UNIVERSITY OF Southampton

University of Southampton Research Repository
ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

http://eprints.soton.ac.uk

# Prologue

## What time is it Eccles?

*Scene* Two plain clothes detectives surround the Crun omnibus factory, secreted in the ground floor attic of a nearby clock repairers.
*Various timepieces ticking, chiming and cuckooing. A chicken clucking. Finally a hooter.*

*Bluebottle (aka Peter Sellers)*: What time is it Eccles?
*Eccles (aka Spike Milligan)*: Err, just a minute. I've got it written down on a piece of paper. A nice man wrote the time down for me this morning.
*Bluebottle*: Ooooh, then why do you carry it around with you Eccles?
*Eccles*: Well, um, if anybody asks me the time, I can show it to dem.
*Bluebottle*: Wait a minute Eccles, my good man.
*Eccles*: What is it fellow?
*Bluebottle*: It's writted on this bit of paper, what is eight o'clock, is writted.
*Eccles*: I know that my good fellow. That's right, um, when I asked the fella to write it down, it was eight o'clock.
*Bluebottle*: Well then. Supposing when somebody asks you the time, it isn't eight o'clock?
*Eccles*: Well den, I don't show it to 'em.
…
*Bluebottle*: Well how do you know when it's eight o'clock?
*Eccles*: I've got it written down on a piece of paper.
*Bluebottle*: Ohh, I wish I could afford a piece of paper with the time written on.
*Eccles*: Oohhhh.
*Bluebottle*: 'Ere Eccles?
*Eccles*: Yah.
*Bluebottle*: Let me hold that piece of paper to my ear would you? 'Ere. This piece of paper ain't goin'
*Eccles*: What? I've been sold a forgery.
*Bluebottle*: No wonder it stopped at eight o'clock.

Transcript from *The Goons, The Mysterious Punch-Up-The-Conker*, first broadcast 7th February 1957

There is a sense in which scholarly researchers, librarians and archivists are effectively emulating Eccles by seeking to migrate and maintain the record of research on the Web while, for a variety of reasons, clinging to its print characteristics.

Take, for example, reactions to JSTOR, the service that digitises copies of old print journals: “It seems there's not much bad you can say about an organization whose mission is to "preserve and maintain journal literature, and to make that material more accessible.” (Phillips 2001)

Yet when we try to add dynamic, time-varying features to the scholarly literature, as the electronic information environment compels us to do, reactions become more polarised, as this thesis and the model of the title that it uses to explore these issues, shows.

Academic research preserves the results of its endeavours in the form of an authorised record, historically in printed journals and other publications. In this system information is fixed in time. Thus readers can, with a greater or lesser ease determined by cost or location, access the individual published records and view them as a true record of the author’s intent.

The benefit of preservation is that a reader today looking at a printed issue of the *Proceedings of the Royal Society of London*, Series B, from 1934, say (the earliest year currently available in the JSTOR archive), will see exactly the same articles as his or her predecessors. But the reader’s context, his or her view of the world or line of enquiry, will be entirely different. It is possible to enhance the purpose of preservation in a digital environment by enabling the user to select time-varying services that can be superimposed without detracting from the original document.

Take a familiar example – in scholarly papers connections with earlier works are established by authored references. Garfield (1955) showed that analysis of these references across a broad section of the literature can reveal influences on later works, e.g. what cites this paper? While references are a standard feature of scientific papers, citation analysis requires additional services. ISI, the information company set up to exploit Garfield’s method, and other recent projects show that electronically it is possible to make citation data accessible either within, or closely linked with, an original paper, although it is clearly not an authored part of the paper.

Yet in the scholarly information chain we remain constrained by the continuing emphasis on writing, content and preservation, to the exclusion of other data services. The problem with digital preservation is not the technology but deciding exactly what to preserve and in what form to preserve it. There are no cost-free options. At one level all that needs to be preserved are an author's original

words and illustrations, simply the digital bits (Negroponte 1995). For the archival journal literature, digital preservation concentrates on saving page images, naturally, since that is the only available version of the data. But then the emphasis switches subtly to page fidelity. For Web-originated materials, preservation may be concerned with saving the presentation, that is, reproducing the user's experience. Add the attention given to technology migration and emulation as a means to preserve both digitised and digital experiences, and barriers begin to be erected – cost, technological and cultural barriers – to other uses of these digital resources.

Alternatively, content can be separated from presentation, in practice as well as principle, recognising that beyond the original authored content it is not necessary to preserve everything from a given presentation in time. Otherwise it will never be acceptable to add new data to presentations.

Clearly, it is facile to write down the time. Listeners may laugh at Eccles and Bluebottle who are revealed, by nuances of language and expression, to be more dim-witted than they. Stripped of its audible expression, the dialogue may be more likely to provoke incredulity. The simplicity of the example may be disarming, but both reactions mask a sense of unease, that we have a very simplistic notion of time and its effects. New media, especially networked, computer-based information services, are beginning to expose that unease.

# 1 Introduction

## 1.1 Motivations

Scientists and researchers, Nobel Laureates among them, have produced the clearest declaration of their requirement for access to published research papers – a comprehensive collection that can be efficiently indexed, searched, and linked: 'a dynamic digital archive' (Roberts *et al*. 2001). This thesis is not about the form of the archive. It is about the 'dynamic', the innovations and services that might begin to emerge if the content of the archive, or collective archives, becomes freely, or openly, accessible. “Unimpeded access to these archives and open distribution of their contents”, Roberts *et al*. continue, “will enable researchers to take on the challenge of integrating and interconnecting the fantastically rich, but extremely fragmented and chaotic, scientific literature.”

That declaration encapsulates what this thesis, and the model it promotes, contend are the vital features of any system of electronic scholarly communication, which must be:

- Integrated and interconnected, making something accessible from something else
- Dynamic, computationally adding time-critical data services

Allied to the idea that electronic access to scholarly papers might be open to *all* regardless of location, background or privilege, these characteristics point to the main inspirations for the work described in this thesis. In an integrated information environment we can expect to reproduce Bush’s (1945) ‘intricate web of trails’ in a way that more closely matches the speed and spontaneity of enquiry, especially in scholarly research. We may at last be able to approach this vision with today’s powerful computer networks.

It is not simply the computing power of a network that matters but how it connects people. Publishing, like a network, connects people. Nelson saw interconnected information, or hypertext as he called it, as a means to create a world publishing system, but his approach was superseded by the leaner, simpler, less imposing model of Berners-Lee’s World Wide Web, which attracted the interest of publishers of smaller, more traditional units of publication. On the Web the core skill of non-fiction publishing, creating 'information interfaces', can be applied more powerfully: “In the old model, the information product is a container. In the new model, it is a core. One bounds a body of content, the other centers it.” (O'Reilly 1996)

Despite calling for a dramatically different information environment, the scientists' statement implicitly supports the continuing role of journals as validators and filters of the scholarly literature that they expect to form the core content of the archive. Fulfilling this expectation will profoundly change journals, for which the scientists are the principal contributors and users. In particular, two issues endemic to the established scholarly journal publishing system will have to be tackled:

- Decoupling content from publishing process and function
- Reversing the fragmentation of the scientific literature

This thesis presents a new model electronic 'journal' that, conceived and developed before this debate, can claim to contribute to the scientists' objective of a dynamic, integrated and interconnected literature. *Perspectives in Electronic Publishing* (PeP) indexes and links selected works, combining the functions of a review journal with original materials and access to full-text papers on a focused topic, in this case on electronic publishing.

A scholarly journal can take many forms, but characteristically they present original, peer reviewed papers submitted by their authors. This may not be a strict definition, but for many contributors these are the defining features. In addition journals seek to address a specified community, to which the selected materials are actively disseminated. Then there is the issue of presentation, most obviously in the editing, layout and appearance of the works, which ideally are consistent for the whole package, not just individual works.

In the PeP model **papers are not bound to the journal physically, by location or by transfer of ownership, but by a collection of links stored in a link database.** PeP targets an audience, and selects scholarly papers – supported by review, comment and feedback – from the corpus of the freely accessible online literature in the chosen field. Thus it fulfils certain elements of the established journal model (selective dissemination), tries to reinvent others (review), and omits services that might be considered inessential or which are performed elsewhere. Whether this model is accepted as a scholarly 'journal' is a point of contention that will be investigated.

PeP is a primary example of decoupling journal processing tasks. The link data is not bound to the textual data, so this service is *additional* to the underlying content. As currently structured, PeP can include papers with lower overhead than conventional journals, because it doesn't perform all the functions – editing and layout, etc. – of those journals. PeP has limited scope but unlimited space, which it can fill faster than any conventional journal. Limited scope suggests a fragmented approach,

but by linking content from over 100 named sources, as well as many personal sites, PeP is less fragmented than conventional journals.

This approach is consistent with the variously argued view that on the Internet it will be intellectual services that have value, not intellectual assets, in other words content (Barlow 1994, Bauwens 1996, Dyson 1994). According to Dyson, “while content won't be entirely free, the economic dynamics will tend to operate as if it were.” Aggregated content, in contrast will remain of high value. Some argue that in the age of information glut the new currency will be 'attention'. This is likely to be especially true of scholarly works (Mermin 1992), but others put the case more generally (Goldhaber 1997).

Using PeP, the thesis examines whether the journal ‘container’ can be expanded without compromising its defining features, and if so whether it might be feasible to approach Bush’s vision, for journal users, of instant navigation and unanticipated discovery in a virtually – in principle – unlimited information universe.

The established structure of scholarly journal publishing, encompassing primary journals and secondary services such as abstracting, indexing and alerting, is now largely reproduced digitally. If this structure is not yet fully integrated and interconnected, significant progress has been made, but it remains too dependent on the restrictions on access . The challenge for Web publishing over this digitised structure is to give *all* users the freedom to explore *all* works in context, to discover new relationships, perspectives and ideas through immediate and direct interaction with all relevant texts.

## 1.2 Publishing background

On electronic publishing two views appear to prevail today: one is that electronic publishing is publishing as we know it with the aid of new technology (Graham 2001); the other is to treat the media integration capability of the new technology, or ‘multimedia’, merely as mixed media. Both views misunderstand the impact of new technology on the publishing product, that is, the product experienced by end-users rather than the publishers’ end-product, profit. The reality is that from the user’s perspective, electronic publishing, principally in the form of networked, online, Internet-, World Wide Web-based publishing, will be publishing reinvented, and that the online medium is a new medium in its own right. These factors alone are sufficient for a major transformation, generating new opportunities and new products and changing publishing practices.

If publishing is changing, what about the products? Multimedia products, typically on CD-ROM and latterly DVD, became a mass market commodity in the 1990s, and the most imaginative examples

demonstrate the distinctiveness of multimedia compared with other media forms. But as one mature technology reaches the mass market, another appears with yet more far-reaching implications. Publishing via computer networks offers a new dimension: interactive, in some cases approaching real-time, communication that transforms product generation and dissemination. Online publishing removes the constraints of physically packaged products such as books, journals and CD-ROMs.

It is in this context that this thesis focuses on academic publishing, mainly journals publishing. One of the less glamorous publishing sectors, academic publishing has nevertheless become, when viewed in the totality of 'professional' publishing, one of the most commercially successful publishing businesses. In the summer of 1995 the largest science, technical and medical (STM) publisher, Reed-Elsevier, announced the planned divestment of many of its constituent publishing companies, including mass-market book publishers and newspaper groups, to reinvest in 'more profitable' professional publishing, probably in electronic publishing, it was reported (*The Times*, 19 July 1995). The results of that investment have included the purchase of a major US rival publisher and the development of online businesses aimed at 'professionals who cannot do without the information services the titles provide' (*The Times*, 10 August 2001).

Academic journal publishing is peculiar in other respects. It is perhaps the only commercial publishing sector that does not routinely pay its principal contributors, academic authors and editors. It has, at its worst, generated large profits from specialised journals with tiny circulations and therefore large overheads by ruthlessly price-gauging already expensive products. At the other extreme, academic publishers have nurtured discipline-wide, sometimes centuries-old journals that confer an unmatched authority on their contributors (Schauder 1994). These contributors are among the first network-literate users, yet they remain curiously conservative when asked to break long-established publishing rituals that are embedded in the academic reward system.

Wily publishers have exploited that conservatism. Recognising the growing popularity of the Web, in 1995 journal publishers began producing digital versions of established journals, inevitably the vast majority copied directly from print originals using portable document formats such as Adobe Acrobat. This may appear, given the support of all major publishers, to be the definitive short-to-medium term solution. Publishers such as Reed Elsevier have even been cited as examples of the first online business models to become profitable (*The Times*, 10 August 2001), something they will be keen to maintain.

The switch to digital journal publishing itself motivates more dramatic change, however. Whether viewed from Negroponte's (1995) perspective on the impact of the transition from a physical form to

a string of digital 'bits', or the McLuhan (1964) view that “we shape our tools and afterwards our tools shape us”, even this partial shift to digital journals represents a major change in the nature of information, its communication and value in society.

Can academic journals survive in their present form? On the basis of a broader analysis of magazine publishing, Randle (2001) is cautionary: “even with all the excitement generated by the explosive growth of the Internet and its exciting capabilities, history shows us that old media continue to survive and prosper – somehow.” Academic journals will challenge the rule expressed by the author and media commentator Umberto Eco, that “in the history of culture it has never happened that something has simply killed something else. Something has only profoundly changed something else.” (*The Observer* 18 June 1995)

## 1.3 Plan of the thesis

The origin of this thesis was a detailed survey of the development from 1990-95 of the first Internet-based electronic scholarly journals in the areas of science, technology and medicine (Hitchcock *et al*. 1996). The survey analysed the features, formats, economics and the prospects for e-journals. The survey noted, however, that the next phase of the development of e-journals would not be to remodel the journal based on these early examples, but to digitize copies of printed journals. It is against this background that a more radical journal than was apparent or anticipated at that time was conceived as the basis for this investigation.

Today’s Web journals are clones of paper journals not just in terms of titles and formats but also in terms of the publishing process. Chapter 2, the literature survey, learns the lessons of early attempts to build electronic journals and identifies the emerging technologies that are redefining the journal publishing framework and have inspired debates re-articulating the role of journals in scholarly communication. Focusing on the period from the 1960s, alternate sections interleave parallel developments in electronic journals and hypertext systems, the latter highlighting the emergence of the Internet and the impact of the Web. Aiming to identify points of convergence between publishing and the supporting technology, the concepts of open systems, and open information, emerge as critical factors.

Chapter 3 considers e-journals specifically, beginning with the main findings of the survey of e-journals. At the end of 1995, the survey revealed, there were just over 100 peer-reviewed journals available in full-text form on the Web, barely over half of these from established commercial or learned society publishers. The process of digitizing established journals means that, by some

estimates, there are now over 10 000 e-journals, so analysis of e-journals must be more focused to spot significant new developments. This chapter examines three ‘snapshots’, collected e-journal descriptions, spanning the decade from 1991 – a particularly intense transitional period – charting progress and identifying the emerging, critical features that may shape the next-generation e-journals. One particular genre – free, open access e-journals – has not been wholly stifled by mass digitization and some especially successful examples are highlighted, notably physics journals published as ‘overlays’ to a collection of over 150k freely-available papers, the arXiv physics eprint archives.

‘Overlay’ journals perhaps highlight the most significant change in journal function: from package to service. Users will continue to browse core journals, but practices are changing. Online, research queries tend to be more precise. It is no longer sufficient to deliver a collection of papers as a journal from a single publishing site, or even to copy the package to the sites of commercial partners, such as journal aggregators. The journal as a package already has less relevance to the user. The emerging network architecture includes new services that transform the data between source and user, interposing machine–machine services in the client–sever architecture of the Web.

Chapter 4 considers the impact of these services on scholarly communication. Change is already happening in the academic library, embracing journal site licenses, services provided by distributed digital libraries, and the means of interconnecting rapidly expanding collections of digital resources, including new standards for linking such as OpenURL. The Open Archives initiative is just one influential organisation promoting interoperability, the ability of cooperating services to share data across networks, with implications for more open and freely-accessible scholarly resources. Development of machine–machine services in scholarly digital libraries is likely to be founded on a new form of distributed computing environment, the emerging standards-based architecture of ‘Web services’, and this is another consideration. Understanding this emerging infrastructure of scholarly communication and its implications will be pivotal to the acceptance of new journal models such as proposed in this thesis.

From the first e-journal experiments, developments in e-publishing span a generation; some of the underlying mechanisms that might support e-publishing have been known even longer; and some of the inspiring visions go back over half a century. There is no certainty that this period of change in the product of publishing is nearing an end. Chapter 5 takes a longer-term view of how scholarly publishing may be reshaped. It is necessarily more speculative than the preceding chapters. A simple view of the migration a new medium, not to just a new format, suggests that the greatest changes may be reserved for users, and leaves open the question of just how able we are to influence the changes that are reshaping scholarly communication, particularly in view of the converging but sometimes

conflicting goals and interests of the contributing partners, including authors, users, publishers and other service providers.

If the network-based services described in chapter 4 are to be effective, a user-based analysis suggests two critical steps are the decoupling of journal content from publishing process, and the defragmention of collections of content, the journals. Decoupling questions the traditional approach to rights, ownership and access. By deconstructing the process of journal production, chapter 5 shows how this change might be achieved. Defragmentation is reversing the process of specialisation forced on paper journals by page constraints and other factors, and the consequent stagnation in the ability of non-specialist users to access these works. Two primary tools supporting defragmentation are search and link. E-journals may have marked the transition to a post-Gutenberg era; defragmentation will accelerate as another seminal influence is assimilated, marking a post-Google era.

The original hypothesis for the new model e-journal of this thesis is reproduced in Chapter 6. This may be unusual, but it places the conception of the model chronologically some years ahead of actual implementation, which is described in Chapter 7. This allows comparison of the two stages, revealing whether changes were due to practical constraints or to changes in the prevailing infrastructure, perhaps with implications for the longer-term stability of the proposed model. That model re-implements the three key elements required of any scholarly journal publishing system: distribution and accessibility; academic recognition and validation; and placing new work within the context of the other works. Comparison with the original hypothesis informs the design of the evaluation of the model, and therefore users' reactions to it, as Chapter 9 reveals.

*Perspectives in Electronic Publishing* (PeP), as well as informing the title of this thesis, was also the title of the implementation of the proposed model. Chapter 7 offers a brief guided tour. A database-driven Web service, the primary unit of the model is the template for the database records. Examination of the template reveals the main design features. This chapter also outlines the ancestry of PeP within the IAM group (formerly the Multimedia group) at the University of Southampton, and compares and contrasts it with similar models elsewhere.

The principal tool for applying editorial management and control in the new model is a link service, and conceptually a key premise of the project was to evaluate reaction to a link-based, more technically a linkbase, journal. This is not the first application of a link service, and the design of the link service was not part of this work. Instead, Chapter 8 considers the requirements specification that shaped the implementation of the link service as part of the model represented by PeP. It was not a simple adaptation and imposed constraints on the editor and user interfaces. The reasons for this are

best understood alongside similar examples from annotations services and hypertext systems. Although not fully explored in the practical work of this thesis, there are intriguing possibilities for this model of linking to participate in more global linking services, based on the emerging OpenURL standard, and these options are briefly discussed.

The impact of any new model must ultimately be judged by the reaction of target users. Chapter 9, the heart of the thesis, analyses the results of a detailed evaluation among invited target users, authors of papers included in the journal, and other specialists in the field. The model is shown to have provoked divergent, occasionally controversial views that provide insights for the future of scholarly publishing generally. Detailed results are presented in Appendices 1–7; follow-up correspondence with evaluators, in Appendix 8, adds a more personal, and colourful, dimension to the analysis.

## 1.4 A note on terminology

One of the difficulties of emerging fields is adopting words and phrases that have a consistent meaning for all readers over a significant period of time: “terms such as electronic library, electronic journal, and electronic publishing all stem from a failure to stress that the core of the revolution we find ourselves in is not that existing systems and activities now have an electronic form, but that library, publishing and journal are archaic and obsolescent, if not yet obsolete, ideas.” (Wilson 1997)

In this field the journal prefix terms electronic- , digital-, networked- , online-, Internet- and Web-, which describe the range of forms for disseminating journals, are often used interchangeably. In fact, they can be used quite specifically. Broadly they describe decreasing subsets of these forms. Thus, 'e-journal' has been used most broadly to describe disc and network distribution, while the term 'Web-journal' is the most specific; almost invariably, non-paper journals today are Web journals.

Terms such as ‘Web’ journal indicate the origins of a particular journal, whether in print or the electronic medium. With the rapid conversion, or ‘digitization’, of most past and present print journals, the term digital journal acquires a distinctive meaning, although a more common description among specialists was to call these ‘parallel’ journals, since the print and electronic versions are derived simultaneously from the same source. This type of journal is recognizable from the page images of the original journal. Early examples of this genre were simply that – images – without the features that users of digital technology take for granted: searchability, text selection, etc. Today’s page images can more justifiably be called digital.

Journals without a print version might be referred to as electronic-only; most of the pioneering electronic journals are electronic-only. Those electronic-only journals that existed before the Web might refer to themselves as Internet-journals, often preserving those pre-Web text-only formats to emphasise the point.

So-called network or online journals are more difficult to define. Online information services, in the form of databases for example, have been available since the time of teletext in the early 1960s, giving the term ‘online’ a longer life-span than any other form defined here. In journal terms ‘online’ was used to indicate that a product could be accessed via a network rather than on a disc, although with the disappearance of disc-based journals this distinction is no longer necessary and the term is often used as a direct substitute for ‘electronic’.

Which leaves ‘network’ journal, an occasional and misunderstood term, for which a definition will be attempted here. Any journal accessed via the Web could be called a network journal, but that is simplistic and limiting. With the network increasingly being viewed as a computational resource rather than a transmission mechanism, there is a chance to establish a more useful form. Let’s consider network journals to have a computational interface, not just a user interface. Underlying this are computer language syntaxes like XML, a more advanced markup language for the Web than HTML, and metadata schemes such as the Resource Description Framework (RDF) and Dublin Core (DC). In this way something more powerful begins to emerge. At one level many journals offer bibliographic descriptions in DC, but let’s be more ambitious and retain the term to describe those journals that make the texts ‘open’, or accessible, for computational analysis too.

Where appropriate this thesis will aim to use the most specific term. Otherwise it will revert to using ‘electronic’. *Perspectives in Electronic Publishing*, the model at the centre of the thesis, can reasonably use the terms electronic-, online-, Web-, and possibly networked- (with some more effort). Whether it is a ‘journal’ is open to examination. Perhaps it might be better described as a “communication system”, which Wilson (1997) uses to describe not a “technological system – but any system of human communication”.

# 2 Literature survey: the convergence of journal publishing, technology and users

## 2.1 Electronic journals: why now?

Publishing is about people. The *Oxford English Dictionary* variously defines the verb to 'publish' as: to make *generally* known; to issue copies *for sale* to the public. In a business sense the two are inextricably linked, but these definitions show that publishing is not strictly limited to commercial activity; simply, in many cases, it is the most practical way. Since there is invariably a cost associated with the process of making something generally known, in whichever medium, it is accepted that a price will attach to the item to recover that cost. This works perfectly well, especially in the lucrative international markets for published products in which publishers put profit before their historical, perhaps mythical, obligation to the 'social good'.

There is an exception, however, to the efficiency of publishing markets. The published output of the worldwide scholarly community has risen dramatically. Odlyzko (1995) demonstrates the effect in the field of mathematics where over a period of a century and a quarter the number of papers published annually doubled every 20 years, but post-World War 2 that doubling has occurred every 10 years. More scholarly papers mean more journals, as new fields, sub-fields and increasingly specialised niches have developed. More journals have typically reduced the size of the market for other journals, however, with the result that prices have spiralled upwards. Some journal publishers have prospered as a result (McCabe 1999).

The consensus among those responsible for journal purchasing, librarians, was that the market for scholarly journals was beginning to break down (Okerson 1992), a process chronicled since 1989 by a newsletter (Tuttle 1991). An ongoing 30-year survey of the usage of scholarly journals offers a less dramatic view. This survey shows that usage of journal articles continues to increase and that the effective cost per reading has increased only minimally, due less to publisher and library costs, which have decreased slightly, than to an increase in scientists' time used to obtain and read articles (Tenopir and King 2000).

Thus, accessibility of the published scholarly literature appears to have diminished, due either to higher journal prices or because users have to process greater volumes of information to find what they need, or need to find more information, or all of these factors. The prospect offered by electronic

publishing of scholarly papers is to improve access by resolving these problems through more open access and improved information management tools.

It is recognised that the cost of printing words on paper is high. Scholars had no alternative but to accept the 'Faustian bargain', as Harnad's 'subversive proposal' (Okerson and O'Donnell 1995) called it, with this 'papyrocentric' publishing industry in order for their work to be distributed and read. The published work of scholars had to be paid for. Until now, that is. Electronic-only publishing, despite false early promise (Turoff and Hiltz 1982), makes free access to scholarly papers *for all* a viable alternative. Some contemporary electronic-only journals are proving to be more robust than their predecessors, although their number remains small relative to paper journals, hundreds rather than tens of thousands.

Rapid growth in the numbers of electronic journals is not due to original e-journals, however. Most paid-for scholarly journals that originated on paper now have an electronic version too. One major learned society, the Association for Computing Machinery, even published a plan to phase out print versions of its traditional esoteric journals and transactions as the demand for them becomes too small, "an outcome that may happen for some journals as early as 1998." (Denning and Rous 1995) It has not happened yet, but the trend towards greater electronic access and usage is clearly established.

Put simply, what has motivated electronic journal publishing on this scale since 1995 is the World Wide Web with its freely-available browser interfaces that enable information to be viewed easily using personal computers connected via a common communication infrastructure, the Internet. How did we reach this point, and what are the prospects for reshaping electronic journals publishing?

## 2.2 The hypertext influence

If publishing is people-oriented, a simplified analysis of online publishing that concentrates on technology is clearly going to be inadequate. In fact, what works in the electronic medium was identified at the end of the last World War, long before any of this technology was possible. Bush's (1945) seminal vision was based on our ability to create, store, organise, retrieve and understand information. His crucial insight was to recognise that although elaborate ways of indexing information had developed, the human mind doesn't operate in that way: "It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails." To present information in this form Bush envisaged a 'mechanized private file and library', a device he called a memex. This device was never implemented, but in its description Bush preempted much current technology. He anticipated the

information structure now referred to as hypertext, involving associative indexing, linking items of information and annotation. The memex's 'large translucent screens' perhaps foreshadowed today's windowing interfaces, and the reference to a 'web of trails' isn't the only parallel with today's Web as Bush also saw that a critical feature of his memex was its ability to be consulted with 'exceeding speed and flexibility'.

It was not until the 1965 that the term hypertext first appeared, and it was credited to Ted Nelson, who defined it as 'non-sequential writing'. In this respect hypertext is not confined to computers, but Nelson imagined that all information would one day be 'equally accessible' from 'one great repository' and readers could follow different pathways, so it was inevitable that computers would be intrinsically involved.

This was the starting point in 1960 for an ambitious project to create not any hypertext system, but a *universal* hypertext system, called Xanadu (Nelson 1987). This was to be that repository, and it appears that there was to be no scope for evolution in a system that would contain 'all of its ultimate features' as part of the design. Not so much a publishing system for the world, but *the* world publishing system.

Much of the initial acclaim for Nelson deflected attention from another pioneer, Douglas Engelbart, who developed a machine for 'the augmentation of human intellect' called NLS (the oN Line System) but which was later marketed as Augment. This system supported not just shared information, but collaborative authorship as well (Engelbart 1975), and has particular significance for those seeking to establish a greater role for authors in the dissemination of their work directly to users: "We have a strong feeling that most people working in such an environment will want to reach through to work with other people at least as often as with computer services."

## 2.3 A short history of electronic journals

While Nelson's massive Xanadu publishing project alternately inspired and floundered through the 1970s, elsewhere scientists were becoming interested in the more mundane goal of reproducing single journals electronically. It is over 300 years since the first scientific journals appeared simultaneously in 1665 in London (*Philosophical Transactions of the Royal Society of London*) and Paris (*Le Journal des Scavans*), as reviewed by Guedon (2001), these events themselves happening approximately 200 years after the invention of print (Schauder 1994). It is just 20 years since the *feasibility* of an electronic journal was discussed by Senders (1977). This discussion of a hypothetical

journal led to the first e-journal project, the Electronic Information Exchange System (EIES) developed between 1976 and 1980 at the New Jersey Institute of Technology (Sheriden *et al.* 1981).

What began as a model for a single e-journal evolved rapidly. A remarkable paper by Turoff and Hiltz (1982) describes what emerged: a newsletter; an unrefereed 'paper fair' or preprint journal; a refereed journal similar to the classic print-based model; and a subsystem for attaching enquiries, comments and notes to published materials. What is remarkable about the paper is that, despite a revolution in the supporting technology since EIES was implemented, in publishing terms it describes exactly the same problems and challenges that e-journals face today. Three elements of the EIES journal were successful and one, the refereed journal, failed. Only two articles were ever published.

