prints all provide access to scanned input from paper originals. Of these PPT was by far the easiest to use. This is partly because the server is located on the London MAN, providing a very fast link to our chosen test sites. In addition, all material (whether scanned, OCR'd, or electronically generated) is encapsulated in PDF. A standard Acrobat 3 Reader had no problem displaying and printing the documents - the speed and quality of print being determined only by the printer. In contrast, both JSTOR and CERN are based on remote servers with relatively slow links into local networks. Coupled with a high access demand on the server, this makes transfer of the relatively large files involved (around 1Mbyte per document) very slow. In addition, neither employs standard Acrobat/PDF as the principal print control mechanism (the 'helper' print program used by JSTOR needs a whole page of instructions for the user to link it into the browser, while CERN uses TIFF and PostScript). The net results are difficult systems that only dedicated users with a high degree of technical know-how (and spare time) would willingly use - that there are such users may be a measure of the value attached to the source material, rather than the simplicity of the operation.

Speed of access (or lack of it) is probably the most noticeable factor in all e-journal services. While network speeds will undoubtedly improve over time, other strategies may still be needed if an optimal service is to be achieved. Site mirroring is perhaps the most obvious approach. Significantly, JSTOR established a mirror site in the UK in March 1998 (hosted at MIDAS and available through CHEST to all UK academic sites on subscription). Several of the other suppliers listed above already operate multiple mirror sites. CatchWord may well have gone furthest in this direction with some 12 mirror sites at various strategic global locations. Even CERN is to some extent a copy, if not a mirror, of the Los Alamos pre-prints archive.

Mirroring entails cost. At some point a balance must be struck between the cost of replication and the likely benefit in terms of performance. This is partly a question of granularity - should every library replicate the universal document store? It is also a question of transparency - why should a user need to know where the optimal location is to be found? There are technical and economic issues here that we have only just begun to address. Caching is a related concept, and generally means storing in some fast-access location in anticipation of demand.

In these terms, printing from an e-journal can be viewed as a caching strategy and should therefore be considered within this general context. Improved access by replicating the server reduces the need for print caching, but at a cost. An obvious intermediate strategy is for users to build their own electronic archives of relevant material (much as they do now in collecting photocopies of interesting articles). This makes considerable sense at the level of a research group, for example, where storing a few thousand PDF documents on the local workgroup server presents no difficulty. In fact, it would be surprising if it were not already happening. The problem for libraries and publishers is that there is no economic or legal/copyright model to accommodate this kind of use.

Another problem with some current e-journal services is the need for special viewing/printing software. JSTOR, CatchWord, ProQuest Direct and CERN Pre-prints all have this as a requirement, or at least as an option. The reason is again that better performance can be obtained by using optimally designed software, rather than a general-

purpose browser/viewer/printer such as Netscape/Acrobat. Unfortunately, support for non-standard software creates serious problems for the library or computer services department, as well as for the individual. Allowing the user to install software or 'plug-ins' downloaded from arbitrary Web sites is problematic on shared workstations. The acceptable solution will come with the introduction of dynamically configurable browsers. Already CatchWord has rewritten its RealPages software in Java to take advantage of these facilities in the latest versions of the standard Internet browsers, and others will surely follow if they have not done so already.

4. The Impact of E-journals on Local Networks and Network Printing

4.1 Introduction

The earlier part of this study focused on the problem of getting access to e-journals and on the difficulties imposed by network limitations such as bandwidth availability and the establishment of access rights. The printing of retrieved material on a locally connected printer showed very few problems - provided only that the PC and printer hardware were of reasonable specification. The more difficult issues concerned the problem of locating material and establishing access rights - both of which were somewhat outside the terms of reference of this report.

A second stage of the study has concentrated on the specific issues of e-journals in the context of a network printing environment, with shared printers and workstations in a variety of configurations and with different underlying operating system software. This is the typical arrangement for general student access to computer and information services in most HEIs. A cluster of PCs (or Macs) are allocated for student use on a time-shared basis, along with a file server or host processor and one or more printers. It may be different for research staff and faculty, where workstations are often permanently allocated to individuals.

While e-journals do not in themselves introduce any significant new technical requirements into the printing environment, by potentially increasing the level of demand and the scale (including file sizes) they will show up any weaknesses or inadequacies in current services.

Adobe PDF is the primary format for e-journals intended for presentation in a print as well as an on-line format. The alternative, HTML, is used for many e-journals but is not primarily intended for print output. For this reason, in spite of the fact that HTML creates some interesting print problems of its own, the present section is focussed on PDF.

Proprietary e-journal formats, such as CatchWord, may also offer print-orientated output along with increased on-screen functionality (in this case based on Java). Unfortunately, this is likely to further complicate the problem of setting up a consistent networked access environment. For this reason, the study will not attempt any detailed investigation of network printing from such sources.

An exception to the latter restriction is JSTOR. The recent installation of the JSTOR database as part the MIDAS services makes available to the UK academic community a substantial collection (> 50) of journal titles. While these are available as scanned images in

PDF, the facility also exists to print them in a higher resolution using a special print facility called JPRINT. This kind of material makes considerable demands on networks and network printing technology. It therefore serves as a useful test case for the capabilities of present systems and was therefore investigated in some detail.

4.2 Handling Large PDF Files in a Shared Environment.

Our earlier studies showed that there were very few real technical (as opposed to administrative) problems in accessing, retrieving and printing e-journal material supplied as PDF files, when the environment consisted of a self-standing PC with Internet link and local printer. The only constraints were the capacity of the link and the adequacy of the PC and printer hardware. The Adobe Acrobat 3.0 Reader runs very robustly under most common operating systems, including Windows 3.x, Windows 95 and Mac. Integration with a full range of Windows printer drivers makes printing from within the package very simple. Printing from the Apple Mac is generally via PostScript, which is again unproblematic (though finding drivers for other types of printer can be a difficulty).

The Adobe Acrobat Reader is freely available and easy to install. The software package is large, however, and it is not therefore advisable to download it from a remote Web site over a slow Internet link, unless that is unavoidable. Many commercial ISPs provide Acrobat as part of their standard CD-ROM utilities package.

In a shared environment, it is likely that the Acrobat Reader will be downloaded on demand from a file server to the local workstation/PC. Integration of network printing from within Acrobat may be more problematic to set up. In some cases a manually controlled two-stage procedure of first 'printing to file' then sending the output data stream (usually a PostScript or PCL file) to the printer queue, may be necessary. From this point of view, any problems are those of network printing in general, not PDF or Acrobat (although care may be needed to ensure that users know to print as Level 2 PostScript rather than Level 1, since both options are likely to be available in the Acrobat Reader).