A retrospective by Naylor and Geller (1995) identified conflicting views for this failure. Turoff, effectively the technical director, blamed the lack of motivation of the academic community at which the journal, *Mental Workload*, was aimed. Neville Moray, the journal editor, in contrast attributed it to the poor quality of the EIES system. One of the two articles to appear was written by John Senders, who was also part of the project team. According to Moray the article "took nearly 12 hours to enter into the electronic journal, and about 4 hours to edit." Users retrieved articles through 'dumb terminals' attached to minicomputers that were connected over a Telenet network. Interface design was primitive and variously described as 'very rich' or 'too complicated'.

Following a failed attempt to collaborate on EIES, due principally to restrictive national telecommunications policies in force at that time, a similar project was proposed in the UK. The Blend project at Loughborough University was funded from 1980 to 1984 (Shackel 1991). Learning from the EIES experience, the project constructed refereed 'journals' but these journals were not *published*, as was *Mental Workload*, merely archived, in electronic form only. Given the level of computer interconnectivity at that time, access would clearly have been limited. Papers could be published elsewhere, with acknowledgement.

Instead the main emphasis of the Blend project was on the editing cycle for papers: authoring, refereeing, leading to editorial acceptance. By 1984 the archive contained 18 papers on computer-human factors, the chosen subject, and 22 'poster', or unrefereed, papers. In addition, the archive stored software reviews and newsletters, and developed a reference database. In its experimentation with comment and discussion forums tied to papers, annotated abstracts and collaborative writing, the project was, again, a forerunner of many of today's electronic journals.

## 2.4 From hypertext to open systems

By the mid-1980s the early mainframe-based hypertext systems had been superseded by hypermedia systems using workstations, what Halasz called second-generation systems. The IBM Personal Computer had appeared at the beginning of the decade, and this provided the platform for the first commercial hypertext products. Ironically, the application that popularised hypertext ran not on PCs but on the then relatively new Apple Macintosh. HyperCard was bundled free with the Apple hardware and, with the aid of the Mac's advanced graphical user interface (GUI), users who had never heard of hypertext began building so-called HyperCard stacks (Hall 1994). "Suddenly, everyone is talking about hypertext", said Nelson (1987).

The scale of this activity can be gauged from Conklin's (1987) review. The most striking features of the work at the time were the differences between hypertext systems. Conklin did not provide a definition of hypertext systems but listed some of their characteristic features, differentiating a hypertext system from other computer applications. According to Conklin a hypertext system should have a single underlying database, a single coherent interface to the database, and a sophisticated notion of machine-supported links: not one or other of these features, but all three are essential.

Despite heightened activity in the field, concern was expressed that a quarter of a century after Nelson had introduced the idea, hypertext was still not being used as 'a fundamental tool for daily knowledge work' (Meyrowitz 1989). The reason, or 'missing link' as Meyrowitz argued, was that hypertext systems as applications were separate from others, forcing users to 'disown' their existing applications should they want to use hypertext. The solution was to enable hypertext to integrate with applications, an 'open systems' approach.

In computing the concept of 'open systems' was developed to enable proprietary applications to interoperate, that is, share data. The need for this was evident in all popular applications and for all platforms – Unix, PC and Macintosh. The key technical idea underlying the Internet was open-architecture networking, allowing different networks to interwork (Leiner *et al.*1997). Since the essence of hypermedia would be to integrate a wide range of applications supporting the production of text and audio-visual materials on a large scale, this was clearly an important concept.

Nelson (1987) referred to the need for open hypermedia to provide what he called a 'balance of rights' between authors or publishers and users. He cited interactive media in which the designer made it impossible for the user to choose a particular pathway through the material: "the author and publisher have a right to show you something *their* way; but users have a right to employ these things

as *they* choose." So Nelson viewed open hypermedia as the ability to add to or reuse parts of materials. Although Xanadu conformed to this view, Nelson persisted in referring to it as a *universal* open hypermedia environment, failing to recognise that in publishing terms the concepts of 'open' and 'universal' begin to conflict. As Cliff Morgan of the publisher Wiley has said in a different context: "the publishing industry is rather notorious for its culture of resistance to standardisation – the anti-Procrustean tendency is strongly represented!"

The first working 'open' hypermedia system, Intermedia, was developed at Brown University in the USA. Researchers at Brown had been experimenting with electronic document systems, all with an underlying hypertext structure, since the late 1960s. Intermedia, in effect Brown's fourth-generation system, was the first to include multimedia capabilities (Yankelovich *et al*. 1985). The system design assumed that users would want to create links as part of their regular work with word processors, spreadsheets, and graphics editors, for example. Intermedia was thus "a framework for a collection of tools that allow authors to make links between standard documents created with heterogeneous applications." Instead of integrating widely-used applications, however, Intermedia used customised front-end tools and in this sense could almost be viewed as a monolithic hypertext package that Meyrowitz, also at Brown, had deprecated. But the system introduced the paradigm of the 'navigational link', which could make connections between applications because the link data were not contained within any single application, and this has since become recognised as the measure of open hypertext.

Taking Conklin's description of a hypertext system as a database, an interface and links, the question an open hypertext system raises is how different applications can interact equally with the hypertext system to share and display links, which are stored in a link database or linkbase. This structure requires link data to be communicated to the applications. A general approach called a 'link service' was introduced by Pearl (1989). The link service is effectively a database lookup service where the data items are interpreted as links between other data items. A link service can potentially integrate a far wider range of third-party applications than realised by Intermedia, although this requires that the applications are link-aware, that is, able to communicate with the link service. There are drawbacks – maintaining valid links when documents change, for example, is more difficult in a system mediated by a link service – but Pearl envisaged, as did Meyrowitz, link services giving links the same utility as computer cut-and-paste operations.

A number of open hypermedia systems reported in the early 1990s adopted a link service approach, including Microcosm (Hall and Davis 1994), Hyper-G (Kappe *et al*. 1992) and Multicard (Rizk and

Sauter 1992), and there are other examples of open hypermedia systems in research (Wiil and Leggett 1997). Two of these systems, Microcosm and Hyper-G, were commercialised.

Describing Microcosm, Hall and Davis noted that advances in operating systems and GUIs were making integration between applications easier, but that this usually involves embedding virtual objects within applications: "To take advantage of the availability of multimedia information within our computer systems, we need the flexibility provided through the use of hypermedia link services." Within Microcosm link data are processed by filters and displayed, for a given application, in a viewer. A Microcosm viewer was developed for AutoCad, and some limited interaction was possible with Word. For applications without viewers, which were thus not link-aware, an intermediate solution providing a limited set of link functions was constructed using the Windows Clipboard (Davis *et al.* 1994).

The power of the link service approach becomes more evident in the range of processing functions Microcosm makes available to the user. Users can create and follow links using menu-based options. Links define a relationship between specific points in a set of documents, which could be text documents or audio-visual materials. In Microcosm a spectrum of link types can be created: the *specific* link is defined at a specified point in a document (when the link anchor is highlighted or coloured this is commonly known as a 'button' link); *local* links are defined on an object, say a common word or phrase, within a particular document, and apply to every instance of the object in that document; while the *generic* link extends the concept by applying to every instance of the object in any document within a defined collection of documents, known in Microcosm as an 'application'.

Within large and complex information environments the importance of this incremental range of link actions is apparent in the degree of control and level of automation that can variously be applied to create link structures. "The local and generic links, which have dynamic source anchors, are the features that enable the Microcosm link service to be applied to other applications." (Hall and Davis 1994) In the context of link structures, linkbases provide an added dimension. Linkbases act as an important filter process, and in the Microcosm architecture there can be any number of linkbases in the filter chain. Users can install and delete linkbases even when Microcosm is running, so different views of the data can be presented to different users. "On an even wider scale, filters can be built to generate links to and from wide area information servers such as WAIS, WWW and Gopher." (Hall and Davis 1994)

Microcosm and Hyper-G, as well as other linking models, have been extended to augment the Web, respectively, as the Distributed Link Service (DLS) (Carr *et al*. 1995) and HyperWave. (Maurer

1996). The DLS was initially aimed at ‘publishers, authors and readers’. In this area its impact can be measured in terms of influence rather than implementations, which can be understood by looking at the character of the electronic publications motivated by the emergence of the Internet as a public utility.

## 2.5 Early Internet ‘journals’

Where the early electronic journal experiments, EIES and Blend, were unable to distribute articles widely due to the limitations of telecommunications networks, electronic journals that published and transmitted full-text, peer-reviewed papers worldwide via a network had begun to appear by 1990. The catalyst for this new development was a global network, the Internet, the character of which today is defined by its broadly-based community of users and can be traced back over 30 years.

The transmission of data between computers requires a physical connection and a means of communicating, a protocol. The Internet grew out of a US military network, ARPAnet, which to conform to military requirements had to be fault-tolerant, that is, resistant to attack. This produced a peer-to-peer network rather than one with any centralised control, and transmitted information in packets using a packet-switched protocol first developed in the 1960s. According to Leiner *et al*. (1997) ARPAnet grew into the Internet based on the idea of multiple independent networks. In the 1970s other networks included satellite and radio networks, and these were enabled to interwork through the adoption of open architecture networking which describes not the network architecture but the way in which the different networks can communicate. This communication came to be based on two protocols, TCP/IP, to control the addressing and forwarding, or mailing, functions (IP) and to manage the information in transit (TCP).

Critically, the adoption of TCP/IP as a US defense standard in 1980 enabled ARPAnet to be split into its military operations and research communities, the latter thus being freed to work with emerging network communities such as exemplified by the national networks in higher education, NSFNET in the USA and JANET in the UK. With the mandatory adoption of TCP/IP by NSFNET in 1985 the Internet began to support a broad community of researchers and developers. At the same time the policy of reducing the dependence of NSFNET on government funding encouraged commercial, non-academic users but these users were restricted regionally. The resulting competitive long-distance networks were only able to buy Internet connectivity upon the privatisation of NSFNET in 1995. Such was the demand for Internet services that even these proprietary networks, commercially designed to lock subscribers in and keep non-subscribers out, built ‘gateways’ to the Internet.

For most Internet users it is not the technicalities of this infrastructure that are important but the universality of information services that can be accessed (Krol 1994). Prior to the Web, the most popular of these services were electronic mail, for one-to-one and list messaging, and news lists for posting to larger groups. For larger data transfer, the Internet could be used to log into remote computers and transfer data files using the file transfer protocol (FTP).

Technicalities did intrude, however, in the earliest implementations of these services. Users needed to acquire what might generally be regarded as specialist computer skills, for example, some knowledge of Unix, or an understanding of computer file structures or database structures – albeit on a limited scale. So when "one of the first electronic journals distributed via computer networks" appeared in 1987, *New Horizons in Adult Education*, readers are unlikely to have found a simple or even intuitive interface. Nor would the journal have had typographical design or graphical images, just plain text.

This aspect of presentation was not the primary concern of the first online journal to make an impact, *Psycoloquy*, in 1990. Its founder, Stevan Harnad, also edited a paper-based journal, *Behavioral and Brain Sciences*, which demonstrated the idea of open peer commentary. It is standard practice for scholarly journals to peer-review submitted articles before agreeing to publish. These reviews are tendered anonymously to authors and are never published. Open peer commentary, in contrast, actively encourages comments to be posted publicly, by invited and non-invited contributors, while also using anonymous refereeing on which to base editorial judgement preceding formal publication.

As a cognitive science specialist, Harnad recognised the potential of virtually instantaneous electronic dissemination over the Internet to enhance open peer commentary. In the new medium he renamed this 'scholarly skywriting' (Harnad 1990), so-called because it is "as if each contribution were being written in the sky, for all peers to see and append to." (Harnad 1991)

*Psycoloquy* demonstrates skywriting. Underlying the simple text presentation is a sophisticated vision of a new type of publication, unconstrained by commercial interests, unfettered by the temporal and spatial constraints of print, and free to explore the cognitive potential of the new medium.

## 2.6 Eprints: complement or competitor to the journal monopoly?

If electronic communication is to liberate scholarly papers to be viewed by all, it was never expected this would be achieved by independent, free e-journals alone. Gardner (1990) proposed an electronic archive that he envisaged being managed by a learned society publisher, in effect a collection of journal papers. In this respect the idea was closer to Bernal's depository from 1948, and the

dissemination of separate articles in paper form, which a number of projects had shown to be ineffective (Piternick 1989). Gardner saw the primary advantage of the electronic archive, however, as the improved facility for discovering, retrieving and using information.

Harnad (1990) argued for archives that go "far beyond providing searchable electronic archives for electronic journals", likening to an 'intellectual perestroika' the prospect of scholarly skywriting combined with freely accessible electronic archives of papers in the PREPUBLICATION (author's capitals) phase of scientific enquiry, "the phase in which most of the cognitive work is done". He claimed this would be "incomparably more thorough and systematic, potentially global... so unprecedentedly interactive" that it would "restructure the pursuit of knowledge." Subsequently Harnad (1996) predicted that scholarly skywriting would "increase individual scholars' productivity by an order of magnitude".

Such an archive soon materialised, based on an automated system developed in 1991 by physicist Paul Ginsparg at the Los Alamos National Laboratory, USA (Ginsparg 1994). Specialising initially in high-energy physics, what became known as arXiv has grown phenomenally, encompassing all other fields in the discipline. Over 150,000 papers have been posted to the archives, and with over 30,000 new submissions in 2000 the physics archives exhibit an approximately linear growth rate.

ArXiv is frequently and inaccurately referred to as a 'preprint' archive, but the physics archives perform a service substantially in excess of a repository of papers awaiting formal publication. Simply in terms of usage, the system serves over 70,000 users in over 100 countries and attracts up to 130,000 visits daily (Luce 2001). Recent studies have shown this level of usage to be significant. Citations to electronic eprints such as those accessible from the physics archives "have nearly doubled every year since 1992." (Youngen 1998) A study of arXiv showed that in the case of highly cited papers there is a significant positive correlation between how often a paper is cited and how often it is downloaded (Harnad *et al*. 2001), showing how access can enhance impact for the best papers. The effect is not confined to physics, but extends to other disciplines that are less well organized on the Web: "More highly cited articles, and more recent articles, are significantly more likely to be online, in computer science." (Lawrence 2001)

This ought not to be surprising. Communication via arXiv is formal. Although papers are unrefereed, the archives can be used by other services as the basis for refereeing papers (Taubes 1996a), most obviously in the form of 'overlay' journals. Ginsparg (1994) always anticipated that the 'data highways' would have to 'reimplement the protective physical and social isolation enjoyed by ivory towers and research laboratories'. Till (2001) contends that 70% of papers in arXiv are eventually

published in journals. In many cases copies of the peer-reviewed versions that appear in journals are posted to arXiv, replacing the original preprints. In this sense it can be seen why arXiv is more accurately referred to as an ‘eprint’ archive. The result of this deceptively sophisticated re-engineering of the dissemination process for papers in physics, according to Brown (2001), is that “eprints are used to a greater extent by physicists than previously measured and that eprints have become an integral and valid component of the literature of physics.”

The success of arXiv was not guaranteed. In the 1960s some biomedical science articles, which were treated as personal communications, were distributed through Information Exchange Groups (IEGs), simply organised mailing lists (Green 1967). In the early 1960s one of the seven IEGs was processing 'at least 90 per cent' of important papers in its field, according to Green. At that point “the editors of five biomedical journals met and agreed to refuse publication of any manuscript previously circulated via IEG. This unaccountable decision turned out to be lethal to IEG.” Green went on to remark that the type of dialogue fostered by IEGs “has virtually disappeared from the scientific journals”.

Till (2001) compares the experiences of the IEGs with arXiv to identify the important factors that led to the latter’s success: “The most crucial factor may have been the extent to which the author, reader, and reviewer communities have coincided, and thus have had an implicit agreement (or ‘scholarly consensus’) about standards of quality for research considered to be acceptable for publication.”

Just as importantly, physics had a pre-existing preprint culture; authors printed and circulated papers to colleagues prior to journal submission. It helped that by the end of the 1980s virtually all physicists were interconnected through the Internet. Thus Ginsparg did not have to market a new product, but instead re-engineered an existing service in a markedly more efficient form.

It is harder to determine the origins of the model on which the working implementation of arXiv was based, perhaps unsurprisingly since the system was simply described by Ginsparg as written in ‘a few summer afternoons’. Papers about arXiv give few clues in this respect, but comments given to the *New York Times* (Overbye 2001) point to Cornell physicist David Mermin as a contemporary influence. A few months before arXiv appeared, Mermin (1991) wrote about distributing a preprint of a paper he expected less than a dozen people to read seriously. The paper was ‘squeezed’ into eight pages so it could be printed on two sheets to save postage and copying costs. Finally he produced four copies that were sent to the journal *Physical Review Letters*, a step he described as “the only one that casts doubt on my judgment, seriousness of purpose and moral integrity”. The process of printing the work on paper was barely tolerable either: “In a rational world, paper, printing, postage and *PRL* would never have crossed my mind. I would simply have E-mailed my essay to a central

clearinghouse for posting on its electronic bulletin board." Ginsparg's archive became that clearinghouse, although it is explicitly not an informal bulletin board.

Antipathy between journals and the free circulation of papers was not confined to biomedical journal editors. In language that Ginsparg would subsequently echo, Mermin pronounced "journals are obsolete except as archival repositories".

Intended to be based on journals, Gardner (1990) did not describe an electronic archive that would be free. According to Ginsparg systems such as arXiv cost so little to set up and maintain that they can be offered 'virtually free'. Following a system 'spasm' caused by Ginsparg's protest over funding (Taubes 1993), and a later grant of $1M from the US National Science Foundation, the physics eprint archive remains free to users.

On the scale of cost to users, the publisher Elsevier, which also publishes in physics among other areas, is at the opposite extreme to eprint archives. Elsevier is has a reputation for its predatory pricing policies, but must be regarded as highly successful commercially. It was also one of the first major publishers to recognise the potential of electronic publishing, and in 1991 it began the ambitious Tulip project. This investigated, with nine US universities, the networked delivery of scanned page images and the 'raw' ascii text of 43 journals, expanded to 83 journals in 1995.

The project concluded that year and the final report (Borghuis *et al.* 1996), although self-serving in some respects, is a rare and candid insight into the operations of a major multinational publisher. Crucially, the report highlighted that "the universities and Elsevier Science have not resolved one critical issue, that of how to make the transition to digital libraries work economically", for which read 'highly profitably'. Despite the thoroughness of the project, it was always likely to be undone by this dogmatic attachment to a commercial model that, however successful for paper journals, cannot simply be applied to e-journals. Since it was predetermined that charges would have to be maintained at the levels of paper journals, page images would have to be preserved to associate the electronic product with the journals to justify the price. So when, unforeseen, the popularity of the Web began to surge midway through the project, the publisher's response was to use it merely as a transmission medium rather than adapt the content to the new information structures now possible. As a result, file sizes for individual papers were large, incurring high storage and cost overheads; in addition there were printing problems, and response times were 'too slow'.

Among 'an accumulation of discouraging factors' was insufficient coverage: not all core journals, those not published by Elsevier, were available in the database and required end users to search

additional information elsewhere. Curiously, such a large-scale project produced ‘no definite answer’ on the relative importance of coverage versus convenience. No doubt Elsevier was reluctant to hear the answer as demonstrated by the physics eprint archives, because there lies the real potential for e-journals publishing and the fraying of the substantial influence of publishers like Elsevier, a factor the publisher belatedly appears to recognize (Guedon 2001).

## 2.7 The Web challenges the printed page

The impact of the Web may have surprised publishers in the mid-1990s; it is now recognized as a mainstream form of communication. By 1991 sophisticated hypertext systems had begun to appear, but the Web was prompted by a ‘home brew’ personal hypertext system used to keep track of personal information in a distributed project (Berners-Lee *et al.* 1994). Simply, if the relationship between two projects changed the information “could smoothly reshape to represent the new state of knowledge”. It is this property of scaling, the originators argue – with the Web it is easy to address an object anywhere on the Internet – that has allowed the Web to expand rapidly. This is not to mention the popularity inspired by the simple and intuitive browser interfaces based on NCSA’s Mosaic, through which Web documents can be viewed, and the simple utility of the HTML markup language, based on the SGML standard, from which the ‘fabric’ of the Web is constructed.

Despite the inference of its full title, the Standardised Generalised Markup Language, SGML is not a markup language: it prescribes how markup should be specified, not what that markup is (Barron 1989). In other words, to generate a printed page or document in a readable, or viewable, form from its SGML description requires an underlying formatter or front-end translator. HTML adheres to SGML as an instance of a document type definition (DTD) that can be parsed by an SGML parser, but where SGML is regarded as a format for long-term storage and maximum flexibility, HTML is simply a communication language. HTML documents are plain text files with 'tags' or codes that are interpreted by the presentation system, typically a Web browser, using a specified DTD. A set of tags is defined in the recommended HTML standard of the moment.

Where HTML is easy to use but limited, SGML is powerful and flexible, but creating and maintaining SGML publishing applications is labour-intensive (Wusteman 1997) and many features in SGML are not needed for Web applications. The solution is XML (Extensible Markup Language), standardised in 1998 by the World Wide Web Consortium (W3C), a broad-interest body formed to maintain and develop standards for the Web. Like SGML, XML is not a language but a meta-language, a tool for writing languages, designed to “enable generic SGML to be served, received, and processed on the Web” in the same way as HTML (Bray *et al*. 1998).

XML effectively marks the convergence of decades of computing development supporting interoperable data across computers and networks. XML derives from a philosophy that data belongs to its creators and that content providers are best served by a data format that does not bind them to particular script languages, authoring tools and delivery engines (Bosak 1997).

The most compelling vision for XML is database interchange, what Bosak calls 'the universal hub', but other critical applications have emerged to assure the success of XML. For example, XML is becoming fundamental to the dissemination and interchange of multimedia documents, with languages such as the Synchronized Multimedia Integration Language (SMIL) (Ayars *et al*. 2001).

Structured documents can be created using the Resource Description Framework (RDF), based on XML syntax (Lassila and Swick 1999). Where Web content was originally mostly textual, one of the objectives of XML was to motivate a more data-driven Web, where data did not have to be human-readable but could be understood through machine processing. The broad solution was to use metadata to describe the data contained on the Web. A particularly important metadata schema for digital library applications is the Dublin Core, which was proposed as the minimum number of metadata elements required to facilitate the discovery of document-like objects in a networked environment such as the Internet (Weibel 1995). At the time the syntax of the Dublin Core was left unspecified as an implementation detail. Dublin Core metadata elements can be represented in many syntax formats, although RDF is expected to be the most widely used because it establishes an encoding scheme, and DTDs and XML schema to validate DC documents written in RDF are likely to become available for wide reuse.

Dublin Core is defined by a large community of users whose primary concern is the description of academic and scholarly resources. Other communities may develop metadata applications described using similar terms but which have different meanings, or semantics. RDF provides the ability for resource description communities to define semantics and to disambiguate these semantics among communities. The property-type 'author', for example, may have broader or narrower meaning in different communities.

RDF is thus pivotal in the development of structured metadata. The hierarchy is: XML (the format) – RDF (an XML application; the infrastructure, imposing structural constraints) – the metadata element set (e.g. Dublin Core). Based on this hierarchy of standards, RDF has become established for a variety of applications from library catalogues and directories to syndication and aggregation of news, software, and content, to personal collections of music and photographs.

XML can also be viewed as a set of rules that enable markup of complex scientific and mathematical notations: two such markup languages (MLs) are MathML and ChemistryML. For journals these markup languages provide Web presentation standards to compete with document delivery standards such as Postscript and PDF. An XML browser has been developed for chemistry.

As well as providing greater support for database applications on the Web and document validation, XML is designed to work with 'classic' hypertext linking mechanisms, the sort that were 'built and proven during the 1970s and 1980s', including third-party managed links (Bosak 1997). The principal linking mechanisms defined in XML are XLink, used to create structures that can describe the simple unidirectional hyperlinks of HTML as well as more sophisticated inter-resource links, and XPointer, a fragment identifier that allows precise parts of resources to be targetted by links and supports intra-resource linking. Systems such as Microcosm and Intermedia were explicitly recognised as formative influences in the W3C Recommendation, effectively the standard, for XLink (DeRose *et al*. 2001).

A feature of Web presentation is that it concentrates on the structure of information, in particular the hypertext link structure, rather than page appearance. Text fills the browser window, reflowing and reshaping dynamically to fit adjustments in window size. For hypermedia documents, which have no printed equivalent, this flexibility is reasonable, yet experience of using Web pages suggests that there is a desire among designers to exert more control. The use of style sheets, which describe how documents are presented on screen and in print, returns some control over page appearance to the HTML designer. Separate tags can be applied for text and style (e.g. fonts, colors and spacing).

The best-known application is Cascading Style Sheets (CSS). The ‘cascading’ feature enables authors to attach a preferred style sheet, while the reader may alternatively have a personal style sheet (Lie and Bos 1997). Level 1 CSS, the original specification, was designed to ensure all browsers could understand basic text and page structure even if some proprietary presentation commands are used. Later levels of CSS support media-specific style sheets so that authors can tailor the presentation of their documents to visual browsers, aural devices, printers, handhelds and other devices. The CSS specification also supports content positioning, table layout, features for internationalization, and some user interface properties. CSS is widely implemented in Web browsers.

CSS can be used to style HTML and XML documents. Another approach to style sheets, the Extensible Stylesheet Language (XSL), is able to transform documents. For example, XSL can be used to transform XML data into HTML/CSS documents. In this way, the two languages complement each other and can be used together.

XSL is a language for expressing style sheets. It consists of three parts: a language for transforming XML documents (XSL Transformations, XSLT), an expression language used by XSLT to access or refer to parts of an XML document (XML Path Language, XPath) and an XML vocabulary for specifying formatting semantics (XSL Formatting Objects). An XSL style sheet specifies the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document that uses the formatting vocabulary.

Like XSL, the Document Style Semantics and Specification Language (DSSSL) is a document tree transformation and style language. DSSSL is mainly intended for processing SGML documents but, since XML 1.0 documents are also SGML documents, a DSSSL engine can process them. According to Martin (2000) DSSSL has traits redolent of both CSS and XSLT. As in CSS, each formatting object is specified with a set of properties. As in XSLT, processing order is controlled by a processing construct. DSSSL is the best choice for transforming XML into documents for print purposes, Martin says. The nearest W3C-blessed alternative is XSL formatting objects.

Style sheets, and style sheet programming languages which can support more complex markup such as XML, will simplify the maintenance of Web sites, it is claimed.

For publishing applications the issue of page appearance against document structure has a longer history. ‘Does page fidelity matter?’ asked Brailsford (1994). There are powerful pressures for a page-based approach: reader familiarity with the printed page, and the desire of publishers to retain control over page appearance, which for many is almost a trademark. Further, most archive material is in page form already, Brailsford noted.

Just as Elsevier’s Tulip project used scanned page images, so other early experiments in network journal publishing used scan methods. The Core project at Cornell University (Entlich *et al.* 1995) reproduced a large volume of journals in chemistry using Bellcore’s PixLook searchable page imaging software. An SGML copy was also obtained. A library of medical journals was scanned using AT&T’s RightPages (Story *et al.* 1992) in the Red Sage project at the University of California-San Francisco (Lucier and Brantley 1995). In addition to converting pages, the RightPages prototype system had a facility for users to order articles. The Torpedo project at the US National Research Laboratory, also involving the American Physical Society, used optical character recognition techniques allied to a database storage system.

The most successful example of a scanned page image service supplying copies of archival journals is JSTOR. For most material in the JSTOR database, which is dominated by black-and-white textual documents, “we believe 600 dpi bi-tonal TIFF images represent a stable archival standard” (Guthrie 2000a). The same level of stability has not yet been achieved for photographs, and “new scans may have to be made to capture acceptable levels of detail”. By 2000 the JSTOR database included the backfiles of 117 journal titles, covering nearly 5,000,000 pages, with more than 650 participating academic institutions from 30 countries (Guthrie 2000b).

In 1993 the CAJUN project at Nottingham University set more stringent requirements, seeking a means to preserve the ‘look and feel’ of a printed journal without resorting to simple page images. The desired system had to support document browsing using hypertext links and text search, and it should be available to viewers on all popular computer platforms (Smith *et al.* 1993).

For nearly two decades Adobe Postscript has been the industry-standard page description language, that is, a means of printing a faithful copy of a page originated in electronic form with, say, a word processor or desk-top publishing system. When it introduced Acrobat, applying a document representation called the portable document format (PDF), Adobe provided the means to disseminate or transmit faithful copies of page originals in electronic form, either on disc or over networks, with the facility to convert to PDF from Postscript. All the features sought by the CAJUN project had thus been added to display Postscript.

A significant feature of PDF is the file compression options, which are robust for network transmission and reduce the file size overhead, the issue that seriously limits most page image formats. Although there are other portable document applications, Acrobat dominates in the production of e-journals.

As the Tulip project revealed, however, too great an emphasis on the conventional publishing framework, as demonstrated in the need for page fidelity among other features, obscures a wider understanding of the demands of publishing across a computer network, as epitomized by the Web. Greater user control over selection and presentation, for many the essence of the Web, fundamentally alters the parameters assumed in previous publishing models.