The problem, in other words, has reduced to that of printing a (often quite large) PostScript (or PCL) file. The high degree of standardisation of such file formats should mean that there is no intrinsic difficulty (provided the printer itself has an adequate specification). The only problems will be those inherent in network printing itself. Unfortunately these are not always trivial.

Few, if any, printers will handle PDF files directly. The de facto standards for the data stream sent to a laser printer are Adobe's PostScript and Hewlett-Packard's PCL. All other document and image formats must be converted into one or the other before transmission. In the past, there were some printers able to accept additional formats such as compressed TIFF (for example, in the early QMS range), but with the emergence of level 2 PostScript (which includes a compressed TIFF mode) this became unnecessary. PCL, likewise has evolved to handle an increasing range of data types efficiently.

Responsibility for conversion to either PostScript or PCL now lies with the application package (Acrobat, MSWord, WordPerfect, etc) and the printer drivers embedded in the workstation or PC software (or downloaded from the network server). In an environment

such as Windows 95 (or 98) the production of a PostScript (or PCL) output file from the PDF-encoded document is usually very straightforward and unproblematic (and the same is true for Mac applications to PostScript).

For shared network printing, since a PostScript facility will inevitably be needed, a good case may be made for restricting output to PostScript alone. All applications can then be mapped to the same printer-independent output format, simplifying the problem of updating and maintaining printer drivers. The downside is the higher cost of PostScript-enabled printers, but this may be offset by easier management. There may also be a slight penalty in slower performance, although on current hardware this should be negligible. Another advantage is that both Mac and Windows workstations can be administered in the same way, without the need to find special printer drivers (sometimes a problem with Mac).

4.3 Network Printing Configurations

Some measure of the complexity of network printing is indicated by the report that approximately 80% of LAN management effort is devoted to trouble-shooting faulty network printers [23] or that “from 40 to 60 percent of all end-user calls to network administrators are for printer-related problems” [24]. This imposes a considerable cost on printing, over and above the already quite high equipment and consumables cost. Anything that would reduce this overhead would clearly be very desirable.

Unfortunately, it appears that we may still be some way from a universally available solution to this problem. No common platform for the management of printers from different vendors and over different networks currently exists [25]. For specific network operating systems, however, there appears to have been considerable progress in persuading printer manufacturers to bundle their control and management software within the Operating System.

While a full discussion of network printing is well beyond the bounds of this study, it may be helpful to review some of the principles that are involved. At the lowest level stands the hardware - workstation or PC, file server, print server and printer. These may be arranged on the network in various configurations:

- (i) printer connected directly to the user's workstation or PC
- (ii) printer linked via direct parallel connection to the file server
- (iii) printer on a separate PC, which is in turn connected to the network
- (iv) printer linked directly to the network

In shared resources, the trend is towards the third option. One reason is the limited distance over which a standard parallel printer connection can operate, coupled with the need to isolate the file server from the workstation cluster (for security reasons). Placing the printer directly on the network allows it to be located adjacent to the workstation cluster, but remote from the file server or host running the print management software.

The print management software is responsible for handling the printer queues and for any additional data conversion that may be necessary, as well as dealing with any error conditions that might arise. In addition, it would be responsible for any accounting and page charging that may be implemented (though this is a separate and more complex issue).

There are two principal points where data conversion must take place: from PDF (or other application format, such as MSWord, TIFF, HTML, etc.) to PostScript (PS) or PCL; then from PS or PCL into the raster image or bitmap from which the final page image is created. Where these transformations are carried out - PC, file server or printer - will be determined by the type of hardware available and the network configuration. Laser printers generally come equipped with considerable processing power and internal memory. They are designed to take in PS or PCL files and convert them into raster images. The alternative strategy is to create the raster (or a close approximation) on the PC itself, which allows for a cheaper printer. This is the strategy adopted with low cost personal ink-jet printers.

Moving from an application format (such as PDF, or Word) to a raster image involves a considerable change in file size. For example, a 10 page word processor file of 20K bytes may result in a PostScript file of, say, 100K bytes, which in turn will generate 10 raster images (one per page) of around 1 Mbyte each (assuming a print resolution of 300 dots per inch).

Clearly, the larger the file the longer it will take to transmit across the network. There is a trade-off, therefore, between processing power (and cost) and network congestion - simple printers need large files to drive them; conversely, more powerful printers correspond to smaller transmitted files.

This is clearly demonstrated by the difference between PS level 1 and level 2. Our test document of 10 scanned pages (from the TIBG database referred to earlier) generated a file of 24 Mbytes at PS level 1, but only 2.1 Mbytes at level 2 (the original PDF file was 1.4 Mbytes). Bitmaps in PostScript level 1 are encoded very inefficiently, with each pixel represented by two bytes of information (corresponding to the position co-ordinates of the pixel). Converted into the format required to drive a LexMark inkjet printer, on the other hand, it generated 8.8 Mbytes. The conclusion is that cheap (or old) printers will mean big files and large network delays.

The problem is that, to some extent, the two factors work in opposition: to minimise transmission times, files should be as compact as possible, but this increases the raster generation time at the printer. To compensate, the printer will need a more powerful processor and larger memory (at correspondingly greater cost). It has sometimes been suggested that a better strategy might be to generate the raster image on an independent dedicated processor and use cheaper printers, at the risk of increasing network congestion. This approach appears to have been adopted only in a few special situations, however, such as to optimise the load on high-end printers.

The general model allows us to begin to understand where the delays in printing might occur. First is the conversion time from PDF to PS; then the transmission time to the printer queue. The waiting time in the queue will depend upon a number of factors such as queue

length and the priority scheduling that has been adopted (are small jobs given a higher priority at the expense of larger ones?). Next comes the transmission time from print queue to printer, and finally the time for conversion into a raster image and its transfer from drum to paper.

4.4 Test Results

Ideally, an evaluation of network printing strategies would examine different configurations and measure both overall delays (from workstation to print) and individual delays at each step in the chain, for a range of test documents. It would also be desirable to consider at least the three principal network operating systems (Unix, Novell and Windows NT) as these differ to some extent in the facilities provided for the management of network printing. This is clearly not realistic in the time available. Simple tests were carried out, however, on two different network configurations, and with a range of document types. The first network was a cluster of Mac workstations. The second was a cluster of PCs (HP Vectras with Pentium I processors running Windows 3.x). Each cluster of around 20 workstations was linked to a single Hewlett-Packard Laserjet4M laser printer (one networked printer per cluster). In both cases the network operating system was Aix 4.1 (a variant of Unix) and the underlying network a standard 10M bps ethernet, running TCP/IP.

To get a measure of the 'worst case scenario' the document used was the ten page scanned PDF document referred to earlier (the TIBG article from the PPT project archive). Retrieval time from the archive (at QMW) for this 1.4 Mbyte file was 30 seconds on the first network and 1 min 30 sec on the second (probably reflecting different network loadings on the two different days).

The overall delay in printing this document showed considerable variability, of between 5 and 25 minutes, even on very lightly loaded networks. Most of this delay is in the queue (waiting for other documents to print), rather than in data conversion or transmission. It also depends critically on whether the file was converted to PS1 or PS2 format. As explained above, PS1 generates a very much larger file, which is likely to be held back by any reasonable prioritisation rule imposed on the queue.

On a stand-alone Pentium PC the conversion time for this document from PDF to PS level 1 was measured at 33 sec, while to PS level 2 it was merely 8 sec. On the networked machines, however, these conversion times tended to be much larger (presumably through the need to interact with the server - or the conversion itself may take place on the server). PDF to PS1 took 7 minutes on the Mac, while from PDF to PS2 took around 1 min on both Mac and PC.

The transfer time to the printer queue of the PS2 file (around 2 Mbytes) was around 2 minutes, and the document began to print after another 4 minutes (assumed to be time spent in the queue). At the printer, PS1 was considerably slower to process than PS2, though even the latter took around 2 minutes to print the 10 pages.

There are too many uncontrolled factors to be very precise about any of these measurements, and too little data to generate any very meaningful statistics. Nevertheless,

these rough experiments indicate that the expected time to print a document of this size and type on such a network cluster would generally be 10 minutes or greater (and often very much greater).

This is a 'worst case' example. In practice, most e-journals in PDF form do not encode scanned page images (a notable exception is JSTOR, discussed in more detail below). Documents in 'native PDF' rather than scanned image form are one tenth the size and print correspondingly faster.

To improve performance would need faster servers, more (and faster) printers and possibly a faster network (moving from 10 to 100 Mbps ethernet). Increasing the number of printers on the local cluster (to minimize queue delays) is one possible option at this stage. At the same time, the general trend of the technology towards faster processing and transmission means that all components will, in time, improve and replacement is inevitable. Given that most of the cost of such systems is incurred in management and maintenance, replacing a single printer with a faster one may be more cost effective than increasing the numbers.

The printer used in these tests (HP LaserJet 4M plus) represents a useful baseline for evaluating current technology. Now obsolete, with an original capital cost of around £1000, it represents the low end of the installed network printer market. Nevertheless, it should still be able to cope with reasonable levels of demand. Machines of lower specification, however, would almost certainly be inadequate for shared use in this type of application (as would non-laser printers such as dot matrix or inkjets). Assuming a replacement cycle of 4/5 years, printers such as the HP-Laserjet4M are now ready to be upgraded to more recent models, which should result in significantly better throughput.

Section 3.2.2 gives details of the tests from the access point of view

4.5 Types of Network Printer

Printers generally fall into three classes: desktop machines - usually inkjet; mid-range laser printers for work group or shared cluster access; and high end systems for specialist applications.

Our tests have shown that the current range of personal ink-jet printers are quite effective for printing e-journal and similar material, but only in a single user context. With a price range of £100 - £200 they represent relatively small capital investment, even for an individual. The running cost, however, can be high.

Work group printers in the price range of £1000 - £2000 form a distinct class targeted at the shared network market. Most will have direct network access (as well as standard parallel interfaces). Increasingly they will provide full duplex connections to improve monitoring and control. The performance range can be considerable and is rapidly increasing. Middle-rated machines, such as the Lexmark Optra S series or the HP Laser-Jet 4000N, claim a real throughput of between 6 and 18 pages per minute from a loaded queue mixing PS and PCL jobs [26] (the nominal rating of both the LJ4000 and Optra S1855 is 16ppm, and each carries a list price of around £1000). At the other end of this middle range is the Xerox

DocuPrint N32 with a nominal rating of 32 ppm and a price of just under £2000.

Nominal print rates and actual print rates can be very different. A 10-page document composed of scanned images will take very much longer to print than a 10-page MSWord document. As a relative measure of performance, however, the nominal "ppm" print rate is very useful.

High end printing systems are generally targeted at the corporate publishing market or at special applications such as high-quality colour printing. Systems such as Xerox DocuTech are primarily designed for high output of multiple print copies. For this reason they will usually include features such as collating and binding. While such applications may well have an important place within the HEI sector, it is a very different kind of operation from that of user-driven network printing of personal material or online sources (although clearly there will be a need for combining both types of service in some applications). The cost of high-end systems varies considerably, from £10K to £100K and beyond. These prices partly reflect the smaller volumes sold and partly it is their additional features. The latest model, DocuTech 6100, has a nominal print rate of 96 ppm and a US price of \$192,000.

Colour is increasingly found as a component of e-journals published in HTML form. On the other hand, few, if any, PDF versions of current print journal publications use colour in any significant way (a cover page in colour does not really count). This may change as colour printing becomes cheaper (both for short-run hard copy print journals and for network laser quality printing). Personal inkjet printers almost invariably now provide colour and may help drive the market in this direction. As would be expected, colour laser printing is very much slower than monochrome and could create serious bottlenecks on network printing services.

4.6 *Estimates of Scale*

The impact of e-journals on network printing facilities could be considerable. Our investigations indicate that while present systems might be able to cope with current low levels of demand, they could prove frustrating as the use of e-journals (and other electronic source material) picks up. The general problem is how to estimate the likely growth, and to match this to the expected parallel growth in the technological capability of services. Can the technology keep up with the demand?

Present network print services were largely designed to handle the printing requirements of user-generated material - word- processed documents, spreadsheets, etc. They were not specifically designed to handle high volumes of non-user generated source material, such as e-journals, Web-based information or other electronic publications. As the balance switches away from user-generated to network-located sources, there is likely to be an impact on more traditional services. At the very least, an increase in the overall demand for printed output will create delays; individuals trying to finish and print their latest research paper will be frustrated by the demands of those trying to print out material retrieved from their network searches.

One rather simplistic way of getting a rough estimate of how great this demand might be is to ask the question "How many papers are read for each paper generated?". At present, in

writing a research paper (or even an undergraduate term paper) each author is likely to photocopy and read, say, 10 journal articles. If all 10 articles are extracted from e-journals (rather than from print and photocopy) then a ten-fold growth in printing requirements could be expected. This glosses over the question of how many times an author might print out during the creation of a paper, as well as the amount of source material that might be retrieved and later discarded (it may be reasonable to assume that these could balance each other out). It also neglects the fact that a paper retrieved from the network is likely to form a much more bulky file than a locally generated one - either by containing a lot of image material or by being generated from the scanned pages of original paper documents.