## 2.8 Open information systems

The Web has become the classic open system, with its standard protocol for information transfer and universal addresses: anything that can be displayed can be interconnected. The Web is not the only

example of an open system denying attempts to impose proprietary solutions. Bank cashpoint networks and airline reservation systems which particular companies used to seek monopoly power were defeated by consumer choice (Browning 1996). Huge investment in online services such as America Online attracted large numbers of subscribers into what was intended to be a closed network, but it had to provide a gateway to the 'open' Web.

The Web though is not an open *hypertext* system. A generally accepted requirement of open hypertext systems is that they do not differentiate between authors and readers (Malcolm *et al*. 1991). Each should be offered the same set of functions, that is, a reader should have the same facility for altering a version of a text, say, as the original author. The implications are enormous. From a technical viewpoint systems such as Microcosm, which gives users the option to make as well as follow links, conform to this requirement.

By encoding links within HTML markup, the Web does not conform to this view, because it restricts link functionality and reduces linking to an authoring task. According to Berners-Lee *et al.* (1994): "The Web does not yet meet its design goal as being a pool of knowledge that is as easy to update as to read. The level of immediacy of knowledge sharing waits for easy-to-use hypertext editors to be generally available on most platforms." This does not extend link authoring capabilities to readers, however. In the standard Web environment a document presented with no authored links can have no links added to it. It is connected within the Web by other documents linking to it, but it is always a dead-end for linking outwards.

One solution is to augment the hypertext capabilities of the Web with an independent link service, as envisaged by Hall and Davis (1994). The DLS incorporated many of the open hypermedia features of Microcosm to enable hypertext links to be added to any document presented on the Web "whether or not it provides link following facilities itself" (Carr *et al.* 1995). The DLS uses the concept of links as independent entities. In a networked environment the links are distributed across a number of linkbases, and the linkbases can be distributed across a number of servers: the document server may provide a basic set of linkbases, the end user may have a private set of linkbases, and a third-party server may provide specialist links.

In addition to the author–reader provisions, the DLS introduces a new publishing activity: publishers can act as 'link brokers' (Carr *et al.* 1995). A user can choose a selection from a document or an entire document, and the link service will return that document with all the applicable links (from whichever linkbase was chosen) superimposed on to the document. This facility also allows publishers to create information packages for niche markets based on a Web of information nodes,

and to recreate a different Web over the same information for different users. The Open Journal project explored the potential for this publishing capability by interlinking published materials for different user groups (Hitchcock *et al*. 1998b).

As with Sun's Link Service (Pearl 1989), the DLS recognises the problem of maintaining link consistency across a range of diverse resources over which it has no control; on a worldwide distributed system such as the Web the problem is magnified. Further complications are added by the functionality of the DLS, which allows the host server software to access linkbases stored anywhere on the Web by specifying a URL. So not only are the source and destination documents not under the author's control, but neither are the links.

This is a particularly open example of an open information system (OIS), which trades link flexibility and the ability to integrate a wide range of documents with the possible loss of coherence. HyperWave, another OIS – open in the sense in that it stores link data separately from documents in linkbases – solves the problem of coherence by combining a closed document management system with an open link service, what Carr *et al.* (1996) call a 'partially open' system.

HyperWave has a superficially similar architecture to the Web, and in this respect sought not to augment the Web, as does the DLS, but ultimately to replace it (Flohr 1995). HyperWave works optimally with its own client browsers that communicate with HyperWave network servers. Each HyperWave server maintains a document management system.

HyperWave was described by Nelson as the system that came closest to realizing his vision of Xanadu, but the prospects for a *universal* publishing system seem no better than before. From the evidence it appears that consumers are prepared to tolerate product competition when that competition is based on, or appears to be based on, an open commercial framework. Anything that might compromise the openness of the framework is likely to be distrusted.

## 2.9 Hypertext and the Web: the new missing link

It is over a decade since Meyrowitz (1989) identified the 'missing link' in hypertext development, the element that would transform hypertext into a 'tool for daily knowledge work'. With the emergence and mass popularity of the Web it might be assumed that hypertext research had fulfilled its mission, but in its present form the Web is disdained by many in the hypertext community. The list of 'missing links' has grown (Bieber *et al*. 1997): enhanced functions such as support for annotation, link typing,

collaboration, synchronisation and non-text links among others. Nor does the Web support Meyrowitz's original missing link, of separating link data from text, or open hypertext linking.

XML and associated linking developments such as XLink and XPointer demonstrate some progress in applying hypertext principles to the Web, and the systems described above which seek to augment the Web with hypertext functionality (Carr *et al.* 1995, Maurer 1996) were commercialised. Yet there is a sense that these efforts emphasise functionality while users remain elusive.

'Linking is harmful ... downright deadly' (Nurnberg *et al.* 1997). The link, the most evident feature of any hypertext system and its chief attraction for users, is at the same time held to be responsible for the two great problems of hypertext systems: cognitive overhead caused by imposing multiple tasks on the user, and disorientation or 'lost in hyperspace' in which the user may find it difficult to discover useful paths through items of information (Conklin 1987). Both effects arise as the result of the relationships between texts that links represent, rather than links per se. Carr (1994) examined the relationship between links and hypertexts by looking at the structure of texts, pointing out that even conventional texts on paper may be structurally complex but that familiarity with this form enables the information embedded in the text to be easily assimilated by the reader.

The adverse user effects of hypertext are often attributed to link design, but the system-based approach to hypertext design is another factor, often constraining link design within system capability, however flexible that system might appear to be: “By focussing on tools and methods for designers, we tend to overlook readers.” (Marshall 1997) There seem, then, to be too few good studies of user needs to inform hypertext system design, but such studies as there are often discover that users manage information in unexpected ways (e.g. Marshall and Shipman 1997).

Halasz (1991) admitted that hypertext had been based on a system-centric view while recognising that it needed to become a more user-centred technology. Halasz sought to broaden the idea of hypertext from the, at that time, standard node-and-link form by advocating hypertexts without links – 'ending the tyranny of the link' – based on computational and virtual hypertexts as well as other sophisticated navigational aids such as structure search. Hypertext was not simply a system technology, however. Halasz also noted a class of hypertext user, the 'literati' he called them, who use hypertext as a writing space, which he differentiated from the 'engineer' to whom hypertext is an information tool.

While some argue that hypertext causes the problems of cognitive overhead and disorientation, others like Halasz insist that the effective implementation of advanced hypertext features will alleviate the

problem. This may be justifiable for Web-based corporate networks, or intranets, rather than the Internet, where it is easier to motivate the prescribed hypertext functionality (Balasubramanian *et al*. 1997, Hill *et al*. 1997) because the user requirements are typically better specified.

The missing link that the Web discovered was the general user, releasing the latent demand for distributed information through its relative simplicity based on popular browser interfaces. The challenge for the hypertext and Web development communities is to make the case for advanced functionality while continuing to make as much as possible transparent to the user. In this sense the two communities should be implicitly interrelated in ways that do not always appear to be the case. To succeed, hypertext system developers must identify appropriate user environments.

## 2.10 Applying open access principles to journal content

If computer systems can be ‘open’, the Web is ‘open’, hypertext linking services can be ‘open’, and software can be ‘open source’, the convergence of these tools and services with information systems anticipates the next step (Barlow 1994, Bauwens 1996, Dyson 1994), which is that information itself will become open in some cases. In 1999 the *Journal of Electronic Publishing* included a paper on Red Hat software (Young), a commercially packaged version of the open-source Linux operating system. No explicit editorial connection with publishing issues was made, except to comment that the Internet and the Web have “brought about fresh business opportunities for those involved with intellectual property.” (LeJeune 1999) Implicitly the suggestion is that the published papers, the intellectual property of scholarly journals, should be available as open information sources. As in the Red Hat case, this does not mean that this information must be exclusively free; nor need the content be exclusively owned, by journal publishers say, as is typically the case now. The original papers may be free; the packaged versions might not.

ArXiv and early Internet journals offered free access to scholarly papers, but the principle is better described as ‘open access’ rather than free. This avoids the inference that such services cost nothing to maintain, which they do not. They are free to users simply because the costs are minimized to such a level that they can be covered by means other than payment by users or their agents. What makes this possible, and characterises these services, is a precisely defined core functionality coupled with varying degrees of automation that together produce high cost efficiency.

The founder of arXiv, Paul Ginsparg, recalled a seminal conversation in 1991 when he realised the preference of many physicists to post pre-publication copies of complete papers to a network of colleagues was becoming unmanageable (Overbye 2001). Ginsparg recast this as an electronic

storage problem, which he solved by introducing an organisational layer to manage access to these papers and nothing else – no peer review, for example – that could not be automated.

ArXiv could make this service available free by exploiting the connection of physics authors to the pre-Web Internet, and automating the remainder of the service on cheap computer equipment. In doing so it became a focal point for a community, as any publication seeks to do. ArXiv is now funded by the US National Science Foundation but remains a dramatic demonstration of how costs to support basic dissemination and distribution of freely-provided scholarly papers can be reduced to a minimum, as shown by Ginsparg's comparison of arXiv's cost per article with other types of publication (Fig. 2.1).

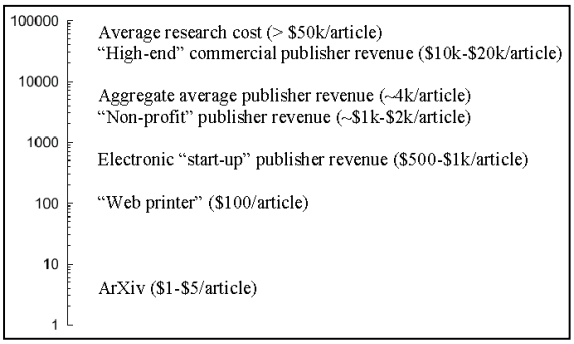



**Figure 2.1 Hierarchy of per article costs and revenues (log scale) (from Ginsparg 2001)**

Open access begins with authors. Researchers could print copies of a paper they had written and send it to colleagues, as Mermin (1991) did. Electronically, a single copy of that same paper posted on a Web server can be viewed by anyone with a Web browser and connection. As a means of publication that confers wider recognition this is insufficient, but as a basic means of dissemination it works well and establishes a new benchmark for publication that, when distribution is ubiquitous, must do more than simply distribute a work.

Open access poses another challenge to conventional paid-for journals, not just from author-posted papers as in arXiv, but from enhanced versions too where 'suppliers' pay for the costs of development and maintenance. This is another consequence of the cost-effectiveness of electronic distribution and the subsequent repackaging of value-adding services and features that this makes possible. In scholarly publishing a 'supplier' may be a university, for example, which already has a substantial investment in its communication infrastructure that it can exploit where there is direct benefit to its members. According to Arms (2000): "When universities are faced with the need to trim their budgets, the scientists will vote for the services that they see as being most valuable. Rather than support the library's journal budget, many will give priority to the staff who edit, design and publish their research on the web." The practice is not yet widespread, as according to Hunter (2001) the "perception of publishing to the Web as a whole business process is currently weak in most university administrations".

Arms was responsible in 1995 for launching *D-Lib* magazine, a conventional, high quality journal in all respects but it is electronic-only and open access by virtue of sponsorship from the Defense Advanced Research Projects Agency (DARPA) on behalf of the US Digital Libraries Initiative.

It is generally recognized that the benefits of open access begin to accrue when a critical mass of content becomes available and can be interconnected, again as demonstrated by arXiv. Elsewhere the volume of open access content is limited relative to the total output of scholarly journals. Given the apparent simplicity of Mermin's procedure for author self-posting of papers, a number of forums have addressed the reasons for this puzzling lack of open access content.

First among these was the September98-Forum, an email discussion list so named after a paper on 'Free Internet Access to Traditional Journals' (Walker 1998) featured in the issue of *American Scientist* magazine for that month. A commentary on the Walker article by the forum's moderator was the first posting to appear (Harnad 1998a). The agenda of the forum is precisely specified as 'freeing the refereed literature', which focuses the debate on open access to journal papers. September98-Forum presents vigorous debate, often from expert commentators, attaining a level of formality that was typified in the publication of an excerpted dialogue from the forum in the journal *Culture Machine* (Harnad *et al*. 2000).

The mantra of the moderator is that freeing the refereed literature is 'optimal and inevitable', a claim that was examined by Kling and McKim (2000), concluding that such utopian visions "do not effectively engage the complexities of the social worlds of the likely users of new technologies".

The forum appears to have been less successful in motivating the changes that it, but not all of its participants, promotes. The reasons for this have been rehearsed many times over by participants and have been summarized by Harnad (2001b). Another reason may be because the forum is overlaid with conflicting values, and is too concerned with the means – journals, refereeing – rather than the ends, free and open access. This is speculative, but it demonstrates how contentious the issues surrounding open access have become.

A significant addition to the open access movement is the incisive *Free Online Scholarship Newsletter* (FOSN), which first appeared publicly in March 2001. This adopts the broader view that all scholarly literature, not just journal papers but books and multimedia presentations too, could be free, both in the sense of free of charge for the reader and uncensored, and on the Internet in a form in which "its full text can be read, copied, linked, printed, crawled, downloaded, and searched". Nor does it make demands: "The only scholarship that should be free is that which its authors want to be

free." The FOSN agenda accepts that providing free online scholarship is "not without problems", notably preservation and author recognition, and admits that rates of adoption will vary: "Different disciplines have different needs." (Suber 2001)

According to Kling and McKim (2000), differences in the use of electronic media "can be characterised in terms of each field's articulation of some key issues faced by all scholarly fields: First, the allocation of credit for work performed; second, selection of target audiences for research; third, access to resources, including data; fourth, speed of work and results-sharing; fifth, allocation of professional status. The manner in which each field deals with these issues are both socially shaped and strongly institutionalized."

A good example was highlighted by Glass (2000) who described the prospects of 'preprint' archives playing a central role in scholarly communications in the 'soft' sciences – social scientists and notably educationalists – as 'dismal'. He attributed this to these disciplines placing "senior educators in less control over the intellectual knowledge due to the ease with which even novices can challenge favored theories and published studies. ... Where even the highest are subject to critique by the lowest, a premium is placed on tribal leaders and gate-keeping and orthodoxy of thought – hence, the premium placed by soft scientists on controlling the messages emanating from the various media."

Against these problems the FOSN asserts that the problems of print scholarship are worse: "The most notable of these problems are that printed books and journals cost money, making them inaccessible to those who cannot pay, and that they are frozen on paper, making them immune to hyperlinks and invisible to search engines." In contrast the costs of publishing scholarship online "are so low that the required subsidy is trivial" because the literature is donated by authors who do not expect payment, and because publication is to the Internet, not to paper. If the required subsidy is small, Suber (2001) says, it can be made by universities, libraries, professional associations, foundations, endowments, authors or non-profit organizations.

Ultimately actions will determine if scholarly works are to be openly accessible. Improving access to scholarly papers has been the object of some recent initiatives. The Open Archives Initiative (OAi) was born at a meeting in October 1999 at Santa Fe. It was no coincidence that a town neighbouring Los Alamos National Laboratories, the original home of arXiv, was chosen because the aim was to work towards a 'universal service for author self-archived scholarly literature', in other words eprint archives. The emphasis was on standardizing the metadata describing the contents of archives, and the means of communicating the metadata, to enable independent services to collect data and present unified views of, prospectively, many different collections. Here was a bold idea, simple and focused,

that could have accelerated the open dissemination of full papers through eprint archives, but not all of its mentors were convinced. The OAi would promote only the open dissemination of metadata, which could describe any digital scholarly resource. Within a year the original emphasis on eprints had been diluted with the result that while innovative OAi services have emerged the volume of freely accessible full-text content has grown slowly.

The OAi could be viewed as the culmination of various proposals emphasizing the role of institutions taking control of the publications of researchers (Bachrach *et al*.1998), such as the National Electronic Article Repository (Shulenburger 1998) and Scholar's Forum (Buck *et al*. 1999), but none was to have the same incendiary impact as the proposal for a collective disciplinary archive to be maintained by one of the most respected science institutions in the USA. The National Institutes of Health (NIH) was to host an 'electronic publishing' site, E-biomed, that would provide free access to unedited eprints and refereed papers in the fields of biomedical research (Varmus 1999). The proposal generated lively debate, not unlike the responses to the earlier proposals, and as a result was reshaped, emerging as PubMed Central (PMC), a name chosen to make explicit connection with NIH's popular abstracting and indexing service, PubMed. Unrefereed eprints would not be allowed; instead journals would be coerced into depositing copies of published papers in the archive up to a year after initial publication.

The incendiary was not quite primed. PMC launched quietly in February 2000 and has grown steadily. Publishers disputed the need for a central archive. HighWire Press pointed out that it already hosted "the world's largest database of free life science articles and second in size only to NASA among free scientific article databases". By the end of 2001 the HighWire archive offered 330,000 articles from 100 scientific journals.

Then a strange thing happened, perhaps the initiative singularly most responsible for broadening the issue of open access to scholarly papers beyond activists in what Kling and McKim called the 'Electronic Publishing Reform Movement'. Some influential scientists posted a three-paragraph open letter on the Web, the aims of which informed the name of the initiative, the Public Library of Science (PLoS) (2000). The letter demanded the full contents of the published record of research and scholarly discourse in medicine and the life sciences be freely accessible, and pledged from September 2001 to boycott any journals that had not agreed to grant unrestricted free distribution rights within six months of their initial publication. The letter stopped short of requiring that PubMed Central be the exclusive repository of the free papers. Nearly 30 000 scientists signed the petition and galvanized a series of viewpoint papers and editorials in the popular science journals *Nature*, *Science* and *Scientific American*.

Some publishers have been shaken. The *Nature* debate on e-access demonstrated an unusual degree of bitterness, scaremongering, and even conspiracy theories (Frank 2001). At one point the President of Blackwell Science compared the proposals for public, electronic archives of the scientific literature by Roberts *et al*. (2001) with Lenin's revolutionary pamphlet of 1902: “I am not suggesting that Roberts *et al*. have written with the same manic violence but perhaps there is a hint in their approach of the same desire for centralized control by an elite (PubMed Central) and the one-track aggression so often linked to revolutionary zeal.” (Campbell 2001) Adopting a similar attack, the Executive Director of a publishing services company claimed: “The PLoS and PMC are increasing friction in scholarly publishing by taking a revolutionary and aggressive stance.” (Pentz 2001b)

Kaser (2001), Vice President Content, Information Today, Inc., warned that PLoS was risking journals slipping “into the ditch out of which no one climbs” and of ‘chucking’ a system “to achieve a dream that sounds so easily accomplished but, upon attempts to implement, could just as easily turn into a nightmare of unanticipated results”. This view of journals about to expire was surprisingly echoed by Tenopir and King (2001): “The journal system is at a critical stage – poor judgement could mean its deterioration or destruction.”

There was a hint of resignation from the Executive Director of the European Molecular Biology Organization (Gannon 2001): “If one believes that all publishers – commercial or non-profit – are the 'enemy', there is indeed no alternative (to PLoS).”

Biomedical scientists were not the first to call for open access to research papers. It was remarkable they should do so at all. Biomedical science is arguably the discipline best served by journals, accounting for over half of the most-cited journals (Garfield 1996). The scientists may regret framing the PLoS letter as an ultimatum to publishers, despite the publicity it generated, as it was always clear that both sides would be calling the other’s bluff. Neither wanted to destroy an otherwise beneficial relationship.

After the September 2001 deadline it was reported that the scientists' efforts had “fallen well short of initial objectives” (Russo 2001). Although ‘several prominent journals’ agreed to make published research freely available via PMC, the PLoS strategy has moved towards the creation of new journals instead of lobbying for change within existing journals.

The prospects for new journals are uncertain, but the impact of the PLoS across all of science cannot be underestimated. Despite having the best journals, biomedical scientists have understood that the

full benefits of electronic publication can only be achieved by allying journal publication with open access. They have recognized that open access will be the defining feature of network publishing, especially for scholarly papers, because open access means that a work is available free to *any* reader, and the benefits that follow are improved access to data – faster, everywhere – higher productivity, better research and, consequently, better journals.

## 2.11 Subverting the journals

Reflecting on the limitations of hypermedia systems, Halasz (1988) identified seven issues for developers. Reviewing these issues prior to the emergence of the Web, Halasz (1991) took a more considered and detailed view than follows, but in the Web we have a hypermedia system for dealing with 'rapidly changing information', while the multiplicity of Web search engines provides query-based access, another of Halasz's issues, to complement the Web's hypertext structure. Open hypermedia systems are a means of 'representing and dealing with nodes and links as unique entities separate from their components'. Sun's Java language is an example of a system that could support secure 'computation in hypermedia networks', and intelligent agents which can 'actively derive new information' from the Web are also emerging (O'Leary 1997). Commercial applications for 'collaborative work' are migrating to the Web, in particular within intranets. Apart from versioning, a long-term problem that involves understanding how to organise and maintain documents and all of the changes to documents, each of Halasz's seven issues has been advanced to a large degree.

It is over 50 years since Bush described the Memex, and if journals are to be electronic, distributed over networks, accessible, interconnected and published in an environment that is as open as possible, the framework is now in place.

One observer has predicted that the transition to electronic journals will be 'the most dramatic change in the working habits of scholars for centuries' (Naylor 1995). From the journals viewpoint, little seems to have changed. So far.

If journals are conservative, it is with the implicit support of the 'professoriate'. Journal authors, not readers, are the influential factor. Anticipating the opportunity for e-journals but dissatisfied with the speed of change, Harnad's 'subversive proposal' sought to change the system from within. The proposal is specific to 'esoteric' journals for which authors neither receive nor seek payment: "If every esoteric author in the world this very day established a globally accessible ftp archive for every piece of esoteric writing from this day forward, the long heralded transition from paper publication to purely electronic publication (of esoteric research) would follow suit almost immediately."

The proposal became the focal point for a spontaneous, electronically-mediated debate, and the resulting book (Okerson and O'Donnell 1995) contains all the reasons why the change to all-electronic journals is inevitable; as well as all the reasons why it is not. "The growth curve (for electronic scholarly journals) has continued upwards, but considerably flatter than that for Internet use among academics as a whole. If the model is so enticing, why the delay?" the editors ask. Authors are well served by journals: "it is not palpably difficult to 'get published'."

This then is the key to the all-electronic scholarly journal of the future. The subversion of today's paper journals is not principally economic, as some would argue, but cultural: journals are for authors who wish to *publish*, that is, gain recognition and credit for their work. In this respect journals publishing works well, so why change it? The needs of readers are less important, their influence in the 'market' for journals distanced by the pervasive role of libraries as purchasers.

Perceptions are changing, however. At its core Harnad's proposal reasserts the basic role publication plays for scholars, to communicate ideas. The 'spasm' experienced when the physics eprint archive was briefly switched off (Taubes 1993) demonstrates that the issue is not the willingness of physicists to submit to complex information systems, but their demand to be interconnected, to receive the most up-to-date results in their field, in short to read their subject voraciously. As soon as recognition that this demand can be satisfied spreads to other disciplines, paper journals publishing will collapse from its present scale, with only the most prestigious journals surviving in paper form.

'And publishing – ah, consider what publishing will become' (Nelson 1987). One does not have to understand the complexities of hypermedia systems to see the direction in which online publishing is heading: "The successful online publisher will most likely license access to other people's content to supplement or enhance his own, whether that content is online books, databases, bulletin boards, graphic image repositories, or online advice columns. What's "for sale" might be the interactive links, the thought structure the publisher puts around the distributed content area." (Fillmore 1993) Like the hypertext community before it, publishing must develop an 'open systems' approach. Fillmore continues: "A publisher's real job has to do with the effective generation and dissemination of ideas – not preserving outdated distribution mechanisms such as paper."

Email-mediated discussion lists, bulletin boards and other community-based information services were the earliest applications to succeed on the Internet. There was a time when publishers cited these as evidence of the unsuitability of the Internet for serious publishing, but others saw parallels with the emergence of print journals: "It is well to remember that the ancestor of today's scholarly

journal was the diary or logbook (the original ‘journal’) in which the scholar or scientist recorded data, thoughts, ideas, meetings, and conversations, much as do today's networked electronic lists.” (Okerson 1991) Today’s networked environment allows all this information to be stored and accessed, supplementing the formal, digested reports of research in the form of journal papers.

The challenge is to shape e-journals to fit the capabilities of the global network, to expand the accessibility of scholarly research information, rather than preserve the format of the printed journal. How this is happening at the journal level, and at the level of the wider network infrastructure that supports dissemination of information, is explored in the following two chapters.

# 3 Innovations in electronic journals

Electronic journal experiments began in the 1970s; real scholarly e-journals were being delivered 20 years later. The 1990s may prove to be unprecedented in the scale and speed of the transition from print to dual print and electronic journal publishing, a pace that may not last: “I’ll be surprised if any but a few pockets of the academic community show the same type of interest as they did for the first phase of (the digitization of) scholarly publishing.” (Peek 2001)

Growth in the number of peer-reviewed electronic journals has continued to accelerate as a series of surveys since 1994, since the emergence of the Web, has shown (Table 3.1). In 1996, when there were just over 100 such journals, it was predicted (Hitchcock *et al*.1996) that all major publishers would embark on a process of digitizing their journals, and that they would continue to use their existing production systems, simply creating exact electronic replicas of the printed journal papers. Both predictions have been borne out, but there are signs that a new phase is beginning in which electronic features are beginning to proliferate, endowing e-journals with a more distinctive purpose.

**Table 3.1 Evolution of a species: the growth of peer-reviewed e-journals 1991-**

| **Source** | **Data published** | **No. of peer-reviewed e-journals** | **Criteria** |
|---|---|---|---|
| *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* (1st edition)* | 1991 | 7 | |
| Clement | Oct. 94 | 25 (4 on Web) | Science, universal access, archived |
| Roes | 1995 | 39 (14) | Full-text |
| Hitchcock *et al.* | Jan. 96 | 115 (115) | STM, full-text |
| Harter and Kim | May 96 | 77 | Scholarly |
| Goldie | Dec. 96 | 508 | STM |
| Hitchcock *et al*. | Apr. 97 | c. 2600 (mid-98, projected) | 'UK' publishers |
| *Directory of Electronic Journals, Newsletters and Academic Discussion Lists* (7th edition)* | Dec. 1997 | 1049 | |
| Ulrich's International Periodicals Directory, 37th edition (Maclennan 1999) | 1999 | 10 332 | Periodicals, ‘exclusively online or in addition to a paper counterpart’ |
| *Directory of Scholarly Electronic Journals and Academic Discussion Lists* (1st edition)* | Nov. 2000 | 3900 | |

* see news release November 6, 2000, http://dsej.arl.org/dsej/2000/dsej1pr.html

This chapter seeks to establish an idea of what the true e-journal will deliver for the user. The emergence of e-journals during the 1990s is reviewed in two stages:

- to the middle of the decade, when individually produced e-journals were most common;
- from 1996 when mass digitization by publishers of established journals began.

In terms of emerging e-journal features, article linking and archiving multimedia are two technical capabilities highlighted by Peek. There is also anticipation that introducing structured markup in the production of journals, using XML for example, will improve the flexibility and usefulness of journals and papers. The biggest impact for the user, however, is potentially the separation of content from a single journal package to enable innovative and cost-effective services to be developed independently of the content production. Examples of journals demonstrating new approaches to 'packaging' that improve access to content, and which could consequently support the interaction of content and services, conclude the analysis of this chapter. The types of services envisaged are discussed in chapter 4.

## 3.1 Online journals 1990-95

Prior to the first wave of digitized journals from publishers there was a greater variety of online journals, exhibiting pre- and post-Web formats, reflecting different levels of financial support from self-publishing to professional publishing, and as a consequence demonstrating diversity in innovation, commitment and quality. Anticipating the avalanche of print replicas, a survey of online journals in science, technology and medicine (STM) from 1990-95 examined the publishing features of those early e-journals (Hitchcock *et al.* 1996) with a view to identifying the longer-term form of the online journal. During the period of the survey 115 journals were accessed through the Web.

From the first online journals the format of an e-journal has typically been a product of the subject of the journal, the time of its first appearance and whether it is an electronic-only journal or an electronic edition (copy of a print version).

By 1995 mathematics had generated the largest number of e-journals of any field. The first maths journals appeared online in 1993, and most were electronic-only. All but two of these journals were available in Postscript, a format designed for page printing but which can be produced from copy supplied in the formatting language TeX. Most of these journals also provided the option to view or download in TeX or the associated DVI ('device independent') formats.

Physics journals make similar demands on mathematical formatting capabilities, but owing to the dominance of the then Los Alamos eprint archives there was just one electronic-only journal and a number of electronic editions. Most electronic editions in this field were the half-hearted attempts of learned society publishers to compete with the eprint archives in terms of offering papers electronically, by reproducing letters and short communications rather than full papers from the print journals.

Medicine and biology were both well represented with online journals. The earliest journals in these areas appear to have begun in 1994. There was one major difference between the two fields. While a number of major biology titles were available online, none of the most prestigious medical journals were available in this form. In biology the number of electronic editions thus reduced the scope for new electronic-only journals, but in the medical area electronic-only journals were in the majority.

It would be hard to claim significant impact of online journals in any other field of science by 1995.