This rough estimate of a possible ten-fold growth in print demand as sources move from paper to electronics is to some extent supported by available usage statistics for photocopiers. Figures from a major university research library show that nearly 4 million photocopy pages were made each year. No figures exist for the proportion of this total that are copied from journals, but it is likely to be greater than 50% (and perhaps substantially greater). If all this material were to be made available electronically then the demand for prints from e-journals could be in excess of 2 million pages. Of course, many of these articles will be photocopied or printed from the online source more than once, so this figure might double. If only 25% of this activity were transferred to electronic form the impact could be considerable. In addition, the ease of printing from the electronic form - compared with photocopying from a printed journal - makes the likely demand even greater.

Four million printed pages at a total cost of, say, 1.5p per page (to cover consumables - but not the capital cost of equipment or service maintenance) is £60,000. Much of this expense would, of course, be carried by individual departments rather than central services. The figure for central network printing at the same institution is also around £60,000, suggesting an annual print rate (from central services) of some 4 million pages. The demand for print could therefore double if e-journals became the standard publication format. Fortunately, perhaps, there is little sign yet of any wholesale conversion, though even small steps in this direction are likely to have some impact.

Anecdotal evidence from users of central workstation clusters indicates that printing facilities are already inadequate. Past years exam papers and on-line lecture notes, where scanning and conversion to PDF has been necessary, have been found to cause difficulty on congested systems. Queuing priorities push large files to the back, and can result in a 2-hour wait for output, according to some reports.

4.7 The Need for Page Charging

Again there is anecdotal evidence to suggest that the absence of page charges encourages computer printout rather than photocopy - the example of past exam papers was cited. This has nothing to do with e-journals in particular, but with print management strategies in general.

It is unlikely that the expected growth in demand that will follow from the switch to online rather than print information sources can be managed without resort to some form of page charging. Unlike user-generated material, without rationing there is no natural constraint on how much network-sourced material might be printed out (except the discouragement of

queue delays). This might not matter if it did not risk the impairment of equitable access for other applications. Details of some of the charging systems in use are given in Section 3.3.

Page accounting (and possibly charging) may well become necessary also as a component of copyright management. Almost all e-journals at present are supplied on subscription (if charged for at all). There is interest, however, in finding ways of providing access to non-subscribed material on a pay-per-view basis. This is a complex issue, both at a technical and policy level, and cannot be discussed in any detail here.

There are also print controls within the PDF structure itself. Basically these allow the document supplier to lock the file and prevent it from printing. This is possible because printing is via the Acrobat Reader software, which provides the overall control mechanism on access. It can also prohibit alteration of the document (including extraction of parts of an article, for example). Most e-journals in PDF permit unlimited printing but prohibit alteration (HTML, on the other hand, has no effective means of restricting access). More complex forms of control – perhaps including payment methods – may well be developed in collaboration between publishers and suppliers of document reading software.

4.8 Initial Results from JSTOR

As mentioned in the earlier part of this report, JSTOR is a project to store in digitized form the complete series of around 100 journals (but not including current issues). In March 1998 a mirror site was established in the UK as part of MIDAS, and it is now generally available on subscription to all UK HEIs. From the point of view of this study, there are two aspects of interest: how heavily is it used (and by whom) and what are the technical problems of printing from the archive.

JSTOR scans the original print and stores it as a 600 dpi digitised image (at 600 dpi the file size will be roughly four times that of an image stored at 300 dpi). It also employs so-called 'rough' OCR to provide a machine searchable file. Articles are available in a number of formats, including PDF and PostScript.

The format originally used for the page images was compressed TIFF (using the Gp4 fax compression algorithms). More recently an agreement has been reached with Cartesian Products Inc. to use their proprietary 'Cartesian perceptual compression' (CPC) techniques to encode the primary data. This gives an average reduction of 5 to 1 over the compressed TIFF format. The saving in storage is to some extent offset by the need for proprietary decoder or 'helper' software to run on the user's workstation. This software can be downloaded from the JSTOR server and installed on the workstation. It produces a PS2 output file that can be sent directly to the printer (assuming it is capable of handling level 2 PostScript). Because the files are smaller (by as much as 80%) delivery time from server to workstation is reduced, while the decoding overhead should not add significantly to the processing time. The only real problem is the need for yet another piece of special (and proprietary) software in order to get access to the data. The risk is that every e-journal supplier (JSTOR, CatchWord, UMI, et al) will develop its own proprietary method for accessing its journals, creating chaos for those responsible for managing and supporting services.

It is not necessary to use the 'helper' program to access JSTOR, however, since both PDF and PostScript options are provided. It is understood that the corresponding files are created 'on the fly' from the primary CPC (or compressed TIFF) files – essentially by running the 'helper' software on the JSTOR server. The increase in transmission time, however, could be dramatic. Documents may be requested either in 'fast print' or 'high resolution' mode, and also in PS1 or PS2 encoding (as well as PDF). The difference in file size (and hence transmission time) is remarkable: a 15 page article from an 1890s issue of the Annals of Mathematics was delivered as a 172Kbyte file in 'fast print' PDF or as a 20Mbyte file in PS1. More interestingly, the difference in quality of print between the two, though detectable, was for most purposes insignificant. This is fortunate as low resolution PDF was also the only format that could in practice be used from the shared workstation cluster tried in our tests. Restrictions on user space or security constraints meant that the 860K install file for the Jprint helper program could not run. Allowing users to install unknown executable code or "plug-ins" from network sources is clearly risky to the integrity of shared systems and may be prohibited. In the case of JSTOR, access to a PDF version avoids this problem.

Usage levels for JSTOR are still quite low. The total figures for June 1998 were recorded as 47 (for UCL) and 486 (for all 20 registered UK sites). Around half of these accesses used the Jprint helper program, while the other half used PDF formats and a tiny group used PostScript. It is assumed that most of this usage was from academic or research staff.

JSTOR access using the low resolution PDF format is quite effective from a shared workstation and network printer. The 15 page test document (172Kbytes) took only 20 seconds to retrieve from the archive and 3 minutes to print. A high resolution PDF version of the same document ran to 614Kbytes and took nearly 8 minutes to print (on a very lightly loaded network). Neither of the PostScript versions could be retrieved successfully, presumably due to limitations on allocated user space (the files were 2.6Mb and 17Mb for the PS2 and PS1 versions of the test document, respectively). Nor could the helper file be installed successfully.

Our conclusion from this series of simple tests is that JSTOR access from shared central services is possible but severely restricted in terms of the methods and formats that can be used. With the normal network/printer loading of term-time use, even a 172Kbyte PDF file may cause unacceptable delays, while high resolution PDF would be impractical in most cases.

JSTOR is an important project in that it is attempting to apply cutting edge technology to the problem of creating digitised archives of print journals. While it is unlikely to be a major resource for undergraduate study it serves as a good test case of what is possible on shared network facilities. Developments in this technology will inevitably have applications to other areas of document storage – including the exam papers; course notes and reserve lists that are more directly relevant to student needs. What our study has indicated is that present network/printer installations may be able to cope, but only just. If demand for this kind of resource were to expand significantly then there would be little chance of satisfying it with the systems at present installed.

The JSTOR service is important in that it pushes us to the edge of current technology. While

utilisation levels are at present quite low, it is a useful indicator of current requirements and capabilities. It is anticipated that the archive will grow from 50 to 100+ titles over the next 4 years. Even so, this represents only a tiny fraction of all published journals. As it does not include current issues, demand is likely to remain limited.

4.9 Overview of the Impact of E-Journals on Networks and Network Printing

The investigation of shared systems and network printing tried to answer the following fundamental questions :

- What is the current level of demand for e-journals?
- Can present services cope with this demand (including volume and technical requirements)?
- Is it likely that demand will grow significantly and over what time scale?
- Can technology and resourcing keep pace with this growth?

Current demand levels for e-journal material from shared central services appear to be quite low. To some extent this may reflect the relative difficulty of accessing e-journal material, or ignorance of what is available. The shared central systems used by students have a very different utilisation pattern from the 'private' systems of research staff and other faculty. During term time they are heavily loaded and the demand for print is high. Queue priorities and shortage of printer capacity will frustrate access to large e-journal files. While tests show that there is no technical difficulty in printing large PDFs, in practice it is likely that many shared systems would be unable to provide a reasonable level of service, particularly from sources of page-image material such as JSTOR. Tests showed only minor delays from non image encoded documents, and no significant problems with current hardware (including both Macs and PCs).

Among students, a more significant source of demand for documents in page image format are the archives of past exam papers that many HEIs are creating. In many cases these have to be scanned in and encoded as PDF image files. At from 6 to 10 pages, they compare with the 'worst case' example referred to above. As heaviest use is from undergraduates, who are also the main users of shared central services networks, the impact on printing services could be high.

Network printers, as opposed to direct PC-linked printers, will normally carry a reasonably high performance specification. The HP Laserjet 4M was one of the first to provide a direct ethernet connection as a standard. In tests it proved capable of dealing with most PDF-encoded e-journal material. Only in the case of scanned images was its performance poor. Its current equivalent (Opra S1855 or HP-LJ4000) should be substantially better.

The need to move to a higher price bracket (from the £1K to £2K of the 'mass market' network printer) for better performance is unproven. High performance printing/copying systems such as Docutech would be quite inappropriate to this kind of application. There might be some benefit in moving to the 40ppm range of machines such as the Xerox

DocuPrint N40 (listed at US \$3500), but this would need further evaluation. A better strategy would be simply to increase the number of printers available or reduce the number of workstations sharing each printer (a ratio of 1 to 10 seems about right and has been adopted by many institutions).

Other sources of electronic material (exam papers, encyclopaedias, reserve lists, etc.) are likely to have a greater impact on central services. And while currently available technology can certainly cope, it may entail substantial investment in updating printing systems to ensure a reasonable response rate. That, in turn, may mean that some form of page charging will be mandatory. No clear best method has yet emerged to provide such a facility, however, and the real technical difficulty of implementing page charging has inhibited its development.

5. General and Economic Issues Affecting the Use of Electronic Journals

The main economic problem for e-journals at present is that their utilisation rate appears to be very low. Reliable figures are difficult to come by, but anecdotal and other evidence points in this direction. At the same time, it is clear that some online articles in certain research areas have enjoyed an enormous readership (the Southampton survey of STM e-journals currently claims over 18 thousand hits!). Extending this level of use to e-journals in general is something of a challenge to publishers. There is now no technical difficulty in collecting information on the number of hits on any online article, but the publishers hold this information and they are, perhaps understandably, reluctant to reveal the figures. Although printing and network problems are incidental to this more basic issue, they are not entirely unconnected.

The only way to increase the utilisation rate is to increase usage among the current readership (that is, those with existing subscriptions, to either print or electronic versions) or to expand the market. Creating added value for e-journals by exploiting the inherent capabilities of the technology is central to achieving economic viability [21].

Increasing usage within the present market means building a more efficient, more responsive service. Apart from the need to create greater awareness within the marketplace, this is largely a technical issue - faster networks, easier access, faster printing. Increasing usage will follow from increasing 'utility'. With 'perfect' access - fast identification and location of source material, easy registration, immediate access, and super-fast interaction with onscreen display - printing should be increasingly unnecessary. In this sense, printing is a response to inherent system inadequacies. Unfortunately, few sources are of this kind -- personal CD-ROM encyclopaedias being one possible exception.

Expanding the market may mean going outside the university sector. As the commercial and public sector becomes more increasingly research-based the demand for access to scholarly information must grow. This applies not only to STM literature, for which there has traditionally been a very clear need and demand from outside the HEI sector, but even arts and humanities publications are likely to see a strong and increasing demand from the media, culture and heritage sectors of the wider non-academic world. Strictly speaking, such considerations are beyond the brief of the present study, which is really only concerned with the implications for HEIs. However, it could be argued that only by expanding the market in

this way will it be possible to create an economically viable e-journal environment for scholarly publishing in general. After all, this is no different from what has been experienced in the past - computers are cheap because of the commercial market, not because of demand from the research departments in which they originated, and the same is true of the Internet itself.

Our study has shown that there are no real technical barriers to accessing e-journals from outside the academic environment - whether via PC and modem, or corporate link to the Internet. Adequate equipment is now surprisingly cheap. The real barrier is administrative - the management of subscriptions and access rights is the really big problem. This is a difficult question that will need to be addressed by both the library sector (HEI and Public) and the publishers. As things stand at present, students can be excluded from e-journals by the limitations of the technical resources available, while the wider public is excluded by the narrow licensing arrangements.