What features did these online journals display in 1995? Some were searchable. Some papers included links to other resources, although few digitized journals did this. Biology journals, especially those produced by dedicated online producers such as HighWire Press, had begun to make extensive use of links between references and a database of abstracts, notably to various Medline services, which in turn link to the already heavily interlinked and freely accessible DNA sequence and protein structure databases which underpin the discipline. Few other journals offered links to resources external to the journal. Colour graphics and photographs, not as costly to include in electronic journals as they are to print on paper, were rarely used in early e-journals, and video and audio, unsurprisingly, were even rarer. Electronic editions can present materials in advance of paper publication, and many did, although unedited preprints were not common. Readers of some online journals were alerted to new papers or issues via electronic mail, discussion lists or newsgroups.

One of the best examples of an e-journal of this time was *Astrophysical Journal Letters*, launched in July 1995. This presented the letters section in advance of publication of the full paper journal. The growing archive was searchable, and in HTML form the papers linked from thumbnail figures and photographs to full images. Threaded links were used to create themes and support subject categories, and external links pointed to a database of abstracts. Papers were also available in PDF form for local printing (Boyce *et al*. 1996).

For many electronic-only journals at the time the number of papers published in a whole year would not fill a single issue of a mainstream paper journal. These electronic-only journals faced a classic Catch-22: no electronic features, no papers; no papers, no features.

There have been some electronic-only successes. Some of the early online journals in mathematics have established a steady flow of papers, as has *Journal of Artificial Intelligence Research* from 1993 (Minton and Wellman 1999).

Reviewing progress of e-journals from this pre-1995 era, Crawford (2001) concludes that "early free electronic journals have done better than might have been expected". Beginning with the 104 titles listed in the 1995 edition of ARL's *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, Crawford found that nearly half (49) of those journals continue to publish a 'significant' flow of articles after at least six years, long enough to be considered 'lasting titles' having survived what he calls the 'arc of enthusiasm'. This has been achieved, Crawford notes, even though only a handful of these pioneering free e-journals appear in ISI's lists of 3,000 indexed serial titles – in 1999 ISI said it monitored 15 electronic-only journals (Kiernan 1999) – and most are ignored by libraries where attention is focused on aggregation of articles from commercial scholarly journals. Despite this, Fosmire and Yu (2000) found evidence that several free scholarly electronic journals have had a significant impact on their fields since the study by Harter and Kim (1996).

## 3.2 E-journals: progress in the digitisation era

By 1995 e-journals were on the agenda of most publishers. Many already had some involvement with e-journal projects but few had committed significant resources. With increasing visibility of free e-journals, and continuing progress of the physics eprint archives, pressure was mounting for publishers to counter with their own electronic products. Journal publishers and editors faced a difficult decision: to build new programmes of electronic-only journals, or copy paper editions. The ACM electronic publishing plan went as far as to consider eliminating paper from the process entirely (Denning and Rous 1995), but no other major publishers openly contemplated this option at the time. Paper journals continue to attract the majority of papers and the best papers (Harnad 1995a, Anderson *et al*. 2001), but an e-journal predicated on the paper journal model cannot have real interactivity and multimedia content, for example, without diverging from the paper production process and content.

Not surprisingly, publishers chose to model e-journals on paper. Since 1995 the mass of printed journals have conceived an electronic copy. Effectively, every journal now produces current issues

electronically. The significance of e-journal copies on this scale, however, is that it unites all those who have an interest in journals of any format in deciding the future for this type of publication.

One catalyst for this journals digitization programme was a pilot site licence scheme set up by the Higher Education Funding Councils (HEFCs) for the UK (Bekhradnia 1995). The idea of a site licence emerged at that time as an agreement between publishers and libraries designating the machines across the library 'site' that were authorized to view e-journals subscribed to by the library. With print journals there is a natural limit to circulation. Libraries might purchase multiple copies of highly-used journals. With no such limitation for e-journals, site licences were structured to provide concurrent access for users while compensating publishers for possible loss of multiple subscriptions.

For the period of the pilot scheme the publishers were guaranteed a portion of their existing subscription income directly from the councils; in return the publishers agreed to deliver both paper and electronic editions to subscribing libraries. Coincident with the resolution of the subscription issue was the publishing industry's apparent convergence on Adobe Acrobat (Smith *et al.* 1993), or Acrobat-like page image applications such as RealPage developed by journal producer and distributor Catchword, as the production format of choice (Table 3.2). Even the growth of HTML could not detract from overall use of Acrobat. As pioneered by HighWire Press, HTML and PDF have been widely adopted as dual e-journal delivery formats for on-screen reading and for local printing.

In other words, through site licences and by using PDF page images, publishers found ways to increase their income sufficiently to cover the marginal additional cost of generating an

**Table 3.2 Impact of PDF as a journal format: range and popularity of formats used by 'UK' publishers to produce online journals in 1997 (Hitchcock *et al.* 1997a)**

| **Format** | **No. publishers using (announced)** | **No. journals using (announced)** |
|---|---|---|
| PDF | 13 | c. 630 |
| RealPage | 10 | 111 |
| Full SGML (HTML 'on fly') | 4 (3) | 58 (c. 1400) |
| HTML | 3 | 50 |
| SGML headers | 2 | 265 |
| Other (Postscript, Printerleaf) | 3 | 39 |

electronically-deliverable format from the original journal pages, at the same time as the underlying technology required to implement the process and deliver the pages had begun to reach maturity both technically and in terms of market acceptance. As a first step this approach appeared to be justifiable from the viewpoints of publishers and users, who benefitted from the flexibility of electronic access and the ease of printing selected pages, although an early user study of digitized e-journal services revealed that people prefer not to read at length on screen, and printing out is slow (Woodward *et al*. 1998). The report blamed usability problems on publishers for allowing technology, rather than human factors, to drive their efforts.

Digitisation is responsible for the dramatic rise in number of scholarly e-journals from a few hundred in 1996, to the 4000 or so e-journals evaluated by the first edition of the *Directory of Scholarly Electronic Journals* in 2000, and the 10000 titles with electronic copies recorded in the 1999 edition of *Ulrich's International Periodicals Directory*, presumably as claimed by publishers (Table 3.1).

Digitisation had a less obvious effect on extending the range of features supported by e-journals. The page-based technology and, more fundamentally, the underlying publishing philosophy that embraces it, are inadequate to fulfill the demand for all-electronic journals. To exploit electronic capabilities fully, digitised e-journals must eventually diverge from their paper counterparts. Elsevier is reported to have acknowledged that “stand-alone on-line journals that are the same as the hard copy without added features such as links to supplementary data have not had a great impact”, but “electronic-only journals, such as *New Astronomy*, which offer entirely new ways of presenting information and do not mimic hard copy, have had a huge impact and have attracted many authors and readers.” (Wilkinson 1998)

## 3.3 Progress in e-journal features: three snapshots spanning a decade

Better, faster, cheaper: these were the watchwords of the emerging e-commerce industry in the late 1990s, and these seem reasonable criteria on which to assess the progress of e-journals. In addition to the surveys cited above, we are afforded three snapshots in time of selected e-journals over the last decade – two of these are special journal issues presenting collected papers about new e-journals, and one is a recent survey paper:

- *Public-Access Computer Systems (PACS) Review*, special section on Network-Based Electronic Serials, 1991
- *Journal of Electronic Publishing*, special issue on Electronic Journals: Why? 1997
- McKiernan (2002)

The most obvious difference between journals described in 1991 and 1997 was the delivery format, which changed from email-based plain text circulated via the Bitnet service, to HTML on the Web. In most cases the journals, when new, could be described as exploratory if not experimental. Tables 3.3 and 3.4 compare the characteristics and performance of the journals covered in the respective issues.

It is remarkable that of the journals in Table 3.3, which began with relatively humble features and span 15 years since the first was launched, all but two continue to publish new issues (and another struggles on after a two-year termination). This is more surprising because only two of these journals could be said to have had a clear vision for what an electronic publication could offer that was distinctive from print: *Psycoloquy* and *Postmodern Culture* both sought to change the form of the writing process and the speed of interaction with users. None of the journals cited lower costs compared with paper as a primary motivation, although it is unlikely many would have been viable as print journals, as the *Newsletter on Serials Pricing Issues* discovered after it discontinued its original dual print-electronic format.

The growth of the Web and wide use of HTML was clearly a significant new factor in journal designs by 1997. Discounting *PACS Review*, which completes the circle between the two issues and belongs in the earlier era, by 1997 novel e-journals were seeking to exploit the facilities of Web formats for enhanced graphics, animation, sound and vision, hypertext links, also the inclusion of datasets and code simulations. Note that since this is a snapshot in time, the motivations highlighted for the 1997 journals are taken from the referenced papers rather than from the journal sites today (Table 3.4).

Only two of the journals highlighted lower costs as an attraction of the electronic format: *TC* and, intriguingly since it charges a subscription, the ACM journal. Another, the potentially feature-rich *Earth Interactions*, was clear that its costs would *not* be less than print. *First Monday* cited rapid review and publication as an advantage of the format, as did the ACM and RSNA journals.

With the exceptions of *RSNA EJ*, which has merged with another journal, and *PACS Review*, all the 1997 journals remain active today, fulfilling Crawford's arc, and suggesting the different visions had some foundation, as did their counterparts of 1992. This might be because all performed the traditional journal function of publishing original peer reviewed papers, rather than because of any novel electronic features.

**Table 3.3 E-journals covered in *Public-Access Computer Systems Review*, Vol. 2, No. 1 (1991), http://info.lib.uh.edu/pr/backvol.htm#v2n1 (data checked Dec. 2001)**

| Title | Motivations – why electronic? | First published | Latest issue | Refereed | Free | Other formats* | Reference |
|---|---|---|---|---|---|---|---|
| *New Horizons in Adult Education* | Part of the Syracuse University Kellogg Project which had "a mission to provide broader access to the university's adult education materials and to facilitate the exchange of information and learning using the very latest technologies where possible … connecting adult educators worldwide; generate dialogue" | Fall 1987 | V15i2 Summer 2001 | Y | Y | | Hugo and Newell |
| *Psycoloquy* | "explicitly devoted to scholarly skywriting, in which authors post a brief report of current ideas and findings on which they wish to elicit feedback from fellow specialists as well as experts from related disciplines the world over" | 1989 (from 1985 as an e-bulletin board called the *BITNET Psychology Newsletter)* | V12 2001 | Y | Y | | Harnad |
| *Newsletter on Serials Pricing Issues* | "to lead the fight against high journal prices", to show librarians responding to the emergence of electronic publishing as a viable alternative to paper journals | Feb. 1989 | i257 Aug. 2001 (since recognized as the final issue) | N | Y | Paper to i13 | Tuttle |
| *Postmodern Culture* | "can serve to encourage more experimental scholarly writing" such as works-in-progress, a video script | 1990 | V12i1 Sep. 2001 | Y | | Disc, microfiche | Amiran and Unsworth |
| *Journal of the International Academy of Hospitality Research* | "The Scholarly Communications Project of Virginia Tech agreed to publish JIAHR as a pioneer effort to explore--in a very practical way--the frontier of electronic communication of scholarly information" | Nov. 1990 | i12 1998 | Y | N | | Savage |
| *EJournal* | "One fine day, as narrators blithely say, I wondered if it would make sense to try distributing some sort of "journal" over the network." | Apr. 1991 | Mar. 2001 (revived after valedictory issue 1999) | Y | Y | | Jennings |

* Other than ascii, HTML

*New Horizons in Adult Education* http://www.nova.edu/~aed/newhorizons.html
*Psycoloquy* http://cogprints.ecs.soton.ac.uk/cgi/psyc/newpsy
*Newsletter on Serials Pricing Issues* http://www-mathdoc.ujf-grenoble.fr/NSPI/NSPIe.html
*Postmodern Culture* http://www.iath.virginia.edu/pmc/contents.all.html
*Journal of the International Academy of Hospitality Research* http://scholar.lib.vt.edu/ejournals/JIAHR/
*Ejournal* http://www.ucalgary.ca/ejournal/

**Table 3.4 E-journals covered in *Journal of Electronic Publishing*, Vol. 3, No. 1 (Sept. 1997), http://www.press.umich.edu/jep/03-01/index.HTML (data checked Dec. 2001)**

| Title | Motivations – why electronic? Features/problems | First published | Latest issue | Ref. | Free | Other formats* | Reference |
|---|---|---|---|---|---|---|---|
| *PACS Review* | Wary of graphics and length of papers. Few links | 1990 | V9i1 1998 | Y | Y | Ascii text, V1-5 print | Ensor and Wilson |
| *ACM Journal of Experimental Algorithmics* | "publish data, programs, animations, and multimedia … search … inexpensive … should have a shorter turnaround time as well as more flexibility … a library of software, test data, test generators, and past results" | 1995 | V5 2000 (accepted articles to appear for 2001) | Y | N | Postscript, LaTeX | Moret |
| *TC: A Journal Of Biblical Textual Criticism* | Lower costs for e-journal than print; increased access for users; shorter production cycle. Features include links, and simulating the process of scribal copying. Problems displaying multilingual text, lack of authors | May 1996 (Nov. 1995 email discussion list began) | V6 2001 | Y | Y | | Adair |
| *First Monday* | Rapid review and publication | May 1996 | V6i12 2001 | Y | Y | | Valauskas |
| *Radiological Society of North America EJ* | Java, animation, sound, moving images, links, interactivity. Faster dissemination, enhanced discourse, improved access. Lack of submitted papers | 1997 | V3 1999 (merged with *RadioGraphics*) | Y | Y | | Ackerman and Simonaitis |
| *Earth Interactions* | Publishers "force me to reduce the information in my scientific papers so that they will fit on a flat printed page." Graphics (animation 3D), data and code segments. 'Forward references'. Not 'less expensive' than print | 1997 | 2001 | Y | N | SGML | Holoviak and Seitter |
| *Living Reviews in Relativity* | Citation analysis and reference links, javascript, article updates. "We cannot look at "publications" as isolated pieces … Living Reviews is a key part of a complex information system" | 1998 | V4 2001 | Y | Y | | Wheary *et al.* (1998) |

* Other than HTML

*PACS Review* http://info.lib.uh.edu/pr/pacsrev.html
*ACM Journal of Experimental Algorithmics* http://www.jea.acm.org/
*TC* http://rosetta.atla-certr.org/TC/TC.html
*First Monday* http://www.firstmonday.dk/
*Radiological Society of North America EJ* http://ej.rsna.org/
*Earth Interactions* http://earthinteractions.org/
*Living Reviews in Relativity* http://www.livingreviews.org/

One of the difficulties of assessing journals based on papers by their editors is how much is real and how much aspirational, especially after a short publication period. Perhaps some of the presumed feature-rich journals of 1997 were the exception – Singh *et al*. (1998) were unable to trace any interactive multimedia journals, and more recently Burg *et al*. (2001) similarly found it "difficult to find examples of the type of (multimedia interactivity) journal predicted since the early 1990's", although they did highlight *RSNA EJ*: "Among the journals we found, two with a significant multimedia interactivity were the *Internet Journal of Chemistry* and the *Radiology Society of North America Electronic Journal*. *IJC* lists images, interactive chemical structures, interactive graphs, movies, etc. ... *RSNA EJ* contains a number of animated radiology images using video, VRML, or Java-based simulations". In addition Burg *et al*. noted a music journal, *Ethnomusicology Online* and "in the creative vein, some journals exploring interactive narrative and art have sprung up, such as *Cultronix* at Carnegie Mellon or *Labyrinth* at the University of Southern California".

Reality and aspiration can also be hard to discern in McKiernan's (2002) description of e-journal innovations. Section titles such as 'empowering', 'entwined', 'explorative', 'expressive' and 'experience' frame an imaginative structure, allowing the article to be feature- rather than journal-centric. Compared with the 1997 journals, the more innovative features tend to be user-centric, and a number of new models that extend beyond the single journal model are highlighted.

Among the notable features of the latest generation of e-journals are visual indexing and search services, and language translation facilities. McKiernan selects pre-eminent examples of journals displaying video and interactive graphics. Users of some services can create personal 'filing' cabinets to save annotated copies of selected papers, and can save bibliographic records for use in local reference management tools.

Discussion forums are notoriously difficult to activate within journals, and McKiernan does not enlighten on how active the cited forums are, although some journals such as the *British Medical Journal* (BMJ) and the *Journal of Interactive Media in Education* (JIME) have succeeded in motivating rapid responses to published papers.

McKiernan notes that most of the e-journal features he describes were anticipated by Lancaster (1985) a generation earlier, but notes that linked access would need to be added to Lancaster's continuum of e-journal development. Publisher-centred reference linking services, such as the Institute of Physics's Hypercite and medical journals linking to the National Institutes of Health's PubMed, have been available for some years, and will soon be widely supplemented by cross-publisher services such as CrossRef and library-mediated linking services, as discussed in chapter 4.

McKiernan's analysis is not quantitative and it can be assumed that the features described are exemplary rather than typical, but individual e-journals have exhibited continuing progression in terms of exploiting new possibilities that are not available to the print journals.

Most authors writing in 1992 or 1997 believed their journals would be better as electronic products; but only a few thought they would be faster. By default rather than design, it might be suspected, almost all are cheaper than any print rivals, as they are mostly free journals.

## 3.4 'Free' e-journals

Mass digitisation of journals may have slowed the growth of free e-journals, but it did not stop them entirely. Wells (1999) found 387 'free' e-journals, defined as independent, electronic scholarly journals, that is, those that are available through the Internet, usually the World Wide Web, free of charge to the reader, and publish academic articles, usually peer reviewed. Wells was sanguine about the future for free e-journals, concluding that unless the current market for electronic journals changed in some way, either by journal budgets devolving to academic departments, or some collaborative action between universities, that the future for electronic journals was as part of "one-stop shops" for particular subjects, with a mixture of free and paid-for journals (or article servers), together with other related services, controlled by commercial outfits.

Using analyses of citation impact, Fosmire and Yu (2000) showed that the prospects for 'free' e-journals are encouraging, providing tangible evidence that many are making an impact, a few spectacularly so. The authors admit that the data sources used, ISI publications *Web of Science* and *Journal Citation Reports*, may have the effect of underestimating impact in these cases, although given that many of the journals studied are small, they also concede the data might contain statistical quirks that overstate impact.

One of the motivations for the study was the claim that, despite their increasing numbers, free electronic journals did not appear to be receiving the attention they deserved from the library community (measured in terms of inclusion in catalogues and on library Web sites), commensurate with their inclusion in major abstracting and indexing services (Fosmire and Young 2000).

In both cases the definition of 'free' is important. All journals accessed through library services appear to be free to the user, but this imposes a restriction on the range of available titles that can only be masked by the efficiency of individual libraries. Trial periods where access is free are also

discounted. More significantly for advocates of open access are the growing number of high-impact biomedical journals that, prompted by the PubMed Central initiative, make their electronic editions free some time, typically six months to two years, after initial publication. Free e-journals of this type can be found from Kamps' Free Medical Journals (http://www.freemedicaljournals.com/); a similar service highlights free access e-journals recognized by the European Physical Society (http://www.eps.org/PhysNet/journals.html), and there are lists of free journals of varying degrees of rigour for chemistry, mathematics and other subjects. As Kamps says: “Pizza, red wine and cordon bleu all have one thing in common: they do not go through telephone lines. That's the reason we will continue to pay for them. MP3, PDF and HTML do go through telephone lines, and that's why we will probably not pay for them in the future.”

Further confusing the definition of free e-journals is free availability in selected regions, as sponsored by the Developing Nations Initiatives. In 2001, a number of organizations began to develop and publicize programs to bring peer-reviewed science journals for free, or very cheaply, to developing nations. Such programmes are identified at http://www.library.yale.edu/~llicense/develop.shtml .

## 3.5 Overlays, open access and omnipotence

“We cannot look at ‘publications’ as isolated pieces … *Living Reviews* is a key part of a complex information system.” Of the e-journal developers cited above, only Wheary *et al*. (1998), perhaps with the benefit of hindsight and experience afforded by a second paper on their journal, looked beyond the journal and how it might work within the larger information environment. Their means of achieving this is through reference linking.

Reference linking and citation analysis is one of the primary ways in which e-journals can extend into the wider literature. ‘Virtual’ e-journals are other examples, although typically these offer users limited selections within publisher catalogues, often focused on a topic, such as the *Virtual Journal of Nanoscale Science and Technology*, a “weekly multijournal compilation of the latest research on nanoscale systems”.

It is no coincidence that these examples cover the physical sciences, because it is in this area that the scale of the physics eprint archives has enabled new ways of managing collections of papers, thus reshaping the production and business model for journals.

A mechanism for basing journals on arXiv is referred to as an ‘overlay’. Ginsparg (1996) advanced the idea for intellectual overlays to implement the filtering role provided by the journal system,

arguing that a unified global collection would dramatically improve the 'artificially partitioned database of the paper journal system'. At that time *Astrophysical Journal Letters* and *Physical Review D* (PRD) used arXiv to create a preprint overlay, linking to papers accepted for the journal but not yet published.

Ginsparg noted the extent to which publishers had defined themselves in terms of production and distribution, "roles which we now regard as largely automated", a charge subsequently accepted by the American Physical Society (Smith 2000): "*PRD* was available from July 1996, and the only links out from the first articles were either to other *PRD* articles, or to the online SPIRES or LANL preprint databases. The accessibility of the preprints in the online world, at least at that point, surpassed the accessibility of traditional published materials, overturning the old hierarchy that associated 'publication', particularly in peer-reviewed journals, with accessibility."

At the time of 'publication' in *PRD* the preprint link to arXiv disappears. The more radical step, Smith commented, is dispensing with journal-specific copy editing, formatting, layout and the distribution of final content. This step has been taken by one startup journal, *Advances in Theoretical and Mathematical Physics* (ATMP). The journal maintains only links to the archive thereby realizing, it claims, the first e-journal as an overlay to the global eprint archives, and therefore the first journal to be modelled on the eprint culture.

Thus, three forms of overlay are apparent: through citations, links to preprints, and linking to the archive from the journal's Web site to provide access to the final versions of papers.

Launched in July 1997, just a few months before *ATMP*, *Journal of High Energy Physics* (JHEP) is another electronic-only journal designed to exploit first submission to arXiv, but appears to stop short of a full version overlay, instead relying on "complete automation of the editorial work that is carried out by means of a software robot" to minimize costs and ensure free and open access to refereed papers" (JHEP Executive Office 2001). In both journals the emphasis is on peer review with editorial processing significantly scaled-down.

In case it might be expected that such radical new models might struggle to attract authors, according to Fosmire and Yu (2000), *ATMP* is a bona fide high impact journal: "It boasts an article with over 800 citations and another with over 500 citations. Certainly, high-energy and particle physics theory papers have high citation numbers within the field of physics, so *ATMP* and the *Journal of High Energy Physics* (JHEP) would be expected to have high impact factors (and *JHEP* ranks in the top 10%). However, *ATMP* excels even among high energy physics titles, as it had articles that were

cited 2nd, 3rd, and 30th most often in 1999, among all citations in the HEP-SPIRES database to articles from any year".

These journals are implementing the mission of true electronic scholarly journals, no longer isolated entities but part of a wider information environment and embodying O'Reilly's vision for creating 'information interfaces'. There can be no doubt that the critical factor is an underlying layer of open access data, in this case the physics arXiv. In exploiting that resource *ATMP* and *JHEP* have tailored their services precisely and cost-effectively and in doing so can present journals that are also openly accessible to all users. Not all services can or will be free, but the model adopted by these journals puts back into the field as much it gains from the facility of the original open access layer.

Recognition that the packaging¸ if not the role, of journals is set to become more amorphous may seem limited, but there are other factors at work. A new infrastructure for networked scholarly communication is emerging, and this is the focus of the next chapter.

# 4 The emerging scholarly communication infrastructure

While the detail of the future infrastructure of electronic publishing may remain unclear, one statement can be made with some certainty:

- Scholarly electronic information will be ‘seamless’ and ‘integrated’

This must be true because a search of the model built for this project reveals that one or both of these words, untypical in common speech, is used in almost 15% (85 papers from almost 600) of a selection of papers on scholarly electronic publishing (and this is without searching the full texts of these papers).

A search of Google (in November 2001) for a typical phrase such as “seamless integration of information” produced nearly 500 results, predominantly companies offering network and inter-application software. A small modification to “seamless access to information” gives almost 1000 sites, with portals and gateways to the fore, which is closer to the topic of interest. If instead the term “seamless linking” is used 450 sites are listed, leading with journal publishers and databases, and this begins to reveal the meaning of the terms in the context of this work.

Although these words and phrases have become clichés, in essence what they mean is that from any given document the user might expect to be able to retrieve any *related* document within one mouse click. Typically what is related is defined, and linked, by the author or publisher or other service provider, and is limited by the tools and information services at their disposal. Longer term the relation may be *anything the user might consider to be related*. While this is ambitious, it is what intelligent agents and the semantic Web project are ultimately intended to achieve.

For scholarly research information, particularly the published journal papers that Harnad defines as esoteric or give-away works, by authors who have no intention of receiving direct payment for publication of the work they produce, this raises two subsidiary questions about the ‘seamlessly integrated’ literature:

- Will it be complete (from the viewpoint of *every* user)?
- Will it be free (or *appear* to be free)?

A work may appear to be free to the user when it is accessed via a library, for example. Completeness was the classical goal of the top research libraries, where users would have access to, if not absolutely everything, at least everything they were likely to need. But with the growth in the output of published research and the 'serials crisis', this no longer seems to apply. The condition will need to be replicated, everywhere, if seamless integration, even on a modest scale, is to be achieved.

This chapter explores some of the initiatives being taken by various interested groups – libraries, publishers and other special interest groups – to build a framework that will support 'seamless integration' for users of scholarly journal papers. It challenges assumptions evident in many discussions on the system of scholarly communication that show, as Lynch (1999) puts it, "We've chosen to emphasize extrapolation rather than to identify and understand emerging discontinuity", and reveals the growing influence of two key factors, access and interfaces, which are subtly changing the form of scholarly journal publishing.

## 4.1 Progress in libraries: preservation strategies

Any analysis of the changing infrastructure of scholarly information should begin where, traditionally, researchers have gone to find information and from which, it should be assumed, they continue to access most professional resources such as journals and books (Tenopir and King 2000). That is the library, not a digital library but the physical buildings that still form the academic centre of most educational establishments. For many the current experience of a digital library is an extension of the physical model. Even if they don't have to visit the actual library building, access to services is controlled and paid-for by the same administration, and the resources are mostly digital replicas of those that are familiar in physical form. True digital libraries are more likely to be based on large-scale initiatives, such as the Distributed National Electronic Resource (DNER) in the UK, which are beginning to provide common cross-library services and networked information, but for now the single university library remains the foundation for most library activities.

Confronted with the threat of disintermediation – where users bypass regular agents and access materials directly, an obvious and early manifestation of user behaviour once the Web became widely used (Wiederhold 1995) – and the opportunity for publishers to exploit this (Odlyzko 1999), libraries appeared to be cautious during the early 1990s to the possibilities of electronic services. Describing the problems libraries are having grappling with the issue of e-books, Lynch (2001a) identifies the reason for this: "libraries are, I believe, confused about what they want, particularly in terms of business models".

For journals an emphasis on business models has meant an obsession with the serials crisis, and it can be seen in retrospect just how much this has dominated library developments in the provision of electronic services, which after the initial apprehension can now be seen as substantial, as shown below, even though the business model – the library pays suppliers on behalf of users – can be said to have changed relatively little. This has to be contrasted with many researchers and developers whose experiments with electronic information services, some described below, have *not* been designed to reduce the serials crisis but to find ways of improving the utility and effectiveness of information.

Thus, early free e-journals and eprint archives were welcomed as potential solutions to the serials crisis rather than as radical new publishing models. Libraries are naturally sceptical of the longevity and stability of these services, and exercise careful selection policies (Weintraub 1998). As a result, free e-journals and archives are not prominently presented as part of libraries' core services. Should they be? Libraries have never presented resources to users on the basis of cost (although cost often acts as a rationing factor, especially for print copies). Yet it is not hard for the casual browser to notice expensively acquired journal gateway services and well-known database services from typical library Web pages; in some cases only the subject-directed user will discover arXiv, a resource of equivalent magnitude to many of the highlighted services.

This might seem reasonable Web site strategy based on anticipated user behaviour, but it is an example, possibly unintended, of the power of the user interface, which we will find becoming more important as this chapter progresses. The effect is that most users remain unaware of the issues, specifically journal pricing, that may be restricting the resources they can access, and unaware of the alternatives. The US Association of Research Libraries (ARL) views this sufficiently seriously that it urges its 123 member libraries to take an active role in educating their patrons about the 'issues of scholarly communication', meaning journal pricing and the serials crisis, anticipating use of substantive library Web pages as an obvious means of doing so. A survey discovered "only 4% of ARL libraries place a high enough priority on these issues to warrant homepage presence." (Lugar and Thomes 2001)

By effectively placing free resources in obscurity, libraries are perhaps reflecting a subconscious agenda, unappreciated even by ARL. Superficially, resources that are free to the user might appear to be the ideal solution to the library serials crisis, but as an organization focusing on its own business model and role as budget centre and power broker in the information chain, the library does not need free sources – remember disintermediation – that will give others the ammunition to attack its budget, but better value sources that enable it to provide users with more pages per unit of that budget.