Online browsing is not something that scientists like to do during quality research time. Just as much reading is left for otherwise unproductive periods such as train journeys and bath time. Late night Web surfing falls into this category; it is therefore essential to have access from home. Publishers are increasingly recognising this and are prepared to include home access within their licensing agreements. Our experiments have shown no technical difficulty in home access and many advantages, such as being able to avoid periods of network congestion. Printing in this context is also viable (even for quite large documents in scanned image form) and avoids the problem of local network and printer congestion in the workplace. The personal cost incurred may be some disincentive, but is trivial in comparison to the cost of commercial document supply services, for example.

Printing costs can be high. For a small laser printer with replaceable cartridge the cost per page is around 4p (based on cost of cartridge and of paper, but excluding the capital cost of equipment or the maintenance charge). With capital and labour included this might well approach the 10p per page charged by some libraries (for others, 5p appears to be the acceptable minimum). Inkjet printers are likely to bear similar costs, even though the capital cost may be less; maintenance and consumables charges could be higher. Personal printing of this kind generally carries the highest cost. Even so, the cost, at £1.00 for a copy of a 10-page article, is smaller by an order of magnitude than for document delivery, and roughly comparable with do-it-yourself photocopying in the library (while being infinitely more convenient).

Workgroup printing can benefit from some economies of scale, by using faster machines with a higher utilisation rate. The technologies used (inkjet and replaceable laser cartridge) are still inherently expensive. High volume laser printing carries a much lower per page cost, being based on different technologies requiring more expensive but longer lasting components, but the capital and administrative costs per machine are likely to be much higher. The usual estimate is 1.5p per page for consumables (paper and toner), but excluding capital cost and maintenance.

Network congestion and restrictions on access will increase the demand for print. If access were available 'on-demand', with 'perfect' response, then users might feel less need to print out material as security against failure of access. At the same time, if more terminals were

introduced and better network access provided, then the higher volume of use would be likely to generate even more print - until some kind of saturation point is reached. This is the basic conundrum of trying to engineer a smooth transition from a print-based information economy to an electronic one. While commercial organisations with their centralised planned economies may impose electronic 'work-flow' upon their workers, the free and 'democratic' economy of an academic institution is less easy to manipulate.

6. Conclusions and Recommendations

6.1 Conclusions

E-journals represent just one aspect of a growing move towards online content, and away from the earlier focus on user-generated material. This is at present frustrated in many sectors by lack of resources or by unreasonable restrictions on access. Within HEIs, a shortage of print capacity on shared public services is a major block on access to online information.

There is no fundamental technical problem, nor any quick technical fix. Current printer technology is well able to cope with the requirements of large PDF files, or their PostScript derivatives. Whether networked laser printers or personal inkjets, the current generation of printer is capable of handling the print files of current e-journals. Obsolete or outdated equipment, however, is likely to fail in this task.

The short survey undertaken as part of this study indicates that printing is not generally perceived as a major problem in relation to e-journals. While it is true that reports of slow or failed printing have occasionally emerged, these usually have an obvious or simple cause. Any printer less than 2 or 3 years old, coupled with Acrobat 3.0 should have little difficulty in coping with any of the current e-journal offerings encoded in PDF. Problems with HTML are largely due to its hypertext-screen-oriented structure. Nevertheless, scanned page images or textured backgrounds will slow things down considerably. Publishers may need to take this more into account in the design of e-journals – textured backgrounds, in particular, have little justification.

Where printing has a problem it is with shared central services. Here the move from user-generated material to online sources is already creating considerable difficulties for shared access printing systems. The higher level of demand for print will mean that increased capacity must be installed.

The basic issue is one of resources and rationing. A policy of free printing from shared services is unlikely to be sustainable in the face of growing demand. Without rationing, the quality of service will decline to the point where many users, including students, will opt for private facilities. Putting more and faster printers on the network can only be justified, however, if it is accompanied by adequate systems of accounting and billing. This essentially means page charging.

Costs can be high, roughly comparable with do-it-yourself photocopying in the library; workgroup printing can benefit from some economies of scale, by using faster machines with a higher utilisation rate. The technologies used are still inherently expensive. High

volume laser printing carries a much lower per page cost, being based on different technologies requiring more expensive but longer lasting components, but the capital and administrative costs per machine are likely to be much higher. There is little justification for continuing to subsidise printing, as happens at present - except for the high cost and difficulty of implementing an automatic charging mechanism. The development of accounting and charging systems appears at present to be proceeding on an individual site basis and some co-ordination across the whole HEI sector would seem to be desirable. Page-charging is technically quite difficult to implement. No standard solution has yet emerged, although a number of products and systems have appeared in the market or have been developed by individual institutions. Detailed discussion of this issue, however, is beyond the bounds of this study.

With private facilities the issues are clearer. Personal workstations are now the norm for research and teaching staff at most institutions. Printers may be shared, but this is usually within a very small workgroup where it is easier to allocate resources to maintain or upgrade them. The problem here is more one of locating online resources and negotiating access.

Negotiating access in the truly private domain is even more difficult. Most licensing is still on the basis of IP address, which restricts home access. No solution to the problem of public open access to e-journals – in the manner of public access to libraries – has yet been proposed.

A real concern for libraries, is managing e-journal subscriptions and access rights on behalf of their users. This is a fundamental issue that will only be solved by continued close co-operation between suppliers and libraries. We cannot yet be certain of the optimal licensing model, partly because there are conflicting demands from users, on the one hand, and publishers, on the other.

Users, above all, want fast response and easy access from any location – library, office or home. Better networks, more mirror sites, and cleverer caching strategies will all help to improve response rates. But the cost is not insignificant and must largely be born by the publishers or universities themselves. In the meantime users will chose to print - or create their own cache of electronic material. The risk for suppliers is that if access is frustrated then users may not buy into the online model at all, or choose to go elsewhere for network information – electronic editions could wither under competition from more accessible electronic-only sources, or private caching from pre-print archives. [22]

6.2 *Recommendations*

A number of these recommendations are for further study. Some are more specifically targeted at providing solutions to immediate problems.

We need better ways of monitoring the print stream to see exactly what kind of impact online sources are having. It is very difficult to disentangle e-journal printing problems from the problems of network printing in general. How serious is the problem of scale? Information Services management do not appear to know, and find it difficult even to guess at the composition of the print stream. User surveys at the print station might help.

Network printers must be upgraded. Guidelines are therefore needed on the cost/performance characteristics of currently available network printers and their suitability for the typical print stream of a university shared resource service. There appear to be substantial benefits in providing a PostScript-only service, but this should be confirmed by a more detailed cost-benefit analysis of PS vs PCL printers. There is also scope for involving CHEST in this area.

A closer look at page charging is required. At what point of demand does page charging become imperative? A comparative study of existing implementations to assess the impact of charging, to provide a basis for the evaluation of commercially available systems and to investigate some co-ordination across the whole HEI sector is desirable.