As a result, despite heralding the movement to 'free the refereed literature', the library community has done little in real terms to promote reallocation of funds as encouraged by that movement to support the new scholarly communication models. Since this would involve reallocating money away from the library, this is hardly surprising for an organization that is naturally as committed to its own preservation as that of the materials it holds.

### 4.1.1 Progress in libraries: dealing with publishers

Casting aside early indecision, libraries have participated in important developments that are contributing to the shaping of the emerging infrastructure for accessing electronic information services. Defying logical expectations, instead of turning to new models to solve the serials crisis, libraries have been lured into new deals with the *bete noires* of their journal budgets, the publishers. As a result the library community has become involved in the some major initiatives:

- Site licenses for electronic journals, and more aggregated content from database services.
- Support for the Scholarly Publishing & Academic Resources Coalition (SPARC), a bold and possibly misguided move to increase competition in the journal market by facilitating partnerships with publishers and other journal producers.
- Fast-track standardization of OpenURL, to link users to these subscription and document services, recognising this vast new array of electronic content would need to be accessible and navigable by users *within the library's* information environment.

The economics of print journals determined that copies were purchased on a per-title basis. Since for electronic products the marginal cost of distributing additional copies of a given title is effectively zero, publishers were able to respond to libraries' concerns over high journal prices by offering deals on 'bundled' collections of e-journals, which might include the publisher's complete catalogue of journals, or a subset. In most cases the bundle would include those primary journals to which the library already subscribed in print, with additional journals thrown in for little extra money. In this way publishers got some more money for access to journals that it cost them almost nothing to serve and that otherwise the library would not have been likely to subscribe to, and libraries satisfied their objective of better value for money in terms of cost per page delivered to users.

This is the principle of site licenses, where the 'site' from which users can access content as part of the deal could be an institution, a state-wide group of institutions, a national collective, such as in Canada, or even all the people of a nation, as in Iceland (Elliott 2001). One of the first site license schemes was the Pilot Site Licence Initiative in the UK, superseded by the National Electronic Site

Licence Initiative (NESLI), neither of which support national ‘sites’ but broker deals between publishers and the individual participating institutions across the UK.

Possibly the most ground-breaking instance of a site license has been OhioLINK, which launched an Electronic Journal Center in 1998 serving a consortium of almost 80 college and university libraries across the state of Ohio. Every user at every participating establishment, from the largest university to the smallest college, has access to the content for which a deal has been secured with a publisher.

OhioLINK has been one of the most outspoken advocates of site licenses, claiming that its approach overcomes “the entrenched and limiting economic practices of vendors to individual institutions, and the library-imposed, self-limiting, collection development mentality of information rationing that pervades our community. By radically changing the value equation of information delivered per dollar spent, consortia can set the evolution of our industry on a new and better, long-term course.” (Sanville 2001) ‘Enablers rather than gatekeepers’ is the theme of the initiative, and Sanville presents compelling evidence that cost per page delivered and value for money have increased notably. Yet Sanville tacitly acknowledges that what these impressive data mark is just the first stage in the transformation of access to journals. If, as stated, the current deal is a means to measure and evaluate what is used and what is not, rather than a concession to publisher demands, then it is likely OhioLINK will have to tackle the traditional gatekeeper role again at some time in the future.

For an arrangement that some publisher-library partners have unreservedly referred to as ‘win-win’, site licenses, or the ‘Big Deal’ as detractors refer to it, have proved remarkably controversial. Principal among the latter is Frazier (2001), who argues that such deals will ultimately vest more pricing power with publishers, remove libraries’ discretion over selection of titles and thereby eliminate a control on the market for journals: “There's no question that the Big Deal offers desirable short-term benefits, including expanded information access for the library's licensed users. In the longer run, these contracts will weaken the power of librarians and consumers to influence scholarly communication systems in the future. Librarians will lose the opportunity to shape the content or quality of journal literature through the selection process. Those who follow us will face the all-or-nothing choice of paying whatever publishers want or giving up an indispensable resource”, raising in a different form the spectre of disintermediation.

How can libraries assess the long-term risks against short-term benefits? By looking at the past, which is how Guedon (2001) brilliantly exposes the outcome of site licenses and some publishers’ Machiavellian instincts in promoting them. Nevertheless, the signs are that short-term interests are winning (Nabe 2001).

### 4.1.2 Progress in libraries: promoting competition

Against the background of increasing concentration of titles among fewer publishers (McCabe 1999), libraries have supported at least one major action to introduce more competition into the journal market. SPARC emerged from a Roundtable in 1998 convened by an alliance of scholarly organisations in the USA, including the ARL, as a response to high journal prices. What is supposed to differentiate SPARC is its commitment to lower journal prices, and it can do this because member libraries in the alliance agree in advance to buy its journals.

“Becoming a SPARC library requires a $5,000 annual fee and a pledge to spend a minimum of $7,500 each year on SPARC publications. Since the average ARL member's annual serials budget is $3.6 million (out of a total $5.5 million materials budget), that commitment is not onerous.” (Rambler 1999) At the time, Rambler noted, 100 ARL members – plus 14 non-ARL members – had made the decision to support this venture.

Typically SPARC supports journals that compete directly with ‘high-priced’ journals from other publishers (SPARC Alternatives programme), with at least one notable success. *Organic Letters*, first published in July 1999 by the American Chemical Society and promoted by SPARC has, as measured by impact, surpassed the established journal in the field, *Tetrahedon Letters* from Elsevier, it is reported. This confirms the emergence of SPARC as an effective publishing support organisation, if not necessarily a radical one.

In addition, SPARC sponsors original electronic-only journals (SPARC Leading Edge programme) and collective information services and portals (SPARC Scientific Communities programme)

The emphasis of the support community on pricing, to the exclusion of almost all else, is puzzling. Pricing is an artifact of current market conditions, and a commitment to low pricing is not a guaranteed long term-strategy unless it is underpinned by strong controls. What is ‘low price’? For publishers now demonized for high prices it has not always been so, and price inflation can take hold rapidly if market conditions permit. For SPARC, greater protection for the venture’s core strategy would be assured by allowing authors stronger rights over their published papers. This aspect has not so far been examined in papers on SPARC, which have instead adopted a largely promotional and uncritical tone. As an organization at least SPARC is effectively immune from commercial takeover because of its status as an alliance and facilitator rather than as a publisher.

### 4.1.3 Progress in libraries: making appropriate connections

One of the effects of the growth of published scholarly information (Odlyzko 1994) is that libraries began using a variety of agents and services to acquire materials, simplify administration and reduce costs. One of the outcomes of the serials crisis is that fewer, non-core journals are subscribed to and libraries resorted to just-in-time document delivery and collections from licensed full-text aggregators such as Ebsco, UMI and others (Inger 2001), as ways of providing users with additional full-text sources. Secondary services used to identify relevant resources include abstracts and indexes, and other databases. Thus, users might discover documents in legacy print or electronic sources via the library's online catalogue, from locally held CDs or network services.

Using one of these services it is becoming increasingly likely that the reader's reference will offer a link to the full text of the required paper. The link might be provided by the service that delivers the reference, possibly in conjunction with another agent. An example of an agent providing journal reference linking data is CrossRef, a cross-publisher organization that collects metadata describing journal papers and their locations on the servers of participating publishers (Pentz 2001a). Any service that delivers references can subscribe to CrossRef and use its data to direct links at recognized papers on these servers.

Given its foundation among publishers, with their interests in controlling distribution and managing rights for digital content, CrossRef links use Digital Object Identifiers (DOIs), which are tagged to article metadata supplied by the participating publishers. A DOI identifies and locates an object on the Internet, but the identifier is opaque so neither property can be deduced by inspection from the DOI and a resolution system is required (Davidson and Douglas 1998). The Handle System (Arms *et al*. 1997) is used to resolve DOIs. Any service can base its document identifiers on Handles; currently only four services, including CrossRef, are authorized to act as registration agencies, and therefore to manage resolvers, for DOIs.

Libraries, with their complex array of subscriptions to full-text sources, may not have paid for direct access to a publisher's service, and using a native CrossRef link could incur charges to view the referenced document. Yet the user may have authority to access the same paper free of charge via another library service. This has become know as the 'appropriate copy' problem (Caplan and Flecker 1999), the appropriate copy being that to which existing library subscriptions or other resources allow access.

A software solution to the appropriate copy problem was described by Van de Sompel and Hochstenbach (1999): “The goal is to present information to the user in the context of the entire collection that is available”. Designed as a ‘generic linking system’ and now marketed commercially as SFX, this established the ideas of context-sensitive linking, where links depend on the available collection, and of extended service-links, which enable multiple cooperating services to act on a link request. Subsequently this led to a generalized framework for communicating and resolving reference links via SFX-like services, OpenURL, described as an ‘interoperability specification’, which provides an interface between (Van de Sompel and Beit-Arie 2001a):

- Information resources that allow for open linking, by providing an actionable OpenURL as an additional hook associated with each metadata description
- Service components that provide extended services based on metadata and identifiers obtained through the OpenURL mechanism.

The library community was quick to recognize how this could offer the prospect of matching links to subscribed services, and OpenURL has been proposed as a National Information Standards Organization (NISO) standard. In essence an OpenURL takes the form:

```
http://(who you are, where you are, your institution)/(where you want to go)
 A                              B                                  C
```

The syntax of an OpenURL is that of a conventional URL mediated by the HTTP protocol (A). Known as the BASEURL, part B is more complex, containing information about the user that is unlikely to be known by a service hosting a link and which therefore has to be inserted during transport between servers. An interim mechanism for determining this information using current Web capabilities is to store the BASEURL in the form of a cookie in the user’s browser. This part could potentially be used to determine authentication for each user; its main purpose in early implementations is to point to a chosen resolver, most likely that of the user’s institution, a server that holds information about services the user is permitted to access.

The final part of the OpenURL, the QUERY (C), more conventionally points to the referenced object. This might be formed of an identifier such as a DOI, explicit metadata derived from an authored reference, or partial metadata that will allow a secondary service to identify the required document. Metadata might be described in the XML-compliant Dublin Core format.

An example service using OpenURLs might be based on CrossRef–DOI as well SFX services, as shown in Figure 4.1. A CrossRef link in the client, for instance, would typically contain a DOI-based reference. In this case the DOI proxy recognizes from the user-set cookie a direction to a local institutional resolver, running SFX software, to which it returns an OpenURL. In principle the number of steps shown in Figure 4.1 could be reduced by returning the OpenURL directly to the local server, with sufficient metadata determined from the DOI by the Handle service. The local server then queries databases of resources for locations of accessible versions of the referenced document.

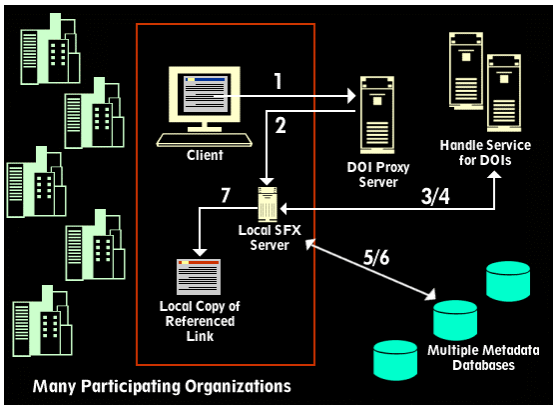



**Figure 4.1 Possible OpenURL architecture for locating metadata services (from Beit-Arie *et al*. 2001)**

The significance of the above example is less in the specific services involved but in the extension of the link from a simple document-to-machine relation to an open linking strategy that embraces multiple machine–machine communication in which the origin and version of the resulting document displayed to the user cannot be known in advance by either the user, author or link service provider. The use of intermediary proxy servers in link resolution, and in Web communication generally, is not new, but this illustrates the growing extent of the machine chain in which each machine does more than relay messages, instead transforming those messages according to the nature of the service provided by the respective machine. This has wider implications for scholarly information services, especially journals, than coordination and control of library services.

### 4.1.4 Progress in libraries: summary

In the library the ‘appropriate copy’ is most likely the one that it can deliver from the journal services it pays for. With the growing emphasis on site licences and support for new publishers that promise to compete on price, and with the OpenURL framework to mediate access, libraries have taken at least three major initiatives to consolidate their role at the centre of academic information services, in particular for journals services.

By focusing on site licenses with the largest publishers as the primary instrument for developing journal collections, however, library administrators may have assuaged long-term problems with short-term solutions, and so have not convincingly been able to extend their traditional interest in preservation of content with similar protection for preservation of access.

Nevertheless, this section has shown that the actions of libraries are contributing significantly to the characterising feature of the emerging infrastructure, that is, the expansion of access to electronic content to more users. According to Hellman (2001b): “eventually everyone will have a link server (or personal link page) operating as a plug-in to their web browser, with preferences customized transparently to each individual. Authentication and rights management will be built in, and the resulting experience will make browsing the professional content as simple and easy as the free-content web is today. It's a sad reality today that paying for high quality content results in a poor user experience because of all the primitive and clunky ways that access control is implemented.”

## 4.2 The foundations of open access

Progress by libraries in meeting the two criteria for a seamlessly integrated scholarly literature, that it is complete and free, may be illusory – the literature only *appears* to be free, and the cost of supporting it remains large. A stronger basis for ensuring that research papers are freely available to all, or at least a version of every paper is free, is promoted by the concept of open access. As was argued in section 2.10, open access will be the defining feature of network publishing, especially for scholarly papers. Open access describes a work that is available free to *any* reader.

Publishing has changed with the advent of electronic communication and the Web. Journal publishers have changed too – some unwillingly (Richardson 2001) – faster and to a greater extent than most would have anticipated in 1995, as chapter 3 showed, and in part this can be attributed to the threat of competition from open access sources. The full impact of the changes has yet to be felt, however.

Publishers may have acted to divert attention from free access to published papers, but there are two other challenges they have to confront: the cost-efficiencies of open access services, and the implications of the emerging architecture that these services will exploit.

The most prominent model for open access is eprint archives. ArXiv is the largest author-archived collection of papers. In scale only Research Papers in Economics (RePEc) comes close, and its model is different, pointing to distributed archives of papers for which it stores the metadata rather than the actual papers (Krichel and Warner 2001a). There are larger archives of scientific and technical papers managed by single organisations, such as Highwire Press, but these are not author-self-archiving sites and use different funding models to maintain free access.

ResearchIndex is not an author-based archive either. Through search engine-like crawling of author Web sites it has accumulated a rapidly expanding collection of some 500k papers in computer science. The service caches versions of qualifying papers, which users can access, but its primary purpose is ISI-like indexing of the citations and references contained in the papers, providing connections between, and context for, the stored works (Lawrence *et al*. 1999). Again, the process is wholly automated, and could not have achieved this scale within three years if it were not so. In turn the project was predicated on the ability to access a freely available, if disorganised, layer of papers on the Web, and thus can be regarded as one of the foremost examples of an open access service acting on the open access literature.

Table 4.1 highlights some major examples of open access services, descending in size, and considers the implications for scope, funding and other critical features for research such as peer review. Inserted in this table is the model built to inform this thesis on the impact of new forms of dissemination in this open access environment, *Perspectives in Electronic Publishing*, which is described in chapter 7. For reference, a conventional journal is included at the foot of the table.

As arXiv demonstrates, providing open access to scholarly papers in not an issue of scale, just of defining core functionality that the expertise of the service can support. At the other end of the scale *D-Lib* magazine, perhaps the best known e-journal covering the topic of this thesis, publishes 50 or so original articles each year, not unlike many other specialist journals. These are two examples that provide different services to contributing authors but are motivated and predicated upon a single shared realisation among those who produce and support them, that scholarly research is improved and enriched when access to the results of that research are available to all.

Table 4.1 is selective to the extent that it highlights open access services (apart from *Serials Review*) at the expense of others, such as bibliographic services, and distinctive abstracting and indexing services such as ISI's *Web of Science*. It is designed to show, in scale, the range of services that can embrace open access. All are products of electronic-only services since 1990.

Only perhaps arXiv and *D-Lib* can claim to have stable funding arrangements. With more libraries agreeing to onerous site license schemes, reliable income is likely to be a continuing strength of the one non-open access service listed in the table, the subscription journal, especially those owned by large publishers.

It is notable that peer review is a contributory feature of only the smallest, journal-scale services, although authors are able to post copies of peer reviewed papers in eprint archives.

**Table 4.1 Examples of open access services in the emerging scholarly publishing infrastructure, listed in order of size (as viewed in September 2001)**

| **Type of service Providing access to full papers** | **e.g.** | **How many papers are accessible from the service?** | **Exclusive papers** | **Peer reviewed** | **Funding of service** | **Scope of service** | **Source, location of papers** |
|---|---|---|---|---|---|---|---|
| Search engine | Google | Unknown, very large | No | No | Commercial, advertising supported | All disciplines, not just scholarly | Any Web-accessible sources, local cache |
| Author self-posting to personal servers | Departmental, institutional archives | Unknown, most | No | No | None (implicitly, institutional) | All disciplines, often disciplinary | Authors, stored locally by service |
| Automated indexing and linking services | ResearchIndex | c. 500k | No | No | Corporate research | Single discipline | Distributed Web-accessible sources, local cache |
| Organised eprint archives | arXiv | c. 180k | No | No (Some author-deposited peer reviewed postprints) | National and institutional funds | Typically disciplinary-wide | Authors, stored locally by service |
| Distributed eprint archives | RePec | c. 150k | No | No | None currently | Single discipline | Authors, stored locally by individual archives |
| Distributed archives | Open Archives Initiative, eprints activity | c. OAi-compliant archives excluding arXiv, RePec | No | No | Institutional, library funding | Multi-disciplinary, individual archives often disciplinary | Authors, stored locally by individual archives |
| **'Supra-open access'** | **PeP** | **c. 600** | **Some original content** | **Open review, not peer review** | **?** | **Focused by topic** | **Web-accessible sources** |
| Open access e- journals | *D-Lib*, *Journal of Electronic Publishing* | 25-50 new papers per year | Some | Yes | Research organisations; professional societies; host institutions | Subject-specific | Authors |
| Print journals | Serials Review | c. 30 new papers per year | Yes | Yes | Subscriptions and site licenses | Subject-specific | Authors |

### 4.2.1 Open access: layers of services

Another way of understanding how open access can make an impact is to view these services as cooperating rather than individual components within a larger framework. A three-layered, multiply-interconnected, hierarchical 'research communications infrastructure' proposed by Ginsparg (2001) shows how these types of services might contribute (Table 4.2).

**Table 4.2 Ginsparg's three-layered hierarchical view of a 'research communications infrastructure'**

| | |
|---|---|
| Knowledge layer | e.g. American Physical Society (APS), *Journal of High Energy Physics* (JHEP), *Applied and Theoretical Mathematical Physics* (ATMP) |
| Information layer | e.g. Google, ISI, PubSci, PubMed |
| Data layer | e.g. arXiv, California Digital Library (CDL), Centre Nationale de Recherche Scientifique (CNRS) |

Explained by means of examples rather than descriptively, this model offers a loose framework for anticipating forthcoming changes. Like Table 4.1, Ginsparg's view is founded primarily on open access services, from arXiv in the data layer to Google in the information layer. At the knowledge level are physics journals such as *Journal of High Energy Physics* (JHEP) that 'overlay' arXiv: services that can add 'synthesizing information', and "partition the information into sectors according to subject area, overall importance, quality of research, …; and can maintain other useful retrospective resources (such as suggesting a minimal path through the literature to understand a given article, and suggesting pointers to outstanding lines of research later spawned by it)." All three levels provide access for users, but perhaps only the knowledge level, with its synthesising objective, might be aimed at both experts and non-experts.

The fluidity of Ginsparg's model is emphasised by the inclusion of paid-for services from ISI, for example, in the information layer. In the emerging infrastructure it is likely that all services will provide direct access to papers, in ISI's case by reference linking and citation analysis. If all papers are accessible in the data layer – disciplinary and institutional archives in the Ginsparg model – then all other services will be providing filtered access of some kind to papers, in which case the distinction between the information and knowledge layers may be moot. In the current framework ISI bases its information services on journal contents, rather than on Ginsparg's data level services.

An interesting feature of both Table 4.1 and Ginsparg's model is the inclusion of Google. Developed within a research establishment, Google might not ordinarily be considered alongside other specialist research services, its user base being far wider. The signs are, however, that for many researchers it is the first service they select when seeking cited papers (Letter to the *Guardian Online*, 13 September;

also, see comments in the evaluation results in chapter 9). This is likely to be a significant factor in the shaping and viability of services that might appear in Ginsparg's knowledge layer.

Whatever analysis of these infrastructure models is attempted, there is one inescapable and startling conclusion: in every layer there is an established service, large or small, that is free to all users. Not every conceivable service that researchers might want or need is represented, but the models demonstrate, Ginsparg argues, "the key possibility of disentangling and decoupling the production and dissemination from the quality control and validation" for scholarly papers. This is ultimately the critical enabling feature of most open access services, the ability to create focused services independently of a grand package, as represented by the printed journal. With the journal package every component of the service is owned and produced by a single provider whatever the cost, with no flexibility or choice for the user.

The task is to fill the data layer envisaged by Ginsparg, in all academic disciplines. Decoupling dissemination from quality control will improve the prospects for this task, but will be resisted by many commercial service providers, and achieving it for more papers, more disciplines and more users will not be straightforward, as chapter 5 explains.

To be truly useful, independent services have to be underpinned by interoperability standards, the ability to share data efficiently through machine interfaces, so that user queries can be handled consistently by multiple services. Preliminary moves to establish criteria for interoperability are discussed below.

### 4.2.2 Legitimising open access: the Open Archives initiative

If there is a striking feature that unites the services described above, other than the principle of open access, it is the lack of a common framework – management, technological, social, or financial. The Open Archives Initiative (OAi) began by seeking to legitimize and provide a technical framework for the development of eprint archives. Table 4.3 shows the milestones, both managerial and technical, that have set the agenda during its short history.

The OAi was careful not to associate 'open' with 'free', and within a year of the first meeting, by September 2000, its mission was formalized and extended to cover all objects that might be found in a digital library. What was to be open was a simple metadata set describing the objects in an OAi repository. This minimal metadata is sufficient for repositories to communicate their contents to other OAi repositories and services using the OAi Metadata Harvesting Protocol (MHP). While this change

of emphasis has had the effect of raising the profile of OAi, reaching out to new and possibly influential supporters among libraries and publishers, its immediate impact for eprint archives has been blunted, at least in terms of the number of new OAi-compliant eprint archives it has stimulated.

**Table 4.3 OAi timelines: development of management and technical frameworks**

| **Timeline of OAi management framework** | **Development of OAi technical framework** |
|---|---|
| • Conceived as the Universal Preprint Service (UPS)<br>• **Oct. 1999** First meeting in Santa Fe, NM, a forum *to discuss and solve matters of interoperability between author self-archiving solutions*. Re-named the Open Archives Initiative.<br>• **Feb. 2000** Santa Fe Convention released, defining the technical and organizational framework. | • Feb. 2000 **Santa Fe Convention**, defines:<br>– Open Archives Metadata Set<br>– Harvesting interface based on a subset of the Dienst protocol<br>– two classes of participants<br>1. data providers expose metadata about content<br>2. service providers issue protocol requests to data providers |
| • **Sept. 2000** OAi extends interoperability framework beyond eprints - *develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content* - and appoints steering committee. | • Jan. 2001 **OAI Metadata Harvesting Protocol (MHP) Version 1.0**, an application-independent interoperability framework that can be used by a variety of communities engaged in publishing content on the Web. Embodies the principles of the Santa Fe Convention but replaces the Dienst-based implementation<br>• Jun. 2001 **MHP Version 1.1** updated for W3C XML Schema specification recommendation of May 2001 |

It could be argued that OAi and its MHP is now just another library protocol, like Z39.50, but even here some intriguing possibilities begin to emerge, as highlighted by Lynch (2001b): "A Z39.50-speaking server can fairly easily be made MHP-compliant, and I would expect to see the development of gateway or broker services that make Z39.50 servers available for open archives metadata." Z39.50 is typically used within libraries to search distributed library catalogues; OAi harvests metadata from collections of documents, or digital objects. Once again, machine–machine communication could be seen combining previously distinct information services.

Throughout, the OAi technical framework has emphasized simplicity: "The Santa Fe group wanted a very simple, low-barrier-to-entry interface, and to shift implementation complexity and operational

processing load away from the repositories and to the developers of federated search services, repository redistribution services, and the like" (Lynch 2001b). As a result, the most immediate impact of OAi has been less in the area of new archives (data providers) than in the area of new services (service providers), with a number of innovative if experimental services emerging:

- Arc, a cross-archive search service (Liu *et al*. 2001)
- Kepler, a Napster-like personal archive creator and central harvesting and indexing service (Maly *et al*. 2001)
- Torii, a portal service focused on arXiv (Bertocco 2001)
- OpCit, a reference linking and citation analysis for Open Archives (Hitchcock *et al*. 2000), again focused on arXiv.

A forerunner and hybrid of these services, the UPS prototype was built for the first meeting of OAi in Santa Fe to demonstrate end-user services such as search and linking (Van de Sompel *et al*. 2000).

As the first generation of OAi service providers, these services are initially targetted at end users. It is interesting to speculate, however, that they could themselves become data providers to other services that will harvest the processed data, transform it and add it to data harvested from other OAi sources, thereby opening up even richer routes for users to find precisely the information they need.

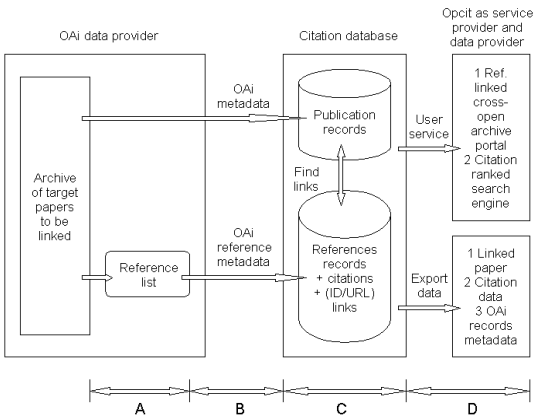



**Figure 4.2 Extending interaction between OAi data and service providers: the Open Citation project example**

The Open Citation project, for example, is experimenting with an OAi-compliant but extended metadata format, the Academic Metadata Format (Krichel and Warner 2001b), to create a machine interface to its citation database. This is primarily to enable these data to be re-harvested by the original document archives and displayed as part of their native services to users; it is also possible for these data to be used by other service providers (Brody *et al*. 2001). Figure 4.2 illustrates the interaction between OAi data and service providers in this reference linking application. In principle this schematic could be extended to represent further collaboration between providers in a chain sequence; at any point the chain could loop back to serve the resulting data to the original OAi data provider.

## 4.3 Digital library architecture: the Web services analogue

Development of machine–machine services in scholarly digital libraries are likely to be founded on a new form of distributed computing environment, the emerging standards-based architecture of 'Web services': "as well as being a problem domain that is well suited to the web services approach, the architecture of a digital library is very similar to the architecture required to support web services" (Gardner 2001). Gardner compares the library focus on information with the emphasis of Web services on applications, noting how the digital library community is used to applying verbs such as 'publish' and 'locate' to documents. This has echoes in an IBM definition of Web services:

> *Web services are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. Web services perform functions, which can be anything from simple requests to complicated business processes.*

Web services are motivated by the need to connect these business processes, especially databases, as was apparent in the origins of XML (Bosak 1997). The basic platform for Web services, as for OAi, is XML plus HTTP, maintaining the ubiquity and simplicity of the Web. Web services are based on three mechanisms:

- to register a service (e.g. Web Service Definition Language, WSDL)
- to find a service (e.g. a registry such as Universal Description, Discovery, and Integration, UDDI)
- to communicate (e.g. Simple Object Access Protocol, SOAP).

Existing public Web services rely on terms that are already well understood by all parties, or on publicly available data on the Web where the interface can be automated – address lookups, eBay auctions, Amazon book ranks (Shirky 2001). It is no coincidence that OpenURL resolvers are most effective when common identifiers (ISBNs, ISSNs, DOIs) can be plugged into URLs for other services (Powell 2001).

Microsoft, too, is building Web services. The most widely promoted of these, Passport (working names can change frequently), enables users to manage their 'identities', and other critical information used for transactions, in network data repositories. For example, a consumer publication will not need to create its own user authentication service, instead delegating it to Passport. In principle the user's Web experience would be instantly simpler, although there are social, commercial and legal (and hopefully standardisation) issues to be resolved.

Digital libraries are developing similar cross-community services (Gardner 2001): “many of the issues that the digital library community has been looking at, such as metadata for discovery, authentication and authorisation, and business models for accessing intellectual property, are also applicable to web services and must be addressed within a web services context.”

These user services become possible by enabling the underlying framework for machine–machine communication to remove the user from all but the primary aspects of a transaction, such as requests, checks and confirmation.