More complex forms of control are required– perhaps including payment methods – possibly developed in collaboration with publishers and suppliers of document reading software. We cannot yet be certain of the optimal licensing model; page accounting (and possibly charging) may become necessary as a component of copyright management. There is also interest in finding ways of providing access to non-subscribed material on a pay-per-view basis.

Queuing strategies and resource allocation greatly affect the response time of shared print services. File size can be measured, but this does not correlate well with the number of pages in the corresponding document. Large files (rather than number of pages) could be penalised by being given a low priority in the print queue. The technical problem is that it is difficult to determine the page count from the PostScript file alone. Nor can a printer response be used since the count is needed before the job goes to the printer. PDF files do include a page count, however, so there may be a way of giving these priority. A review of existing systems and their effectiveness could be helpful.

Finally, more research is needed into more radical solutions. What would be the consequences if institutions abandoned the provision of shared PCs and printers? Would it be more cost-effective to subsidise student purchase and spend money on network support and associated services instead? What proportion of study bedrooms in university accommodation are already wired up to campus networks? What proportion of students already have access to their own PCs and to the Internet? These issues are critical for future planning and carry very significant budgetary implications.

7. References

- [1] HEFCE. Report on Phase I of the Evaluation of the UK Pilot Site Licence Initiative. HEFCE, April 1997
- [2] Steve Hitchcock, Leslie Carr, & Wendy Hall. A survey of STM online journals 1990-95: the calm before the storm.
<URL: <http://journals.ecs.soton.ac.uk/survey.html>>
- [3] Steve Hitchcock, Leslie Carr, & Wendy Hall. Web journals publishing: a UK perspective. *Serials*, 10 (3), November 1997.
<URL:<http://journals.ecs.soton.ac.uk/uksg.html>>
- [4] Hyperjournal, Directories of Electronic Journals.
<URL: <http://www.ukoln.ac.uk/isg/hyperjournal/director.htm>>
- [5] E.doc, e.journal pages maintained by WILMA.
<URL: <http://www.edoc.com/jrl-bin/wilma/spr>>
- [6] Tony Kidd. Are print journals dinosaurs? *Ariadne*, 12, November 1997.
- [7] John MacColl. The professional magazine and parallel publishing. *D-Lib Magazine*, February 1997.
- [8] Derek Law. Parlour games: the real nature of the Internet. *Serials*, 10 (2), July 1997.
- [9] Judith Wusteman. Formats for the electronic library. *Ariadne*, 8, March 1997.
- [10] Catchword Web pages
<URL: <http://www.catchword.co.uk>>
- [11] OCLC Web pages
<URL: <http://www.OCLC.org>>
- [12] Paula Berinstein. Text and graphics on UMI's ProQuest Direct: The best (yet) of both worlds. *ONLINE*, March 1997.
- [13] JSTOR. Why images?
<URL: <http://www.jstor.org/>>
- [14] Paul Ginsparg. Winners and losers in the global research village. *Conference on Electronic Publishing and Science*, UNESCO, Paris 1996.
- [15] Cliff McKnight. Designing the electronic journal: why bother? *Serials*, 10 (2), July 1997.
- [16] Hugo Brailsford. Parallel publishing for transactions. *Ariadne*, 11, September 1997.

- [17] VirLib Project <URL: <http://www.ua.ac.be/MAN/TO2/root.htm>>
- [18] Fiona Williams. Electronic document delivery: a review and comparison of different services (Part 2). *Ariadne*, 11, September 1997.
- [19] John Kirriemuir. Interview with Clifford Lynch. *Ariadne*, 10, July 1997.
- [20] Thom Lieb. Caution: Speed Zone. *Journal of Electronic Publishing*, 3 (2), December 1997.
- [21] Malcolm Getz. An economic perspective on e-publishing in academia. *Journal of Electronic Publishing*, 3 (1), September 1997.
<URL: <http://www.press.umich.edu/jep/03-01/getz.html>>
- [22] The Economist. Publishing, perishing and peer review. *The Economist*, 24 January 1998.
- [23] Jon William Toigo. Network printing management: holy grail or holy terror? Boucher Communications, Inc., 1996.
- [24] Hewlett Packard. Network printing solutions: the value of a network printing strategy. Hewlett Packard Inc.
- [25] John Howerter. Internetworking and interoperability: network printing is not pretty! *Journal of Electronic Publishing*, February 1996
- [26] NSTL. Lexmark Optra S: Performance evaluation. NSTL Report, Lexmark, April 1998.

8. Appendix A: Glossary of Acronyms

FTP	File Transfer Protocol
GEDI	Group for Electronic Document Interchange
GIF	Graphics Interchange Format
HEI	Higher Education Institution
HTML	Hypertext Markup Language
JPEG	Joint Photographic Experts Group
JSTOR	Journal Storage
MAN	Metropolitan Area Network
OCR	Optical Character Recognition
PDF	Portable Document Format
PNG	Portable Network Graphics
PPT	Parallel Publishing of Transactions
ppm	Pages Per Minute
PS1	PostScript Level 1
PS2	PostScript Level 2
RLG	Research Libraries Group
TIFF	Tagged Image File Format
TIFF/IT	Transport Independent File Format/ Image Technology

9. Appendix B: Technical Notes

9.1 Equipment Configurations Used for Testing

The following note gives some further technical detail on the various systems used to test access to e-journals and their printing requirements.

The experiment began with what was assumed would be the worst-case scenario. This involved using a 486-based PC running Windows 3.1 with a PostScript Level 1 laser printer and a 14K bps modem connection to the Internet. Software consisted of NetScape 2, Acrobat 3 and an assortment of TIFF readers. In the event, even this rather obsolete combination proved to be surprisingly effective for access to most of the e-journal sources available – partly because it was available at all hours, and access times could be chosen when demand was likely to be lowest. The only significant failure was its inability to cope with printing TIFF images, either in the form of PDF documents or as plain files. Colour printing was also absent. On-screen display of images, on the other hand, was usable if rather unresponsive.

The next combination tried was a library-based Windows NT workstation linked to JANET and the London MAN by what appeared to be a relatively high-speed connection. Printing was through a directly attached ink jet printer (shared between two PCs). Colour was not available. Access to e-journal archives was very time-of-day sensitive. North American sources were virtually unusable after mid-morning and even some UK sources were unresponsive during normal working hours. Printing was slow and the quality poor, suggesting that printer maintenance may have been a problem. Access to the PC was on a reservation basis, which made time-shifting difficult. Software available included the Internet Explorer (version 3) browser with Acrobat 3 as a plug-in. Scanned documents from the PPT archive were easily accessible with very good response time (presumably because the server is on the London MAN). The response time to some other JANET-based servers, however, was poor.