**Figure 4.3 Gartner technology hype cycle (from http://java.sun.com/features/1998/03/images/year3/original/gartner.curve.jpg)**

Web services are currently on the steep upward curve of the Gartner technology hype cycle (Fig. 4.3), suggesting expectation is ahead of implementation by quite a margin, leading some observers to question the value of existing applications – “Let's just hope the real advantages of sending data between small or private groups doesn't get swamped by the hype of perfect but unachievable automation” – and others to caution about the effectiveness and costs of XML-based applications (Bosak 2001).

Digital library services are rarely accompanied by hype on the scale of other Web technologies, and some of the better-established applications in the areas of metadata (Dublin Core), search (Z39.50), authentication (Athens) and rights management (DOI) could be said to have hurdled flatter curves. Web services are effectively the ‘plumbing’ that will connect these services. Thus, Web services may not fulfil all expectations, but will serve the digital library infrastructure where the applications and requirements are already well defined.

## 4.4 Access and interfaces: implications for journals

For scholarly publishing this scenario of cooperating Web services raises new questions about the role and pre-eminence of journals. In print form, the limited availability and the packaging of journals encourages browsing, but for digital products how valid is this beyond perhaps the user’s core journals? Digital information, rich in media and resources, formal and informal, mediated by multiple services, presents the user with an array of choices that might answer his or her queries most efficiently.

Those queries might be expressed as input to a search engine, or by selecting a link. Where might these citations come from? Personal emails, discussion lists, open access services such as OAi, eprint archives, newsletters, library services, Z-gateways and academic subject portals, as well as formal research papers and commercial indexing services. There will be many more. Once users become accustomed to a service, a citation becomes a *recommendation*. The journal with its veneer of peer review is no longer the exclusively trusted source, nor in this scenario can it remain the exclusive provider of peer-reviewed papers.

Harnad has long argued that the role of journals should be scaled back to the single essential function of quality control, in the form of managed peer review. In this context the magnitude of the implication becomes apparent.

Beyond peer review and a list of contents, what can a journal offer? Users typically want to list all papers on a topic, or all papers by a given author. By tagging eprints with details of the refereed journal sources, as is done for papers on high-energy physics in arXiv, it becomes possible to mimic a given journal title too. OAi should be sufficient to allow the journal contents to be replicated; eprint archives and other open access sources will support access to full-text versions of the listed contents. This is one way of building journal ‘overlays’ on existing document databases such as eprint archives, as described in chapter 3.

If paid-for journal services can’t offer a cited document that is known and available on the network it will be found by other means, and used. For the time being users will have to be careful to check the provenance and versioning details for sources found via these new services, but mostly they will prefer to be given a full copy of the recommended document, rather than not, and be allowed to judge its veracity for themselves.

The journal package has traditionally been bound in issues and volumes. With the advent of multiple networked sources mediated by services such as OpenURL, the binding has been unstitched.

If exclusive access is to continue to be the primary attraction for journals, it will depend not only on the ability of the library framework to deliver seamless access to everything, but on allowing access by automated services, which are becoming the new intermediaries in the scholarly information chain. Otherwise users will be directed elsewhere. However unwilling some publishers may be to give up exclusive rights to papers, exclusivity will become untenable once users become aware of the capabilities of multiple cooperating services, of which current OAi services are an early example. It makes no sense for data, in the form of scholarly research papers, to be locked into sealed containers,

the effect of exclusivity, when the cost of committing these papers to managed open environments such as eprint archives is minimal and for which Google searching, autonomous indexing, Internet Archive preservation, etc., come for free.

There is an alternative, and it is not just in the creation of high-quality, value-added versions of papers, although that is one possibility. It is the creation of interfaces.

The primary interface for a computer user is the screen, which defines what the user can see and do. The interface should enable the user to make sense of the services available. For many users, especially new users, personal computers are regarded as complex. The operating system removes some of the complexity, mostly management of the machine, but still the user has to manage a range of programs and documents as separate components.

With the emergence of the Web, a single component, the Web browser, has come to dominate the user interface. Once on the Web, however, the user has a new level of complexity to contend with: "Because basic features are designed differently everywhere users go, they never learn how to operate them and thus usability is reduced." (Nielsen 2001) Web sites look different, and have different navigation schemes; beyond search and links there is little interaction or cooperation among sites, so users may have to repeat operations on multiple sites to find what they want; and whenever a transaction is necessary the user has to provide personal information such as a password that is unique to each service. According to Nielsen, the "most eagerly awaited" Web services "are probably the unified login and the ability to release funds without using credit cards".

With its Windows XP operating system, Microsoft has the objective of taking the interface beyond screen appearance and machine control towards managed network services. On the Web the most obvious interfaces are the ubiquitous portals. Some such as Yahoo have been massively successful in attracting users. Yet Microsoft's approach with XP and its supporting .Net services are predicated on its belief that no portal yet has an unassailable hold over its users.

As on the Web, portals have become important interfaces in the scholarly environment. Portal strategies – by publishers (e.g. Elsevier's ScienceDirect) and associated networked information services (e.g. Ingenta), and by library resource discovery networks (e.g. JISC's RDN) – have yet to establish a pre-eminent model. This is because all have concentrated on content, mostly owned content. The best next-generation portals will build services on top of content, and for researchers will become the starting point for *all* lines of enquiry.

The JISC RDN is a good example of building on content to provide new services and adaptable interfaces. The individual subject networks, in medicine, engineering, humanities and others, can now be searched as though they were one unified repository, and an interface presenting users with this search facility can be embedded in any library Web page: “students and staff can use the RDN search facilities while remaining within the familiar look-and-feel of their university or college's Web site”.

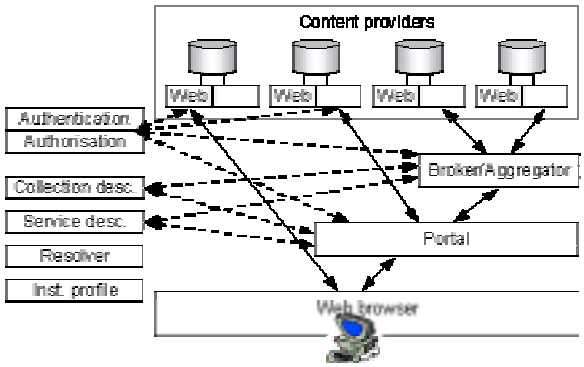



**Figure 4.4 Technical architecture components of the Distributed National Electronic Resource (from Powell and Lyon 2001)**

Guiding the implementation and organisation of these services is the JISC DNER. The scoping architecture of the DNER embraces all possible sources of information described in this chapter (Fig. 4.4). Powell and Lyon (2001) show sequentially how this architecture might develop and how the different service layers might interact using OAi, Z39.50 and Rich Site Summary (RSS), another metadata-based means of describing the content of a Web site. In the DNER, content services might be provided by portals and aggregators or any Web-accessible site, mediated by descriptions, authentication and resolvers.

Demonstrating how far these ideas have progressed since the JISC-funded electronic library projects of the mid-1990s, a set of standards and guidelines covering content creation, preservation and presentation, including access for disabled users, as well as interoperability services, are recommended and in some cases mandated for contributing partners. Innovative projects can no longer be conceived in isolation.

At this level there will be no distinction between library services, commercial services and open access services. Users will want to know everything that is relevant to a query. Not every version of every document will be accessible, and competition between scholarly interfaces will be based on giving users what they want as often as possible and minimising redundancy. Users will choose interfaces that integrate, both explicitly in the user interface and through the underlying architecture, the services that most often produce the results they demand. Interfaces must therefore offer as many services as needed, while hiding the complexity of too many services. In other words, they must become the Windows operating system of the Web user.

## 4.5 Emerging infrastructure: summary

This chapter has illustrated three examples – based on OpenURL, the Open Citation project and the DNER; many more could have been chosen – of the emerging architecture of the scholarly information environment. Each shows that the client–server architecture of the Web is being supplemented by multiple intermediary services. These are not merely concepts but are in development.

This chapter has emphasised the role of open access publishing for a certain type of resource, the scholarly journal paper. It is important to note there is no compulsion towards open access for publishers, libraries or authors, who can all make their own decisions about what is in their interest. Their decisions, however, should not be based on the presumption that nothing has changed. They need to consider these model architectures and changing user behaviour in judging the validity of their actions.

Journal publishers continue to compete for authors and for exclusivity, and can claim some success in preventing a significant migration of authors to open access sources such as eprint services, other than in physics and economics, and to a limited extent in mathematics (Krichel and Warner 2001a). Other factors such as negative peer pressure are limiting participation in eprint services by authors in some disciplines (Glass 2000). The transition from print to predominantly electronic forms of scholarly dissemination is still in its very early stages, however. The impact of digital publishing, in terms of moving the boundary in scholarly communication between freely available and paid-for services, has yet to be fully appreciated beyond the first adopters.

Control of the peer review process by journals remains critical to the publishers' strategy (Guedon 2001), with the consequent grab for copyrights. The precipitate moves to digitise journals in the mid-1990s, the 'miraculous' collaboration to form CrossRef to support journal reference linking, the introduction of library site licenses, even the concession of time-lapsed open access by some powerful biomedical journals, have all cleverly served to deflect attention from the limitations of the long-established journal model for digital applications.

While taking important steps to reinforce their role in the provision of paid-for content, libraries and have been relatively restrained in introducing open access sources into their controlled information environments. Economic considerations, and the success of open access providers in demonstrating stable and reliable sources, will determine to what extent this caution might be overcome. OpenURL offers a framework in which all sources, whether free or commercial, can be integrated. In addition,

libraries with institutional portals can enhance these as user interfaces with selected components from distributed library services.

The process of change within libraries may be longer and more treacherous than even this example suggests, however. For the ‘new genres’ of scholarly communication that fall outside the traditional canon of print scholarly literature, which include teaching as well as research materials, “the inherent uncertainties of these materials will compound the difficulties libraries face in defining their roles in the management of the new scholarly communications.” Lynch (1999) As a result, Lynch contends, “it will be necessary for libraries to reconceptualize their activities virtually from first principles.”

Whether served from the library or externally, interfaces to services will replace interfaces to journals as users’ primary starting points for literature research. These interfaces will be enabled by Web services founded on the Semantic Web (Miller, *et al*.): “The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines *not just for display purposes, but for automation, integration and reuse of data* across various applications" (italics added).

The emerging infrastructure of scholarly communication does not exclude journals, nor commercial journals or paid-for library services, but it will preclude journals in all but exceptional cases from seeking exclusive ownership of the knowledge content of published papers, and will require a commitment to cost effectiveness and interoperability between services.

As was observed in *Nature*: “The ability to click from an abstract or citation to the full text of an article is prompting a shift in the way that journals are used. Scientists often care less about the journal title than the ability to track down quickly the full text of articles relevant to their interests. Increasingly, users view titles as merely part of hyperlinked 'content databases' made up of constellations of journal titles.” (Butler 1999)

# 5 Reshaping scholarly publishing

The scholarly journal literature has gone through a process of digitization: "Right now we're only dealing with the equivalent of scholarly digital molehills, but they will soon become mountains of data – disconnected, unevenly linked, and not effectively searchable stuff." (Peek 2001). Chapter 4 showed this is changing, but having spanned at least one generation since the first e-journal experiments, there is clearly still a long way to go, a transition spanning more generations perhaps. In which case it is one thing to know what is happening in electronic publishing now, another to project what shape it might take in the, possibly distant, future before stability and consensus about the purpose and form of scholarly publishing, which the printed journal has enjoyed for over 300 years, reemerges. This chapter will try. It is necessarily more speculative, and more argumentative, than other chapters. The analysis leads to a proposal for a new model of scholarly publishing,

## 5.1 Understanding 'online': the impact on journals

The conventional arguments for publishing journals online might best be summed up as follows: journals should take advantage of new online features – speed of distribution, multimedia attachments, etc. – while barely deviating from the paper format that has proved so enduring. Electronic journals, to avoid remaining isolated in a 'ghostly netherworld', must conform to the popular paper model, argue Kling and Covi (1995).

There are qualitative reasons for believing that this view, which even Kling and Covi admit has 'subversive' potential because it alters the scholarly communication system 'while seeming to be a routine part of the dominant paper systems', is too conservative. Merely by virtue of making journals available electronically, established publishers are accelerating the change towards a product that will be wholly different from today's journal. Central to this change is to recognise the impact of a new medium, the online medium.

The Web represents the new medium today. The Web is the service through which e-journals are delivered and, by migrating to the Web, journals are participating in the new medium. Information is the lifeblood of professionals, so any change is bound to have a major social impact (Wilson 1997).

It is over 35 years since McLuhan (1964) famously pronounced 'the medium is the message'. It is no coincidence that McLuhan's theories can be interpreted as foretelling the impact of the Internet in its current form as a global communications infrastructure. Press (1995) speculated on McLuhan's

thoughts on the Net, noting his view that “the hybrid or the meeting of two media is a moment of truth and revelation from which new form is born”. The implications have yet to permeate, otherwise how could it be assumed – as does the Acrobat electronic journals model, for example – that the revolution in computers and communications has happened simply to facilitate the delivery of something that is built and destined for the older technology of paper?

Imposition is the prerogative of a new medium: “We shape our tools and afterwards our tools shape us”, McLuhan said. Understanding the effects of a new medium enables us to anticipate and exploit the change, but also to recognise that our ability to control the change is not straightforward.

Since this ‘tool’, currently in the form the Internet and its related services, is largely shaped, it follows that it in turn must now be shaping us, its users. The clue to the longevity of McLuhan’s observations is his emphasis on their psychic and social consequences: “For the ‘message’ of any medium or technology is the change of scale or pace or pattern that it introduces into human affairs.” This change of pace is the crucial feature of the Web. From speed comes the ability of users to act on information and respond instantly, what is frequently and glibly referred to as ‘interactivity’.

Publishers of electronic journals recognise the need for speed, but see this need as reducing many weeks delay in publication to fewer weeks delay (Taubes 1996b).

What would journals be like if they were free and available instantaneously? There may be good reason to hesitate. Speed, in the context of information, is a double-edged sword. Information, if it is useful, demands a response –an idea, a note, an email, a revision, new data, another document (Harnad 1995b). Given the inherent speed of the medium, the first response will also tend towards the instantaneous. The effect of the online medium is first to invite and then demand that the response is integrated within the information chain. Even journals, with their well established editorial routines, cannot resist this trend as they venture online. Speed of publication will become a competitive edge for many academic authors, just as it is in other areas of publishing (Hitchcock 1996).

If the desire for speed among authors is not new – “Fast-moving fields have always engendered a sense of urgency” – then the added dimension of the new medium introduces new threats, as one mathematician cautioned: “in the past the people who moved on too fast did not seriously damage the literature. Instead, they reduced their own long-term impact on mathematics. Now it is technically feasible to damage the literature.” (Quinn 1995)

Interactivity is not now an optional add-on but an essential part of the process. Early online newspapers discovered this. According to *Wired*, when the *New York Times* went online in June 1995 "hundreds of users tried to message @times and many asked for e-mail addresses for reporters, (but) there was no one for them to talk to". In contrast *Time* magazine online actively encouraged contact from readers, causing some observers to say that the online culture is changing the magazine. Echoing the point, the president of one online newspaper said: "Our communication historically has been: 'We print it. You read it'. This changes everything." (Katz 1994) While newspapers are quite different from journals, 'scholarly skywriting' (Harnad 1990) is an equally potent vision of academic journals adapting to the speed of online communications.

Nor is the process of writing immune to the demands of the online medium. It takes time to compose clear and unambiguous expression, that is, grammar, phraseology, also logical structuring of argument. But this is all sequential, and is not the way scholarly texts are read (Ubell 1997). Imposing speed leaves less time to construct sequential argument. The research paper will no longer be an isolated entity but, in active fields, ideas, results, data will all become part of a continuum (Rzepa and Murray-Rust 2001). Links offer non-sequential possibilities for presenting new perspectives (Landow 1992). Responsive writing will be integrated within large link structures.

## 5.2 Reshaping the 'journal'

A complete reappraisal of the basic journal suggests great possibilities and major changes to the framework within which scholarly publishing operates.

For many scholarly publishers their principal product will disappear: not the paper journal yet, but the component on which such journals are based, the exclusive paper. Journal revenues are generated through sales to libraries, but what publishers primarily sell is space to authors, with an associated imprint. Limitless electronic space is cheaper than paper, and with a multiplicity of formats available to authors, fewer of today's imprints will routinely attract exclusivity.

Given the demands of speed of publication and the removal of exclusivity, eprint archives can be recognised as a natural consequence of the new medium. With eprint archives the emphasis is on communication rather than publication, but communication is the principal purpose of most scholarly authors (Mermin 1991). Without electronic communication, authors had to go through a process currently recognized as publication. Exactly what constitutes scholarly publication is now unclear. Is it dissemination, or validation? Must it be for sale? The Web is blurring the distinction.

Within the physics archives dramatic evidence is emerging that speed of access to new papers is changing usage patterns. New papers are being cited sooner and more frequently. On average the citation peak for papers added to the archives in 1999 occurred after just four months (Open Citation Project 2000, http://opcit.eprints.org/tdb198/opcit/citationage/, see section 3.2), far shorter than the formal process of publication. In other words, authors are confident of citing works without the prior validation of peer review and journal publication.

### 5.2.1 Constraints on the new model journal

Not that communication is the only purpose of scholarly papers; the established hierarchy of paper journals demonstrates this is not enough. Authors want validation of their work and recognition for themselves, and this is conferred by publication in the most prestigious journals in their field.

What is becoming clearer is the pivotal role of peer review and all it confers in determining the inclination of a particular community towards the model of the physics eprint archives and the consequences – instant open access – or its rejection in favour of open, but moderated, access. Those communities that instinctively prefer the latter route immediately find their options limited. Why is this? Submission of a paper to a journal for peer review without first making the paper openly available is a clear indication of an author's priorities, and is tantamount to a contract with the publisher, typically on the publisher's terms, which are unlikely to include open access.

The effects can be seen in different communities. Biomedical scientists want open access to research papers after peer review and publication (Public Library of Science 2000), and as a result as yet have little open access within the preferred six month limit (Russo 2001). Most physicists too value peer review, but only after open communication of their work. As a result they have arXiv.

The prospect of eprint archives becoming the dominant form of first publication for new scholarly papers thus depends on the willingness of individual scholarly communities to consider new ways of supplementing quality control. While something approaching instantaneous delivery cannot be consistent with processes of quality control established by conventional journals, most obviously realized through peer review, new measures will have to develop.

### 5.2.2 Links: the bounding quality filters

One possible filter mechanism for quality control is the link. This reappraisal of the product of scholarly journal publishing remains incomplete in one crucial respect, thanks to another peculiarity

of the market for academic journals. Scholarly publishers do not sell information, but sell access to information. The most obvious manifestation of this is making connections between different items of information, collecting papers within a journal issue for example. Creating links, in other words, between information resources is the most natural publishing activity in any medium (Hitchcock 1998a, Hunter 1998). The emergence of the online medium presents an unparalleled opportunity to capitalise on this skill using the hypertext link.

Links are the currency of the Web. Creating, presenting and maintaining links is a difficult process. Link databases simplify link management, and link services can potentially present links from any word, phrase or object within a document (Carr *et al.* 1995).

For users of a link service the selection of link databases enables them to define the information environment in which they wish to work: there is no need to become lost in the information space of the Web, or to be artificially constrained by the physical limits of a paper journal. While links establish relevant connections between different pieces of information, managed linkbases can be used to define the boundaries of a navigable information space.

For publishers, when the link is freed in the form of managed link services it becomes a commodity. As the hypertext link ushers in a new authoring paradigm, so link services promise to become a powerful new publishing paradigm. Link services can also resolve the problem of quality control that the online medium introduces: let link services become the quality filters. Linkbases can be the ‘binding’ for the selected, or ‘quality’, information that customisable online ‘journals’ want to present. In other words, one application of linkbases is as an analogue to print publishing but in a form that naturally suits the dynamics of the online medium.

CrossRef (Pentz 2001a) demonstrates the emergence of link publishers, which charge users, or other service providers, a fee or subscription for continually updated link data, but without the associated rights to access the linked materials. The objective of link publishers must be to enable beneficial use of links on such a scale that the volume of use creates a cost-effective environment for information and link development.

### 5.2.3 Forces for change

Authors and readers of scholarly papers face the prospect of scholarly publishing becoming entirely electronic. Few publishers have admitted the possibility, among them the ACM (Denning and Rous 1995) and the American Physical Society (Langer 2000), but the realisation of all-electronic

publishing is well advanced. The transformation of journals to digital print mimics may have been uninspiring, but the most important fact is that it has happened.

Yet there is no obvious grand strategic vision from journal publishers. While publishers struggle for competitive advantage in this new environment, self-appointed guardians – authors, editors, users and librarians – ironically seek to protect the integrity of the scholarly literature from the forces, both visible and medium-dependent, lining up to reshape it. A series of initiatives, from PubMed Central (Varmus *et al*. 1999) to the redefined Open Archives initiative (Lagoze and Van de Sompel 2001) to OpenURL (Van de Sompel and Beit-Arie 2001a) to SPARC (Johnson 2000), can be seen less as challenges to publishers in the marketplace than attempts to protect and preserve scholarly publishing as we know it. These organised responses, which are concerned with archiving, controlled distribution, the role of the library and simple old-fashioned journals, may do more to save the journal publishers than the publishers can to save themselves.

So what forces are causing publishers to react in this way, and why?

- The network, and its associated computational infrastructure
- Users, acting collectively and individually

Elsevier's Tulip project (Borghuis *et al.* 1996) involved extensive evaluation of user responses to familiar journals delivered in electronic form, discovering that what users want includes:

- access to all information from one source;
- effective search capabilities;
- *high publishing speed* (timeliness of the information);
- sufficient journal and time coverage;
- *linking of information*.

The effects will be seen in the busisness processes of journal publishing, and in the collection and distribution of content, respectively:

- decoupling the different stages of publishing
- defragmention of collections of content, the journals

The critical hub for both changes is peer review. Peer review is both a part of the publishing process and a determinant of the character of the collected journal package. The assumption, based on the weight of published opinion and the interpretation of the conservative intent of initiatives cited above, is that peer review should be maintained as the primary arbiter of quality in published papers.

## 5.3 Decoupling journal content from publishing process

"We need to encourage a vigorous free-market competition based not on the value of the archive a publisher controls but on how much new value can be added to a free public resource." (Eisen and Brown 2001)

The realization that authors can so easily distribute their papers via electronic networks fundamentally alters the culture of scholarly publication, that if distribution is no longer a necessary element then neither is exclusive ownership of works. It might surprise non-academics not that academic authors give their papers for free to journals that they then buy back, often at very high prices, but that they also often give the publisher all their proprietary rights in the works as authors, which the publisher then owns in perpetuity. The consequent restriction on the ability of authors to reuse *their own works,* the 'Faustian bargain' as Harnad called it, of pre-electronic journal publication, can now be challenged by open access publishing.

Of course any good publisher will argue that publishing has never been about distribution, but about marketing, packaging – context and prestige – and finding and satisfying an audience for the work. If these are now the essential elements that justify publication, however, mere distribution from any lesser source cannot properly compete. Except, if the author-posted version is free and accessible, the reader has a choice. Journal publishers that continue to demand exclusive ownership of works do so to minimise the risk of their investment in the value-added version. Exclusive ownership doesn't benefit the author; it is simply an economic hedge that authors and the scholarly community can challenge. Competition with free versions of papers compels versions that are genuinely enhanced yet highly cost-effective because they provide added features that users will want.

The journal attempts to be more than the sum of its parts, that is, the journal framework is more important than any single paper, yet exclusivity is demanded to control those papers that fall within the package. This is how it happens. It is well known, but is worth elaborating because it affects the analysis of decoupling content from publishing process.

The essential steps in the traditional process of journal publishing are:

1. Shape journal: define, design, target
2. Attract authors
3. Review and select papers
4. Production: subedit, format, collate
5. Market and deliver

The publisher leads stages 1 and 2. By stage 3 there is a third-party, the author, whose interest is co-joined with the publisher’s, or so it is believed. Successful review of a submitted paper assures the author that the paper is fit to be viewed by his or her peers. In fact, if fulfilled correctly, the journal review has principally established that the paper is consistent with the profile of the journal. Nevertheless, joint interests have been confirmed between publisher and author, and a small cost has been incurred.

What are the implications of the peer review process? The publisher has learnt from a non-contracted expert witness, the peer reviewer, that the work probably has some value in its market. That value may be imprecise, but it is something that can be exploited in the journal framework. In terms of the paper, the publisher so far has only a small cost to recover, but in terms of stages 1 and 2 a bigger investment needs to be recouped and the paper has to be able to make a contribution to those costs.

### 5.3.1 The pivotal point: assignment of rights

What happens next is pivotal. The publisher needs the author's consent to publish the work. For many publishers that is not enough and the author may be asked to assign all rights in the work, that is, transfer complete ownership and beneficial interest in the work to the publisher. In part this is not unreasonable. If the author's journal selection has been good and the review has strictly been associated with the journal, then that journal ought to be the most appropriate publication vehicle for the paper. In an era of multiple media, however, authors would be better off reserving some rights that journals alone cannot serve, as we see in the steps below.

With all rights acquired the publisher puts the paper into its journal production process, stage 4. Invariably this incurs a larger cost than that of the peer review process and will be passed on to the end user. The end user is never consulted on the value of the production process. A journal may have especially high design and production standards, which can improve reading, but this is a sideshow. The main deal has already been done, and if the peer review has been effective the target end-user

needs to see the paper whatever the cost and effect of production. The evidence, presented by Tenopir and King (2000), is that while journals continue to be accessed at high levels, personal subscriptions to journals have been decimated over the last 30 years.

### 5.3.2 Eprint archives: the need to assert divisible rights

In print journal publishing the process is indivisible, but not electronically. Consider an author in a field that supports eprint archives. Prior to journal review the author deposits the paper in a freely-accessible archive. Within 24 hours the author's peers will have been alerted and are able to access and read the archive version of the paper. After review the author retains the right to self-archive the work and supplements the original eprint in the archive with a revised version that has satisfied the reviewers. The work – the words, the presentation – is entirely that of the author. The process of deposit in the archive takes a few minutes of the author's time so long as it was produced in a format that adheres to the requirements of the automated archive. Within 24 hours the author's peers are able to access the new version of the paper.

It is possible the purpose of the review might be diminished outside the scope of the journal. In practice the distinction is rarely made as a Journal-Ref, a note of the journal that has reviewed and accepted the paper, is added to the metadata for the paper in the archive.

The author benefits in every respect. Instant readership for the preprint is guaranteed in the archive; for the best papers the status conferred by the Journal-Ref tag assures continuing higher numbers of readings in the archive (Harnad and Carr 2000).

The publisher feels short-changed. Should it? That status has been attained by its efforts over many years, but what the publisher is left with is a commodity, which is what the paper has become, of dubious value now that exclusivity has been lost.

This is not necessarily the case. In physics, the only discipline with a large, universal eprint archive, journal publishers have continued to thrive in exactly this scenario. Data on profit growth may be hard to find, but there have been no reports of exceptional fall-off in subscriptions to journals in physics since the archive was launched in 1991.

Other publishers in other fields are more sceptical. Physics is a special case, they argue. Some journals invoke an embargo at stage 2, denying consideration to papers that have been submitted elsewhere, such as eprint archives (Harnad 2000). Or, with all rights in hand, publishers could

feasibly demand withdrawal of any archive versions prior to publication. It appears that non-physics authors might be denied the benefits of their physics colleagues.

Except authors can claim the benefits for themselves, by reserving the right to self-archive their own versions of their own works. Should authors feel any twinge of guilt towards the journal publisher they could surmise the following: if the journal has an appreciably high status – there seems little point in submitting if it has not – it has no doubt been achieved by fortune of exclusive publication, an advantage it no longer needs; also, if publishing is difficult and the publisher adds significant value that the self-archiving author alone cannot, then the publisher should be capable of competing with sources based on author self-archiving.

In effect, the paper has been decoupled not from the peer review process, which the author values most, but from the production process, a superficiality of less importance.

### 5.3.3 The publishers' dilemma

Driven by cost-cutting requirements over many years some publishers have allowed, or required, authors and editors to take greater control of the journal production process (Fig. 5.1). With the arrival of the Web this process has extended significantly, with editors able to take control of the publishing process too (Fig. 5.1c).

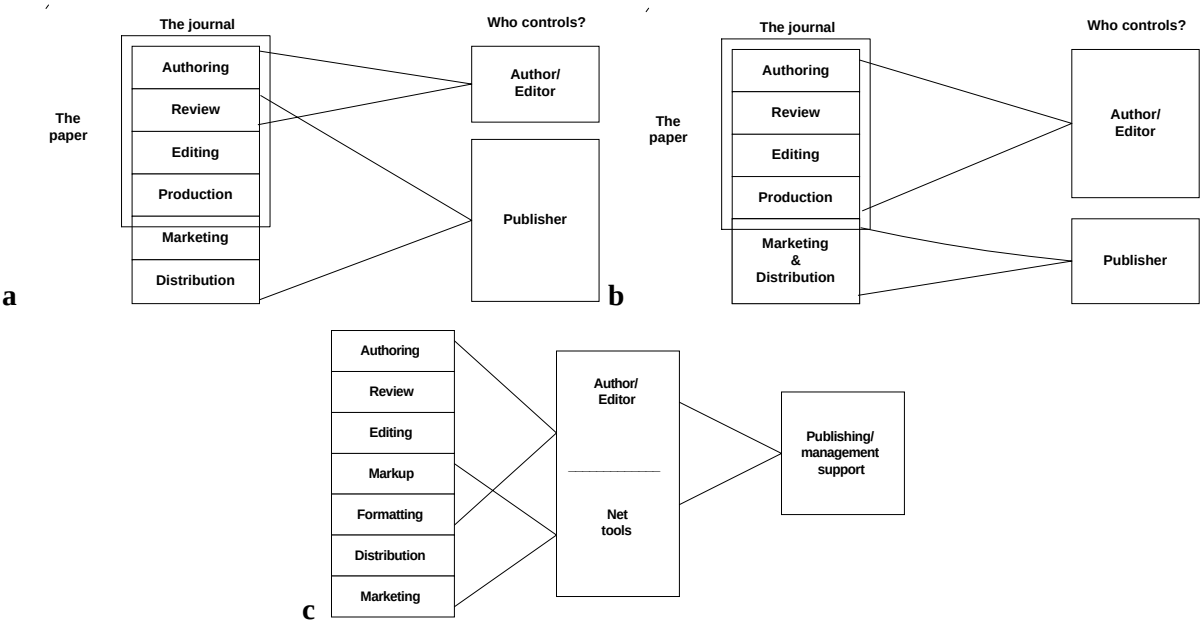



**Figure 5.1** a**, traditional journal publishing process;** b, **modern publishing process – the diminishing role of the publisher;** c**, publishing process for the Web journal**

A simple model of the traditional journal publishing process (Fig. 5.1a) shows who is responsible for the each stage in the development of the fundamental unit of published work, the paper. The publisher is involved directly throughout, and indirectly through the journal editor, often an academic but who is accountable to the publisher. Papers may be submitted by authors or, depending on the nature of the journal, may be commissioned, extending the publisher's influence into the authoring process. Clearly, the publisher takes responsibility for the collected work, the journal issue.

In the mid-1980s, desk-top publishing (DTP) tools failed to support the promise of small-scale journal publishing but widened the scope for authors to become involved in the production of their papers (Fig. 5.1b), although the practice is more typical of conference proceedings than journals. The publisher remains responsible for the core functions of marketing and distribution – selling the packaged product into the market – on which any good publisher can justify its role.

The Web has motivated the most dramatic change in the process (Fig 5.1c), with new e-journals produced in conjunction with professional publishing services just beginning to emerge in significant numbers. Some producers of free e-journals have developed tools to manage their electronic-only operations cost-effectively, and share these tools on an open source basis. One example, the International Consortium for the Advancement of Academic Publication (ICAAP), allows editors to track manuscripts through peer review and produce HTML or PDF using services on its Web site.

Disaffected journal editors or aspiring new journal producers will also find publishers such as SPARC and BioMed Central (BMC) offering the publishing infrastructure. SPARC was motivated to introduce price competition for journals (Case 1998). BMC in contrast will support free, 'smaller circulation' online biomedical journals under the editorial control of researchers "who believe that their particular (sub)discipline needs a publishing outlet operated on the free access model". BMC's business model for these journals includes charging for reviews and opinion pieces, advertising, and processing charges for authors, although the aim is to defer the latter fees to institutions.

The dilemma for publishers is, having allowed the process of decoupling to percolate prior to the Web, to what extent do they wish to resist its incarnation now?

### 5.3.4 Theory and practice of decoupling

Not surprisingly physics publishers have embraced the prospect of decoupling faster than others, adapting to a system based on three fundamental elements: a preprint server, an electronic peer-reviewed journal, and an electronic archive of past published papers. According to Smith (2000) "the

tension concerning responsibility for public distribution and communication of new work has been resolved in favor of the electronic preprint databases". In this sense journals focus on peer review but are no longer the primary communication medium, he concedes.

Other commentators have theorised on the process of decoupling and the consequent benefits. Describing the 'deconstructed' journal, Smith (1999) argued that most activities involved in journal publishing are independent ("quality control activities are not concerned with distribution") and therefore there is no obvious need for these roles to be controlled, and the resulting product owned, by a single publisher.

Publishers are alert to the danger but may be vulnerable to a change in perceptions by library administrators, according to Odlyzko (1999). The Scholar's Forum proposal (Buck *et al*. 1999), which aimed to wrest control for peer review and publishing value-adding tasks from journals to an arXiv-like 'document database', seems to have been still-born but indicated the concerns of influential academics. Phelps (1998), a university vice-chancellor, supported Shulenburger's (1998) proposal for a national electronic archive as a means to introduce competition: "we must find ways to introduce competition into every phase of the process that journals once performed as a bundled effort – quality certification, editorial improvement, distribution, indexing, and archiving."

In contrast to Smith's (2000) view of the reduced role for journals, both proposals go further, decoupling peer review from journals too: "the most important step ... is to create an alternative mechanism to provide the refereeing/certification process now provided uniquely by the editorial boards of print (and occasionally, electronic) journals." (Phelps 1998) The call for independent review boards is echoed by Edmonds (2000), who wants to reduce the effects of cost and ownership that often restrict works in journals to single audiences.

It could be argued that decoupling review from journals is a necessary step where there is no preprint culture and not likely to be one. In such cases control of peer review is critical in determining the ultimate rights owner, and has proved detrimental to demands for open access publishing such as Public Library of Science. Nevertheless, it contravenes the case advanced above that review should not be carried out in a vacuum but with a sense of the intended audience, and so may be best performed within a journal framework, unless a 'review board' or similar body can offer this facility.

Decoupling restructures the business of journal publishing; it is a not a technical process. A proper analysis separates people from process. Labour-intensive processes requiring personal

communications should remain within the journal framework. Specialist design or skilled subediting, for example, may be a competitive edge against services delivering the author's unvarnished version.

### 5.3.5 Resistance to decoupling

Few journals will survive the loss of exclusivity intact. Those that do will mostly have high impact factors, distinctive design and editorial, and large circulations, at least thousands if not tens of thousands of subscribers. The majority will struggle to survive in print as costs have to be cut to compete with free delivery. Production will ultimately be reshaped to add electronic features and to process papers in larger numbers.

Publishers may not be wedded to paper, but they will want to preserve their business models until proven alternatives are found. They have a number of weapons to protect their interests and resist change if necessary:

- Control of the peer review process
- Ownership and identity
- Contracts, obligations, loyalty
- Political and economic pressure

The most powerful will be the ownership of journals with strong identities established over many years through effective management and marketing. The academic reward structure is predicated on such titles. Highly motivated scholars can transfer the interests of a given field to a new journal or service, but without ownership cannot take established titles with them.

The political influence of publishers was demonstrated against a free search service for the physical sciences literature, PubScience, operated by the US Department of Energy (DoE). Butler (2001b) reported that a congressional committee lobbied by the Software & Information Industry Association on behalf of member companies – including Reed Elsevier, ISI, Chemical Abstracts Services and Cambridge Scientific Abstracts – had proposed a budget cut that could close down PubScience. The cut was “likely to have a chilling effect on other government-operated services, including the National Library of Medicine's PubMed Central”. PubScience had survived in tact to the end of 2001.

Beyond journals, publishers and other multinational media companies have lobbied successfully for draconian new measures to restrict dissemination of copyrighted content on the Web, most starkly in

the Digital Millennium Copyright Act (DMCA) in the USA, which not only reinforces existing copyright law but adds an extra layer prohibiting the use of machinery and software that can bypass copy protection systems. In 1995 many believed that cyberspace would kill copyright, but "the issue is not anymore whether copyright is dead. The issue is how many other values get sacrificed in the name of protecting copyright." (Lessig 2001)

The greater power afforded to copyright law might incline journal publishers to continue pressing restrictive copyright agreements on authors. If they relax copyright requirements, decoupling journal content from publishing process is one step towards reshaping journals, which will be accompanied by the defragmentation of content services.

## 5.4 Defragmentation: tackling access inefficiencies

There can be few users of Windows-based personal computers who have not used the disc defragmenter tool. Performance degrades rapidly when hard disc storage space nears capacity. Logical organisation of information held on the disc begins to break down as new files are stored where space dictates, so speed of retrieval of these files suffers. Applying defrag can ameliorate the process, reorganising the files into some logical and accessible order.

A similar process of defragmentation is underway in scholarly publishing, reversing the process of specialisation forced on paper journals by page constraints and other factors, and the consequent stagnation in the ability of non-specialist users to access these works. Mostly this is attributed to economics, the 'serials crisis'. For example: "less scholarly material is being subscribed to in absolute terms. Furthermore, since the amount of scholarly material extant has increased, as the 55% increase in serials titles demonstrates, faculty at Association for Research Libraries institutions have access to a reduced relative proportion of the scholarly literature than in 1986." (Shulenburger 2001)

This is a distraction, however. There are also the cognitive consequences of grouping knowledge into subject domains and the tendency to subdivide items within these domains into small units "in order to isolate individual thoughts. This drive to facet knowledge is represented by the increasing range of specialties found throughout the domains. When items are separated in this way relationships between them is minimized, and contextual meaning is lost. This isolation blocks the act of cognitive connection that leads to serendipitous discovery within the wider environment." (Allard 2001)

As some recognize, the need for defragmentation is acute. "With print based publishing the information contained in the niche journals, each with its own subscription barrier, has become

hopelessly fragmented. We now have a golden opportunity to reunify biology by being able to search, link and match across a swathe of papers, breaking down the boundaries between disciplines, and sub-disciplines." (Hopkins 2001)

Defragmentation improves access to works when dissemination of individual works is no longer tied to single fragments, the journal packages. According to Gardner (1990): "The dispersion of archived texts through the reproduction and distributed storage of serial journal issues – whether through print or electronic media – is the primary inefficiency of traditional scientific publishing." Defragmentation tackles this inefficiency.

The Association for Computing Machinery (ACM) saw that users wanted access to the whole corpus of its publications: "The business model and marketing campaign ... de-emphasized subscriptions to individual journals in favor of a single, annual access fee for unlimited usage of our Digital Library." (Rous 1999) This has produced faster growth of individual subscribers, the ACM claims, reversing steep declines in personal subscriptions affecting most print journals (Tenopir and King 2000).

The importance of integrating access to journals to offer comprehensive coverage within fields has been established in user studies (Borghuis *et al*. 1996, Kirstein and Montasser-Kohsari 1996) and seems to be the choice of scholars and librarians (Bailey 1994). A unified, or a richly linked, archive will boost research productivity "as fewer wasteful duplicative experiments are done", in the view of Eric Swanson, a senior vice president at publisher John Wiley & Sons. The time this saves can be "allotted to an increased number of promising points of departure identified by fast, sophisticated, and comprehensive searching." (Wilkinson 1998)

The most vivid evocation of defragmented scholarly publication is 'skywriting' (Harnad 1991). Even this is a small step in the direction of Engelbart's (1975) remarkable NLS oNLine System, described as a 'workplace' for knowledge workers, supporting dialogue and collaboration as well as access to texts and information services: "publication time is very much shorter; significant 'articles' may be as short as one sentence; cross-reference citations may easily be much more specific (i.e. pointing directly to a specific passage); catalogs and indexes can be accessed and searched online as well as in hard copy; and full-text retrieval with short delays is the basic operating mode. The end effect of these changes is a form of recorded dialogue whose impact and value has a dramatic qualitative difference over the traditional, hard-copy journal system."

Defragmentation offers the flexibility to create new packages that build on journal branding but are not dependent on it. Examples of this are the subject-focus portals such as BioMedNet and so-called

'virtual journals' promoted in many cases by these portal operators (McKiernan 2002). Virtual journals appear to offer new services to the user, notably personalisation, but in some cases are simply vehicles for pay-per-view: access remains tied to the original journal or publisher, the journal that performed peer review and most likely obtained exclusive rights to publish. At the journal level defragmentation is not new, and has been practiced by journal aggregators for many years.

### 5.4.1 Defragmentation: search and link

Utility, and the prospect of new tools to manage and organize the literature, is one of the primary drivers for defragmenting the journal literature: “for example, by providing useful new tools for searching and navigating this vast body of information, or by finding new and better ways to organize and interlink it, to track the development of ideas and new understanding, to identify errors of fact or interpretation, or to add new commentaries or syntheses.” (Eisen and Brown 2001)

The dominant mode of information retrieval on the Web is search. “Search is the most popular service on the Web because it fits the fundamental nature of the new medium: users choose where they want to go today.” (Nielsen 1998) Web search, especially site search, is not always effective: “Even if users can't find what they're looking for through well-labeled links, many companies offer site searching as a shortcut to information. But (a) study found that site searching was not only useless but also detrimental for information gathering: users were 50 percent more likely to find what they were looking for if they never hit the search button.” (Festa 1998)

This news story pre-dates the seminal Web search service, Google, which appeared in September of that year, or it may have reached a different conclusion. “Traditional search was based on finding all of the most relevant articles about the user's query. This approach worked well for scientists searching databases of research papers, but it has failed on the Web. We don't want *all* articles, and we don't even want the most "relevant" ones as determined by the number of times a certain keyword is used on a page. We want the *best* pages about a topic and the pages that are the best starting points for further hypertext navigation.” (Nielsen 1998)

A survey ranked Google the most effective search engine, with 97% of users saying they located what they were looking for ‘every time’ or ‘most of the time’.

Web links are the basis of Google, which ranks the results of a search by analysing the ‘back links’ pointing to a given Web site; in terms familiar to science researchers, it finds the most cited pages containing a search term. The accumulation of link data from millions of Web pages is a potent tool

for analysing the quality of sites that are linked *to*, as Garfield (1955) discovered in the case where the 'links' were the authored references in scientific papers: "by using authors' references we are in reality utilizing an army of indexers." Thus the quality of a site might be judged by how many links point to it; iteratively, the quality of the sites pointing to a given site are also assessed and rated in the search result.

Google is effectively defragmenting the Web, masking the inadequacies of much site-based navigation, but it doesn't provide deep access to papers on journal publisher sites, which continue to represent isolated 'islands' of information.

Defragmentation of the journal literature will instead depend on the accumulation of published papers into consistently tagged and searchable archives, such as Open Archives, and through services providing reference linking and citation analysis. Scholarly papers are not written for use in isolation, and the ubiquity of reference lists within science papers is one way in which this is represented.

Reference linking services first appeared in the areas of biology, physics, astronomy and cognitive science (Hitchcock 1997b). Astronomy provided an early example of reference linking between publishers and distributed sites with Urania. In 1998 astronomy had three principal electronic information resources: the electronic journals, the Astrophysics Data System (ADS), and the various astronomical Data Centers. The ADS presents the full-text pages of the major historical literature with the accompanying references and citation lists which are linked to the articles as well as to the current electronic journals. "Finally, the current electronic journals are linked to each other and to the ADS system of abstracts and full text page images, and the holdings of the data centers." (Boyce 1998) Urania was founded on a working system of common standards, naming conventions and cooperative protocols, an infrastructure recognizable in later initiatives such as OAi.

The real benefit of a defragmented science literature that follows from the collection of reference data is the ability to track the development of ideas forward in time as well as backwards. This is Garfield's citation indexing, described as an 'association of ideas' and bearing remarkable similarity with Bush's (1945) 'association of thoughts' which anticipated modern hypertext. By mapping both reference data and citation index data on to electronic texts in the form of links produces a qualitatively different information environment to those that preceded it: "When information about essentially all references in an article is available, with forward links to other articles, we will truly have added value to the paper form. Finally, when the entire text of a referenced article can be accessed by a mouse click, we will have reached a new level of information access." (Austin 1996)

Defragmentation is about building large collections of accessible resources – not physically bound, tied to single location, or exclusively owned – and adding new services. Before the journal digitization programme of the late 1990s there was a warning: “The segregation of e-journals into an electronic space that isn’t (yet) integrated into the scholarly document systems of libraries, indices, abstracting services, is a formula for continued marginality.” (Kling and Covi 1995) The warning holds true, but it is the established systems that risk becoming marginalized if the producers do not participate in the process of defragmentation.

## 5.5 Reshaping scholarly publishing: summary

Powerful forces are driving the growth in e-journals, clones or not, but the motivations for reshaping scholarly journals, by decoupling journal content from publishing process and defragmenting content services, are not universal. The prospects for progress depend on the perspectives of the three main players: publishers and other commercial suppliers, librarians, and the journal users. The primary motivations for e-journals are:

- Publishers: adding value to journals
- Librarians: improved information retrieval
- Users: faster, more direct access to information, and the ability to act on information

These motivations are not new, nor are they unique to e-journals, but e-journals serve these motivations better than other formats, and the continuing development of e-journals will be predicated on them.

Scholarly publishing and journals are set to change more than most of its participants are prepared to admit. One response to this view of decoupled and defragmented scholarly publishing is to argue that few people support it: not publishers (where are open access journals?), not libraries (where are open access services?), and not yet authors (where are the eprint archives?), but it would benefit all researchers who use the literature actively to inform and shape their research.

For those expecting change to be rapid, there are some significant cautions. These concern misconceptions about the likely impact of e-journals and the viability of e-journal models, and are typically predicated on an unchanging view of conventional journal models:

- Lower cost predictions for e-journals are unproven and insignificant: studies over 30 years reported by Tenopir and King (2000) show that it is hard to justify e-journals on economic arguments alone, especially if those e-journals mimic print journals in all important respects.
- Authors prefer established journals to new e-journals: studies indicate the continued preference of authors for established, i.e. print, journal titles (e.g. Anderson *et al*. 2001, Bjork and Turk 2000) despite findings that often these same people, as readers, want access to free Internet journals.
- Electronic publishing models are not universal: electronic publishing enthusiasts are warned against 'a one size fits all' approach for modelling electronic publishing across academic disciplines (Kling and McKim 2000).

Negroponte (1995) recognized the pervasive impact of digital technology: "In an open system we compete with our imagination, not with a lock and key." Successful online publishers will recognize that generating revenue will depend not on demanding ownership of the raw literature as now, but in imposing services that make sense of and improve the accessibility of this vast information resource.

Bauwens (1996) is more explicit: the first law of cyberspace is that "on the Internet the price of information will tend towards zero." Instead, he says, revenues will be generated by enhancing the information chain – 'from databases to decision bases', for example – and combining free access with for-pay value-added services. Just as the decoupled and defragmented publishing model described here anticipates. The information itself is given away in the quest for influence, much as academic authors do now, but the new twist is that publishers will be expected to do the same.

The impact of open information, as we can now see, has acute relevance for the scholarly publisher. While the price of scholarly papers published online may tend towards zero, thus liberating the literature, this model ensures that within the academic community it will not be valueless.

# 6 A new scholarly publishing model: the hypothesis

This chapter introduces something of a time warp, because the template for the model that became the basis of the implementation tested for the thesis was conceived some years ago. Previous chapters report developments in the general framework for electronic journal publishing to the present time. To understand all the features of the implementation, described in subsequent chapters, and some of the reactions to it, the chapter necessarily steps back in time to illustrate the original concepts as first elaborated in Hitchcock (1997). The chapter is structured to illustrate those ideas that are preserved in time, presented in italics, but also to allow brief assessment of the validity of both the publishing framework and the model with regard to subsequent developments described above.

## 6.1 Anticipating a radical three-track electronic publishing framework

*Senge's (1990) popular business book* The Fifth Discipline *identified five 'competent technologies', or disciplines, for successful corporations. While none of these disciplines was itself new, it was argued that only when all five disciplines converge within an organisation will its highest aspirations be reached. Similarly, the aspirations of the academic community could be satisfied by the proposed framework for electronic scholarly publishing, which has three* inter-dependent *components, but will only be effective when all three components are embraced in a unified model.*

*The purposes of the framework are threefold:*

1. *To provide unlimited access to the complete corpus of scholarly literature, because the fundamental purpose of publishing research results is to support communication, information and progression.*
2. *To enhance the presentation of scholars' work through: validation at definitive stages of the work, based on conventional methods of refereeing and supplemented by other methods such as open peer commentary; and design, production and editorial support.*
3. *To organise journal-like collections of papers mediated by link services.*

*The three elements of the framework that derive from this analysis are, as outlined in Fig. 6.1,*

- *An eprint archive*
- *Refereeing and editorial services*
- *Link publishing*

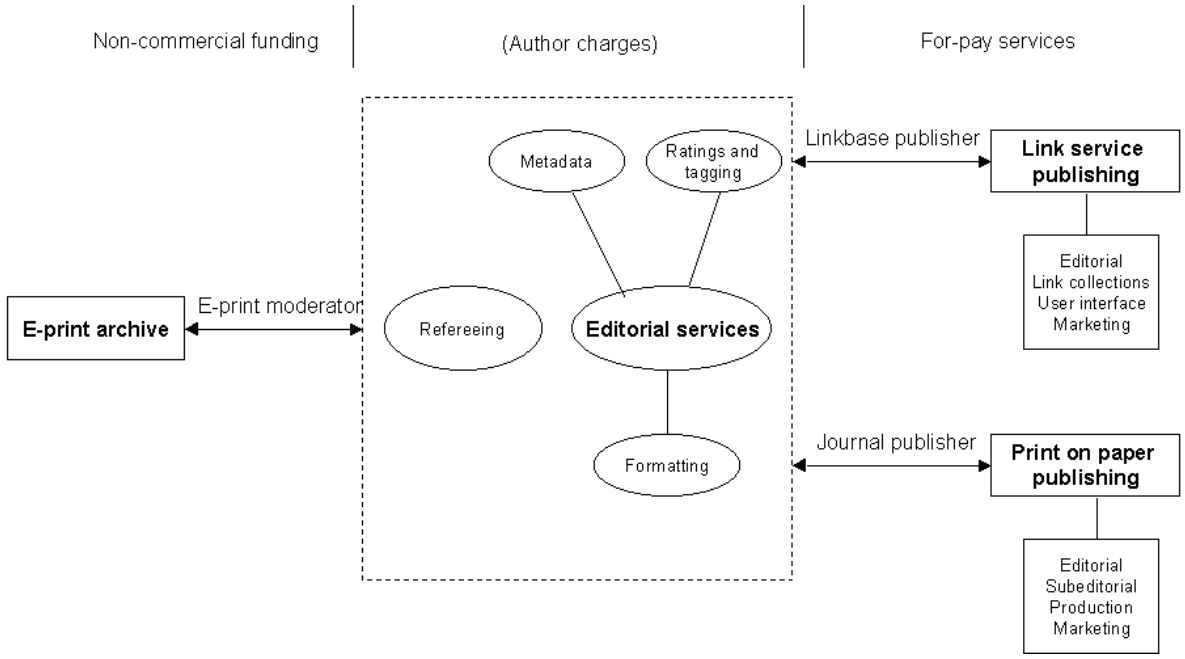



***Figure 6.1. Inter-dependent three-track publishing framework for online journals***

### 6.1.1 Update on the three-track publishing framework

This framework naturally differentiates the three critical requirements – access, organisation and presentation – of any scholarly publishing model. It also identifies how the different activities in the framework should be funded, by placing the obligation on the principal beneficiary of the respective services. Access to the unmoderated, or lightly moderated, literature is funded directly by the research community as a whole. This is the arXiv model. Linking and editorial services are value-adding publishing activities that should be funded by users.

Linking services on a large scale were envisaged as subscription-based products for individuals or libraries. Commercial reference linking services such a CrossRef (Pentz 2001a) have emerged since this framework was conceived, and it can be anticipated that standardization of OpenURL will see such services integrated with library management systems (Hellman 2001b). In this way the costs recovered for linking services are likely to be included either in subscriptions to journals or investment in library systems, rather than as products purchased directly.

It was anticipated that authors would pay for enhanced presentation and refereeing through page charges. In 2001 BioMed Central (BMC) announced a range of journals that would be funded in this way, but significantly is attempting to move the funding burden from authors to their institutions.

It is harder to assess the interdependence of the elements of the framework. There are still relatively few eprint archives of significant size (Krichel and Warner 2001a), nor a substantial OAi-mediated corpus of eprints, and little sign of these services encroaching on other activities of the framework. There is tentative interest from commercial publishers, however, in supplementing their role by building *pre*print archives – such as Elsevier's Chemistry Preprint Server (Bradley 2000), and Netprints, a collaboration between the BMJ Publishing Group and HighWire Press (Delamothe *et al*. 1999). If pursued more aggressively these could attempt to change the character of such archives (Guedon 2001).

Large-scale, disciplinary eprint archives provide effective data sources for reference linking services, as shown by the Open Citation project for example (Hitchcock *et al*. 2000). Smaller, or poorly focused, collections would produce too few links to be useful. Enhancements such as refereeing and editorial judgement will continue to be important in the new model, both as a filter and for organisational structure. In turn, the model anticipates refereed papers being added to the eprint corpus, either replacing or updating the original 'preprint', or pre-refereed version, as can be seen happening in arXiv. In most disciplines other than physics there remains a tension between commercial and open access services, between journal publishers and eprint archives, which could be reduced if payment to publish rather than to read was adopted on a much larger scale than BMC alone can achieve, and if journals were decoupled from publishing process, as argued in chapter 5.

Despite the appearance of fee-based reference linking services, it must be admitted that these are not exactly what was envisaged as the 'link service publishing' activity highlighted in the framework. The model also anticipated *editorial* links, in effect alternative perspectives based on an informed view of a specialist topic. This is the type of knowledge possessed by journal editors, and is why a journal-scale service was anticipated, rather than a publisher, multi-publisher or disciplinary-scale service that can be achieved with reference linking.

If Web interface researcher Jared Spool is correct, editorial links will produce rewards for users: "information essentially has a 'scent', and as users link from page to page they pick up the scent of the data they're searching for." (Koman 1998) Web design features such as navigational bars, left-hand sidebars, back/forward buttons, site maps, do not help the user, says Spool: "They all serve to break the scent, to take the user away from finding what they're looking for. Backing up to the home page is a sure sign that the user has lost the scent."

Reference links are authored links and can be automated once the source data has been extracted from the written papers. Reference linking is an important component of any linking framework, but

was not considered as part of the implementation described in chapter 7 because it was already being developed elsewhere. In contrast, there are relatively few examples of *editorial* linking, as discussed in chapter 8, and fewer examples in scholarly publishing.

There is another dimension to editorial linking. It was anticipated that branded collections, perhaps using established journal titles, might be mediated through such link services, fulfilling O'Reilly's vision of information products centred on information interfaces. The value of the collection will be determined by the quality of the links, i.e. the quality of the resources the links point to as well as the navigational cues these provide.

Given these different approaches to linking it hard to speculate at this stage whether widely-available linking services will be mediated via journals, either by CrossRef publishers or new models such as developed in this work, or by library services.

Despite the uncertainties, the incentive to adopt the framework remains the same as when it was proposed: the ability for *everyone* to *access* and to *act* on *every* item of information within the scholarly journal literature. Any model of scholarly publishing based on the historical structure of the paper journal, even if that model is transferred to the Web – primary, secondary information and everything else – as is happening now, will *always* break one of these criteria unless the existing library purchasing system has unlimited resources.