The third combination was a Unix workstation running the Mosaic browser and linked via a local area network to the London MAN. This was part of a work group arrangement used by a small research group. Response times were generally good, except to North American sites. The main problem was in configuring the software. Acrobat was not immediately available and PDF documents had to be first downloaded then printed via another PC linked to a fast colour inkjet printer. It could have been made to work very adequately had it been possible to spend more time optimising the software configuration.

The fourth configuration was a Pentium-based PC, running Windows 95 and Internet Explorer 3, with a Lexmark 2050 personal colour inkjet printer directly attached. Internet access was via a 50K bps modem. Acrobat, along with an extensive range of other plug-ins (including a TIFF viewer) was available as part of the standard packaging. Speed of access to some sites was considerably slower than with the network connected configurations, principally those directly on the London MAN. With many others, however, it was no worse, and in any case this could often be compensated for by using it at unusual hours. Printing of scanned images and documents with colour photo inserts was surprisingly effective.

9.2. Test Samples:

The full details of the e-journal documents used in the tests discussed in Section 3.2 are given below.

Example 1. Page image article from the Transactions of the Institution of British Geographers (TIBG) archive - (Appleton, "The communications of Watford Gap"). This consisted of 10 pages, with many graphical images, from a 1960 issue of the TIBG, scanned at 300 dpi and encapsulated as a compressed TIFF image within a PDF envelope. The file size of the document in PDF is 1.3MB. Expanded to PostScript it runs to over 24MB. The location of the server is on JANET at Queen Mary and Westfield College, giving rapid access from other university sites in London.

Example 2. Native' PDF article from the TIBG archive - (Amin & Graham, "The Ordinary City"). An article from TIBG, the same journal as Example 1, but encoded in 'native' PDF, rather than scanned page image format. This was a 19 page article converted from the OCR'd image and stored in PDF. The file size is 185KB.

Example 3. A complex 'multimedia' article - paper from the e-journal *Music and Anthropology* - (Fenlon, "Music, ceremony and self-identity in renaissance Venice"). This is an example of complex HTML encoding, with many coloured pictures and sound clips included within the document. The server is based at the University of Maryland and mirrored at the University of Bologna.

Example 4. A 'scientific' paper from IDEAL. A typical science research paper - (Wong, et al "A Fourier transform study of phase transitions" from the *Journal of Colloid & Interface Science*). This was a 7 page document with tables, graphs and many complex textual components from the Academic Press IDEAL archive. This is encoded in PDF, but with smaller graphical components. The server is based at the University of Bath on JANET (and mirrored in the US).

Example 5. An example of background image overprinted with text. This was the front page of a journal (*Economic Perspectives* - from the US Information Agency's archive) with coloured text headers and a background 'watermark' image, encoded in PDF. The file size for the entire (40 page) issue ran to nearly 0.5MB. The server was based in the USA.

Example 6. A plain text paper from the same journal issue as Example 5.

Example 7. Papers from the humanities collection of Project MUSE. These were mostly text, with occasional GIF inserts, encoded in HTML. The journals are primarily parallel publications, with full-length text articles (each around 20 pages) and only limited graphics. The server is located at Johns Hopkins University in the USA.

9.3 *E-journal Suppliers*

The list below is in no way comprehensive; they were chosen to cover a reasonably wide range of the different kinds of service available and how relatively successful these might be, as discussed in Section 3.4.

For each supplier, the list gives the following details:

1. Name of publisher or service
2. Type of service offered
3. Number of journals and major subject areas
4. Document formats used
5. Location of servers

1 Academic Press - IDEAL

- 2 an electronic archive for all the AP titles
- 3 175 journals (23 titles in physics) all of which are electronic editions of print titles
- 4 Document formats are PDF, with abstracts and TOCs in HTML
- 5 [http:// www.janet.idealibrary.com](http://www.janet.idealibrary.com)

1 Institute of Physics

- 2 An electronic archive of all IoPP publications
- 3 33 journal titles (mostly Physics) plus newsletters, magazines, etc.
- 4 Formats are PDF and compressed PostScript (Gzip). File sizes are around 100K+ (PDF) or 50K+ (compressed PS); many also include links from references to INSPEC abstracts
- 5 [http:// www.iop.org](http://www.iop.org)

1 UMI ProQuest Direct

- 2 a one stop shop involving many publishers
- 3 Several hundred titles, with particular strengths in Computing, Electronics and Telecommunications
- 4 A variety of formats, including HTML and PDF. ProQuest Direct Version 1.6 offers colour and full-text graphics from a select number of journals (4/5 medical journals, including the New England J of Medicine, and Brain)
- 5 [http:// www.umi.com](http://www.umi.com)

1 Johns Hopkins University – Project MUSE

- 2 a University publisher
- 3 around 30 journals (mostly Humanities)
- 4 mostly HTML but some in PDF and even TeX (for some mathematical journals)
- 5 http://muse.jhu.edu/journals/index_text.html <http://muse.jhu.edu/demo/index.html>

1 BIDS - JournalsOnline

- 2 a one-stop shop for e-journals, plus a document delivery service for those not available in this form
- 3 500+ titles in all subject areas
- 4 mostly PDF, with TOCs and bibliographic information in HTML
- 5 <http://www.bids.ac.uk/JournalsOnline>

1 JSTORE

- 2 A consortium of several universities and publishers
- 3 Around 100 journals (complete runs except for current 3/5 years)
- 4 All stored in high res. TIFF plus rough OCR for searching. PDF also available and HTML for bibliographic information and Tables of Content
- 5 <http://www.jstor.org>

1. PTT project

2. Electronic archive of a print journal --
3. Complete run of one title (Transactions of the Institute of British Geographers)
4. Scanned or OCR'd images in PDF
5. <http://ppt.geog.qmw.ac.uk/tibg/>

1 CatchWord

- 2 Large-scale document warehousing using multiple mirrored servers (3 UK sites alone) with many individual publishers
- 3 150+ titles from 15+ publishers
- 4 All based on proprietary RealPage format and associated browser
- 5 <http://www.catchword.co.uk>

1 CERN Pre-prints

- 2 A pre-print archive for papers in physics
- 3 Many thousand articles in Physics from 1994
- 4 Each scanned item available as PDF, TIFF, compressed PostScript and GIF pages – if non-scanned then only PDF and PostScript are supplied
- 5 <http://preprints.cern.ch>