## 6.2 Anticipating the 'unbound' journal: the model

*To examine the thesis that an academic e-journal can be reconstructed as an overlay on open access papers and presented as a database and associated link service, the work will create a service 'unbound' from the traditional journal package but bounded by focusing on selected papers (Fig. 6.2).*

*Since the thesis reports developments in online journal publishing it seems appropriate that this should also be the focus of the journal to be constructed to demonstrate the model. Progress in Online Publishing (POP) will cover the history, social role and all aspects of the development of the traditional journal towards the online journal, the influence of new media, the form and impact of new content, new technology, new publishing models as well as issues affecting the digital library.*

*In effect, this will be an unbound work within a bound work, the conventional thesis.*

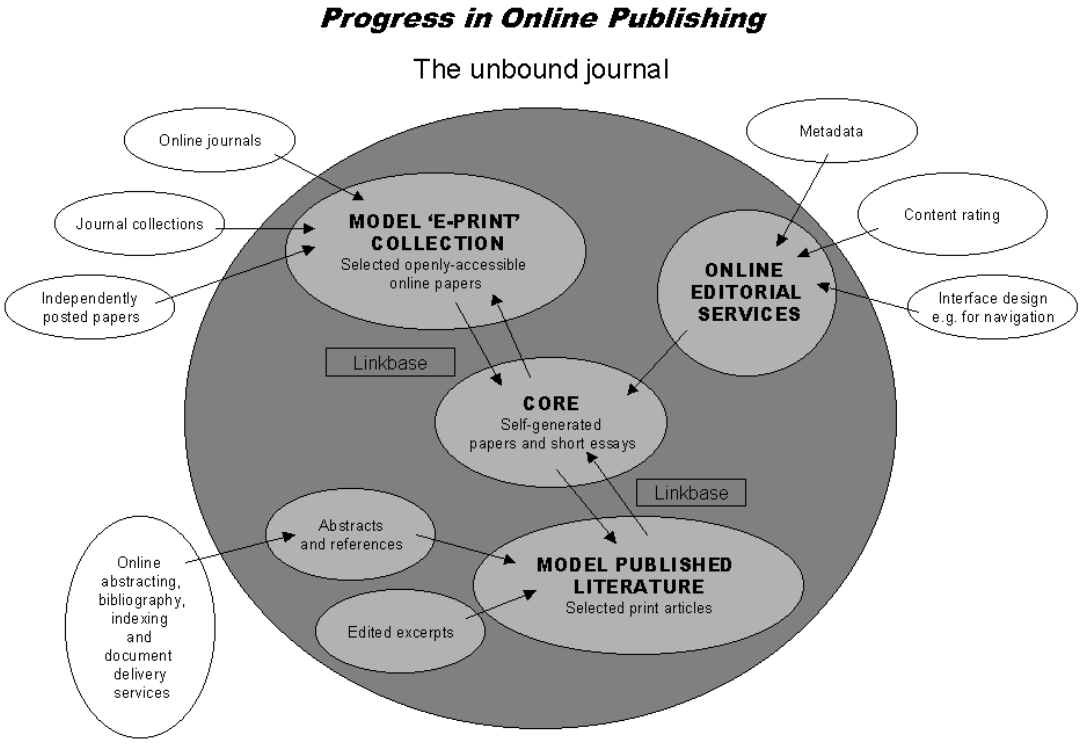



***Figure 6.2 The original model: a schematic of* Progress in Online Publishing**

*A large body of material relevant to the coverage of* PoP *is already available on the Web. In addition, development of The Journal will be proactive as well as reactive, and through announcements and existing contacts will seek to obtain advance notice of new material to be posted on the Web. The core of the journal will be a collection of state-of-the-art surveys, reports and essays by the 'editor'.*

*This, then, is the fundamental unit of the journal. Enhancements will include testing methods for developing the user interface: metadata such as the Dublin Core (DC) for building records; tagging and rating based on aspects of XML and PICS; and building the look and layout of the user interface, which might, for example, consider services such as the D3E Publisher's Toolkit from the Knowledge Media Institute at The Open University for automating frames-based presentations.*

*No rights will be sought over any contributed work, as is consistent with the model that in the scholarly world much of the raw content should be freely available.*

*The value-adding in the model will be created both by forming links across this distributed body of information and informing the links with editorial comment and original contributions. Both the links*

*and the comment are intended not just to allow users to access the works, but to provide them with a broad overview from which to pre-select materials to view.*

### 6.2.1 Updating the model

By analogy with Senge, what distinguishes the 'unbound' journal is the combination of all three elements from the publishing framework, which might be called the three 'competent technologies' of electronic journal publishing. Each element is a major activity practised elsewhere – but not all in one place – so the work considered each one in a rather looser form than would be the case within a formal publishing framework, but in a way that the interaction of the three parts can be judged.

Thus the aim was not to construct an eprint archive, but to consider an 'eprint' in this case to be any freely accessible, and relevant, work on the Web. The editorial services activity was not concerned with the subeditorial or text formatting role that is essential to the journal publisher, but instead concentrated on the role of editorial focus, selection and comment, the selected articles being held together by a collection of links made available as a linkbase.

The model was ambitious, and not every feature shown in Figure 6.2 was implemented, as the next chapter reveals. The name envisaged for the implementation – *Progress in Online Publishing* – is one noticeable change, although more substantive differences will become apparent, in many cases due to practical constraints and other developments in the intervening years,

At a broader level some of the main features in Figure 6.2 can be addressed here:

- Model 'e-print' collection and linkbases: the principal components of the implementation, described as database records.
- Core essays: just two essays produced, a commentary and a review, to test the format.
- Editorial services
    - Metadata: the database records are not described in a standard format such as DC. Given subsequent developments in standards for interoperability – the DC metadata element set (standardized in 2001), which can be described in RDF (1999) encoded in XML (v1.0, 1998); unqualified DC is the minimum requirement of the Open Archives metadata harvesting protocol (v1.1, 2001) – this could usefully be reviewed.
    - Content rating: usage of rating systems has grown rapidly to block rather than review sites. By the end of 2001 almost half of US public libraries had installed net filtering software to block prohibited sites (Oder 2002). There are no major examples of

rating being used to filter the content of scholarly sources, although it has often been suggested that simple tagging could differentiate refereed papers from non-refereed. The database records in the 'unbound' journal are not tagged for ratings purposes, but could be adapted if an accepted standard emerged.

    - Interface design: the emphasis of the design was on the data structure to give users as many views of the data, i.e. means of ordering and querying the data (Spink, *et al*. 1998), as possible. In terms of page presentation one fundamental decision was not to use HTML frames, so the D3E Toolkit, a frames-based interface, was not used.

- Model published literature: one of the strongest principles of the implementation was that it would only describe freely-available versions of selected papers. No attempt was made to model the print literature in the implementation, which is therefore a partial view of the literature. How partial it is can be judged by comparison with Bailey's scholarly publishing bibliography, which covers print and electronic sources with similar, if not identical, scope.

## 6.3 A new journal model: summary

Three critical elements of a framework for electronic journal publishing have been identified. The strength of the framework is its clarity of purpose for scholarly publishing and the way in which it identifies the financial support required. It challenges the existing system of scholarly publishing, not least by presupposing that electronic dissemination of scholarly literature will dominate. Each element of the framework is practiced elsewhere, and major examples have been cited. A model implementation has been proposed, arguing that an integrated approach, embracing all three elements of the framework, will create a new service that is qualitatively different and will offer benefits to users that the individual services cannot. Chapters 7 and 8 explore the implementation of the model in more detail, and chapter 9 examines the reactions of users.

# 7 A new scholarly publishing model: implementation

If the prospect of enhanced access is the motivation for new models of electronic scholarly communication and publication, the need is for a model that can deliver this rather than mimic existing forms.

Enhanced access means instant access to a comprehensive collection of scholarly papers, to services that inform users about the collection and help users to identify and explore new relationships between works quickly and efficiently, and which, most importantly, can always find a version of any cited paper that any user can access.

To fulfil this demand requires a series of interoperable services, from low-cost services that provide free access to papers, such as eprint archives, to conventional and expensive publishing services such as high-impact journals and bibliographic services. Intermediate are electronic-only services such as search and reference linking.

It could be argued that none of these elements is new; that enhanced access is already offered. But the picture is incomplete. Outside physics there are few eprint archives that provide organised access to papers, and publishing services are not interoperable, publishers preferring to build competing and overlapping services, adding to already high-cost products such as journals.

What is new is the concept of the universal access layer, based on which users and purchasing agents, such as libraries, can tailor packages based on individual services and preferences, an approach that would be both economic and efficient, underpinned by competition to provide useful services integrated with, but independent of, content rather than competition to own content.

This ‘layering’ approach to services is elaborated by Hellman (2001a): “The best way to add function is to do one function very efficiently, and through open standards allow other generic technologies and organizations to layer on added functionality.” Search services such as Google, and annotation services as were exemplified by Third Voice, are cited as examples: “for a scientist, exposing an e-journal to Google indexing adds more value at lower cost to an e-journal than almost ANY of the innovative functions” of most e-journals.

Similar visions have been advanced in unlikely places, by publishers in the medical field, a discipline that strongly resists access to unreviewed papers and therefore imposes higher costs on the universal access layer of papers. Yet the editors of the prestigious *British Medical Journal* predict a radical reshaping of journals based on non-exclusive access to databases of research papers (Berger and Smith 1999): “Medical journals that comprise mostly research articles (most of them) are almost certain to disappear. Instead research studies will be published on a huge electronic database. The primary job of the surviving journals will not be to publish research studies but rather to visit the database, scavenge the studies that are important for clinicians (a small minority), and present them in as sexy and appealing a form as they can manage.”

The science editor of *Nature* has speculated on a high added-value approach but one that again emphasises community, shared data and automated databases (Butler 2001a): “there will increasingly be sophisticated and novel forms of publications built around highly organized communities working off large, shared data sets. These hubs will stand out by their large investment in rich metadata and sophisticated databases. The future electronic landscape should see such high added-value hubs evolving as overlays to vast but largely automated literature archives and databases.”

In December 1998 the first record was added to a database built to examine the model described in chapter 6, and which supports the type of services that were later envisaged by these medical publishers. In this case the database contains some full papers, but manages a much larger collection of records describing selected papers stored elsewhere. The records also contain data used as the basis for services that add an organisational ‘overlay’ to the selected papers, which thus become accessible to all users of the service.

In a distinctive twist to the model envisaged by others, URLs stored for each paper described in the database are read by a link service. This service adds links – displayed as link graphics to differentiate them from conventionally authored, underlined and coloured, text links – to any database view or record displayed at the request of the user, *and* to the full texts, wherever they may be accessed on the Web. In this way users can, in principle, jump directly from record to paper, and from paper to paper, fulfilling Bush’s (1945) promise of letting users take unanticipated paths through the literature, changing direction as new findings dictate because any linked paper is immediately accessible.

This example implementation of the model is *Perspectives in Electronic Publishing* (PeP), the distributed journal of the title and the subject of the investigation of this thesis.

## 7.1 Perspectives in Electronic Publishing

Researchers need to act on information, much of which is in published papers. The selected full-text papers accessed using PeP are not contained on a single Web server, but come from many sources distributed on the Web, from which they are presented in their original form. This distributed collection is bound by the database and links. In this way PeP informs research, not just by providing access to papers but, most importantly, by enabling the researcher to find new relationships between these texts, what Nelson (1999) calls ‘deep re-use’.

PeP can be used by anyone with a Web connection and browser, optimally with Microsoft Internet Explorer (for Windows) v5+.x, and can be found at http://aims.ecs.soton.ac.uk/pep.nsf.

The papers described in PeP are all on the subject of electronic publishing. Many papers in this field, experience shows, are freely available from a wide and diverse range of sources, from online journals and conference proceedings to personal and institutional servers. This is a fertile environment for the organisational overlay that the model creates. PeP already links to content from over 130 distinct sources, not including personal or institutional servers, which would make the total substantially higher. In PeP, papers are editorially selected. Only freely accessible papers are selected.

For the purposes of this investigation the implementation was shaped by a number of practical constraints. PeP was intended to be a real, usable service, not a hypothetical model. To achieve this in a personal project the scope had to be tightly focused, and manageable. A journal-scale model is the ideal size. It appears somewhat incestuous that the subject of the model is also the subject of the thesis, but the chosen topic fits this requirement.

PeP covers all aspects of electronic networked publishing, with an emphasis on academic publishing and on journals. It covers the publishers, the publishing process and intermediary services. Coverage extends to research and technical development that will impact on publishing, including some aspects of digital libraries. Changes to the legal framework of publishing for the network environment are another important component.

As an experiment, in the version evaluated, URLs for the papers were not shown directly to users as part of the record but solely by means of *added* links.

A simple guided tour shows what the implementation can do (Hitchcock and Hall 2001). What follows is a brief description of the main features.

PeP is based on a database built using Lotus Notes. This is proprietary software, so the legacy of the project may be limited in terms of the design and general applicability of the database. A more open, XML-based format might have been better, but the tools to build it were not adequate at the time PeP began. The database is secondary to the principle, however, and Notes was chosen because it was shown to be effective in a related project, which had implemented a database and accompanying link service architecture (Hughes *et al*. 1999).

At the front page users are presented with a list of the papers most recently added to PeP. When PeP was first announced to evaluators early in 2001 the database described a few hundred papers. At least an equal number of known papers are still to be reviewed. Users can find specific papers using search, or browse using an index. This index is presented in a second browser window, for reasons

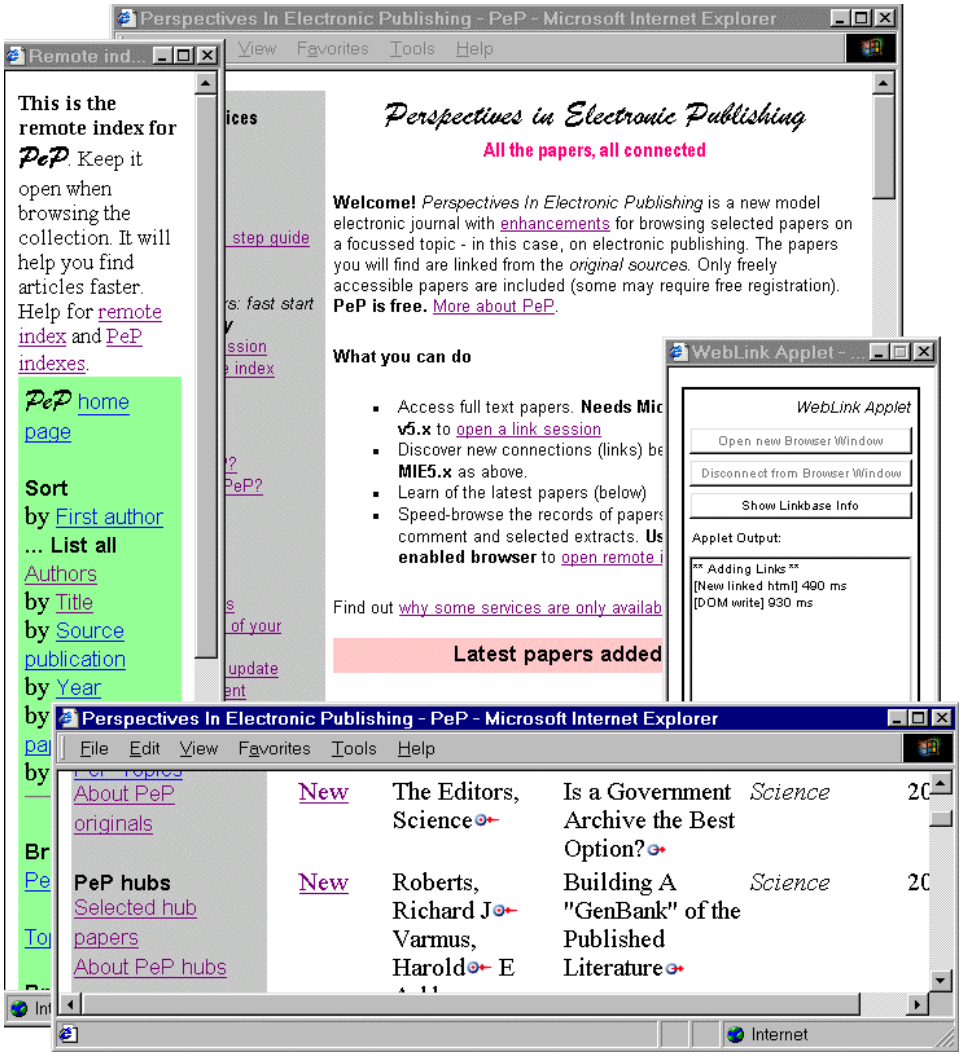



**Figure 7.1 Full featured PeP service, showing the original front page, the linked version of that page with clickable link graphics added (enlarged view), the remote index (left) and the WebLink Applet**

described below, so is referred to as the 'remote' index. Compared with a print journal the remote index was designed to act as the permanent ‘binding’ of the distributed journal papers.

As described so far PeP sounds like a simple library-like catalogue. This is transformed by requesting the added links. Technical constraints limit this part of the service to MIE5+.x users. For those users, the browser downloads an applet containing the link data. This WebLink applet interacts with pages subsequently downloaded by the user and attempts to insert links.

There are two types of added link:

- direct, pointing from PeP to the full-texts of the selected papers
- indirect, querying the PeP database for more information on the chosen subject, which may be an author name, or a key word or phrase

The specification and operation of the link service are described in chapter 8, revealing the reasons for the restriction on which browser can be used to view added links.

PeP’s collection of services is shown in Figure 7.1, with the link applet window and a second text browsing window. The remote index is open too.

Following the PeP or New link to the left of the listing in a contents view opens the PeP record for the chosen paper (Fig. 7.2). Alternatively, the graphical link inserted beside the title of the paper retrieves the full text (Fig. 7.3). Notice that although this paper is served from the originating site – it is not copied to PeP! – it also has added links. This is an example of a well-linked, or ‘hub’, paper. Hubs point to many other locations. PeP hubs aren't created hubs by their authors, but contain key terms specified in PeP that are common to other papers. PeP hubs are useful because they provide context for a larger number of papers and lead users to those papers.

The applet will try to add links to any page displayed in the linked browser window – it is not discriminating – but it is more likely to find text fragments matching link data, by definition, in a page with some relevance to electronic publishing and containing familiar e-publishing terms.

It now becomes clear why the index is remote from the main text window. It would be inappropriate and potentially misleading to frame pages from sources other than PeP. The first cases concerning the legality of Web linking – conventionally authored Web links, not the added links of this project and others – invariably resolved to ‘framing’, a presentational feature of Web browsers that allows

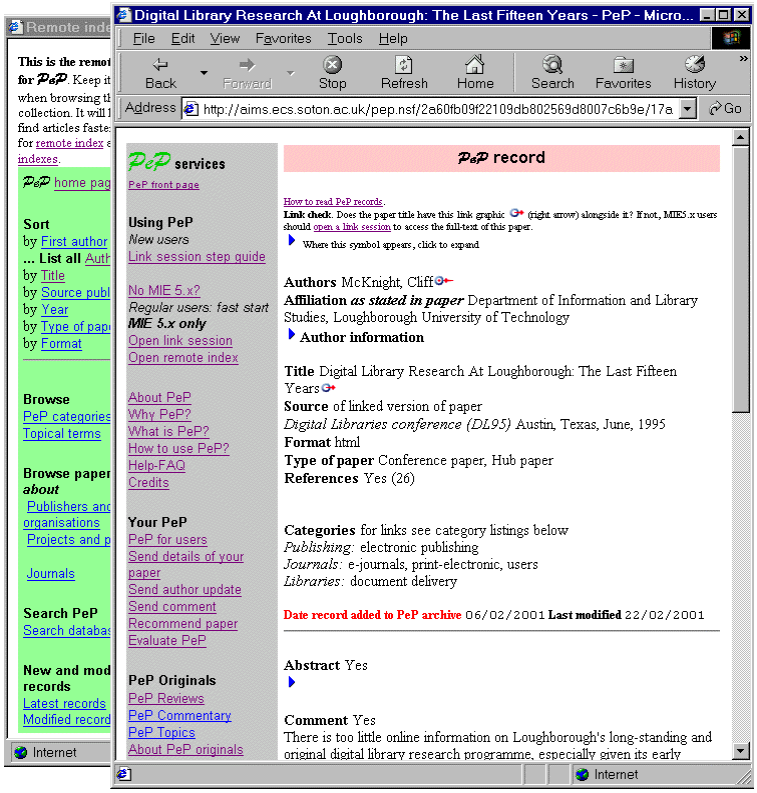



**Figure 7.2 A record for every paper entered in PeP, with bibliographic details, comment and notable extracts, and a link to the full text**

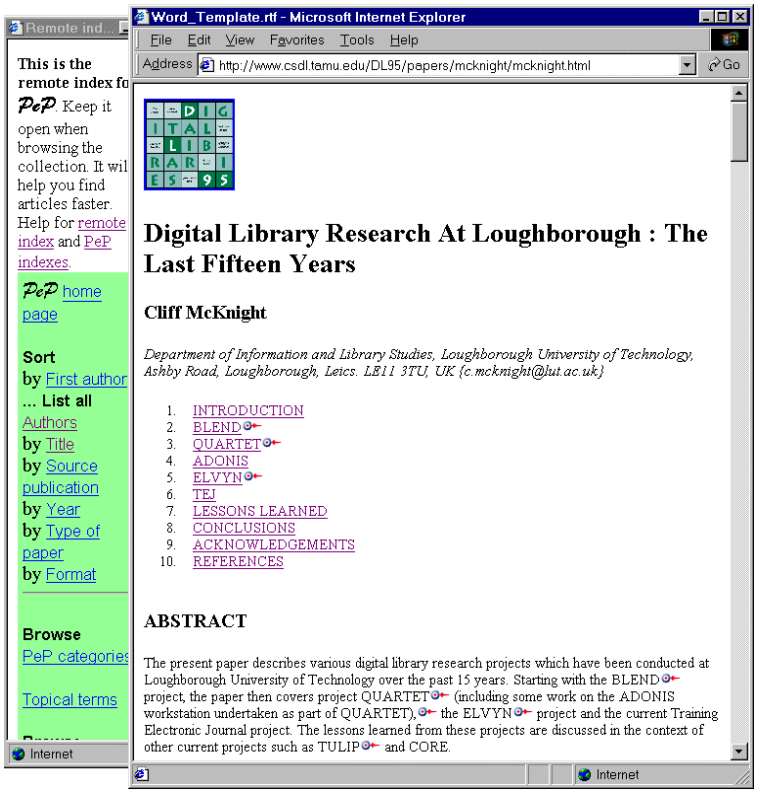



**Figure 7.3 Original full text with PeP links added, or return to PeP using the remote index (unlinked browser window and applet window both minimised)**

independent content segments to be arranged and displayed together in a single window. Misused, however, this can lead to a legally prohibited Web imitation of ‘passing off’, the presentation of content produced by others as though it were your own. PeP was designed to link to, but not ‘pass off’, original works; the identity of the original must not be subverted to the linking service. For authors, PeP does not knowingly point to sources that misrepresent a work or are unauthorised copies of a work, which continue to be accessible against the wishes of the author.

Ideally, the added links enable the user to explore the literature on a focused topic, one paper linked to another in a web-like form. For small or poorly focused applications, this is more likely to be experienced as a tree structure, using the index to return to the database when a Web path leads nowhere. No other service presents such relationships between the full texts of papers, so PeP is the first example to reveal the design requirements of a web-like rather than a tree structured link journal.

## 7.2 Design of PeP

### 7.2.1 The PeP record

McKiernan (2001) reviews the functionality of PeP as seen by the user, covering PeP services (Fig. 7.2, left-hand column in main window) and indexed views. To understand the design of PeP in more detail it is worth considering the form and content of the template for the generic PeP database record, including those fields that are not presented to the user.

The full record (Table 7.1) contains bibliographic details of papers, including a version history of each paper, and information on authors. Original comment, extracts from papers, feedback from authors and readers are all intended to help users assess papers prior to reading the full text.

The form of the record reveals a number of design decisions that seek to exploit the novel features implicit in the model, also to pre-empt anticipated criticisms. One restriction has been introduced deliberately for this project, and certain desirable refinements to the record have not been implemented (due to time constraints, technical limitations or insufficient knowledge). None of these are inherent limitations of the model as envisaged.

**Table 7.1 Template for a database record describing a paper linked from PeP**

| |
|---|
| Author details |
| **Authors** full names as given in the paper<br>**Affiliation** as stated in paper<br>* **Email** latest known<br>* **Author Web page** |
| About the paper, versioning information |
| Bibliographic data<br>**Title** [+subtitle]<br>**Access warning** e.g. registration, need to scroll first page, etc.<br>**Source** of linked version of paper [title] [data, V, N, pp., etc.] [year]<br>***Source url***<br>***url*** of paper<br>Describing the paper<br>**Format** HTML, PDF, etc.<br>**Type of paper** journal paper, conference paper, book chapter, viewpoint, etc.<br>**References** whether included or not, number of refs<br>Versioning information<br>**Publish-Ref** reference data for authoritative version, if different from linked source<br><br>**Other versions**<br>**Other title**<br>**Other source**<br>***Other url*** |
| Classifications for linking |
| Classification of papers<br>***Categories*** select from pre-defined category listings (view from remote index)<br>Data for automatic linking<br>Open choice<br>***Linkwords***<br>Linkwords selected from pre-defined lists (view from remote index)<br>***Topical terms***<br>***Publishers and organisations***<br>***Journal titles***<br>***Projects and products***<br><br>***Links*** whether or not the original paper has any authored links |
| Content |
| Edited content<br>* **Abstract**<br>**Comment by the editor**<br>* **Extracts**<br>* **Author update** response to comment, later work, etc.<br>* **Reader comment**<br>Automatically generated by database<br>**Date record added to PeP archive**<br>**Last modified** |

Key: bold, database fields; italics, not viewable by users

* shown at user request (requires an additional mouse click)

### 7.2.2 Principal design features

#### *7.2.2.1 Electronic-only*

PeP is an electronic-only product of an electronic-only information environment. The implications are profound, but have yet to be realized fully.

#### *7.2.2.2 Added links*

Works are fixed in time but added links can point to other works before or after that moment in time. By reading selected data from the PeP record for each paper – title of paper, authors, URL of linked paper, URL of source, linkwords (as described in chapter 8) – the PeP link service maps new connections between the works that are intended to add to and, in some cases activate, the authored connections, the references, pointing backwards in time.

#### *7.2.2.3 User forums*

Long expected to be a natural element of electronic publication, user forums are vital for a broadly based information service such as PeP. Discussion forums on the topic of electronic publishing have proliferated and thrived (for examples see section 7.5), but few conventional peer-reviewed journals have succeeded in establishing user debate about their contents, except those with a mission to do so, e.g. *Journal of Interactive Media in Education* (Buckingham Shum and Sumner 2001). At the journal level there are typically two reasons for the inability to establish active user forums:

1. Scope: few journals can present anything other than partial snapshots of their community and fail to offer enough context for debate that must necessarily extend beyond their contents.
2. Moderation: journals are highly formal and find it hard to reconcile the rigorous selectivity applied to primary content with open debate.

By definition the PeP model solves the problem of scope, but as a formal resource it struggles with the problem of moderation. It would be beyond the capability of most personal projects to build the community necessary for user forums to succeed, but the implementation of PeP has constructed simple forums and, through the process of evaluation reported below and in the subsequent correspondence with peers (see Appendix 8), can demonstrate some preliminary successes. Authors have contributed updates, corrections, and in some cases acknowledgements of limitations of the original versions of their papers. Readers have submitted assessments of services described in PeP papers, or queried specific points. Debates in other forums referring to specific papers covered in PeP have also been highlighted. Responses to user forums are linked from the PeP front page.

The next task is to identify the features that would improve the user forums. PeP has not resolved to remove all moderation from forums, nor what degree of moderation is appropriate, nor the mechanisms for implementing it. Almost certainly the forums need to be more open and immediate than has been achieved so far.

### 7.2.3 Further improvements

As with the user forums, it has not been possible fully to implement all desired features into PeP. Some new services that would enhance PeP are outlined below.

#### *7.2.3.1 Date granularity*

A drawback of the PeP front page is that it lists the latest papers *added to PeP*, which could be a paper from 2001 or 1945 since PeP can be retrospective, rather than the very latest papers. This is a limitation for users who might want to use PeP as a current awareness service. Both lists are needed, but PeP cannot properly list the very latest papers because the lowest common denominator in its chronological index is 'year', and this is not granular enough. With hindsight the Source subfields of the PeP record could be more granular, but the bigger problem is that across all sources covered by PeP there is little consistency in describing publication dates. How can a paper that appeared in a *month*, be ordered in a list with one described by a *quarter*, or with those that have simply *Vol. No.*, or those that have no data at all? This is something that requires consistency across sources.

#### *7.2.3.2 Versioning*

A characteristic feature of an electronic-only information environment is the regular posting of new versions of papers, most visibly in the preprint–postprint sequence. In a dynamic, electronic-only environment there can be multiple preprints; more significantly, the postprint does not have to be the final version. Some in the scholarly community will resist this, seeking to preserve the process of formal dissemination and publication of a fixed document. A new process of faster publication and continual updating can be a strength, however, of a model of electronic publication unbound from print constraints, allowing authors and readers to update the current context and relevance of the work, for example. Such a process requires vigilance and a new culture that has yet to emerge. A model like PeP can uniquely track the changes to these materials, and must commit to doing so. As currently implemented the versioning information stored in the records is rudimentary, but it requires more and better information from others, as well as improved management of the original sources.

#### *7.2.3.3 Are papers listed in PeP refereed or not?*

For many users a field indicating whether a paper has been formally refereed could be the most important feature of the database, allowing them to pre-select only refereed papers for display. It is not yet included because from the range of sources encountered in PeP it is rarely straightforward to identify a version of a paper unambiguously as peer reviewed. In a PeP record there are two indicators of the quality of a paper, the Publish-Ref field (details of the formally published source) and the editorial comment, but the prevailing culture of scholarly publication suggests that the 'refereed' tag will be necessary.

#### *7.2.3.4 Author names*

These are stored in a conventional database format: *surname, first name or initials,* as given. This presents two problems. The first is the classic equivalence problem, i.e. how can the database know whether 'Hall, W.' and 'Hall, Wendy' are the same person? The second is that where 'Hall, W.' may appear in database records, it is not a common way of writing author names, except perhaps in structured references, and so cannot be linked in pages not generated by the database using the text matching technique employed by the link service. A better approach would be to store author details, especially those details that pertain to the author rather than the paper, in a second database, in which connections and equivalences could be authored.

#### *7.2.3.5 Enhancing the link service*

The ability to link from author names in pages not generated from the database is limited, as explained above, but other fields intended for linking are not yet linked at all. These are the category links, which are too generic for automatic linking on any page. The simple work-around for this has been simply to point all category links to the categories view in the remote index, but this is too general a solution. As a result the category classifications have proved to be less used and less useful than intended. Another linking solution is needed. A solution might be to adapt the link service to include an additional link condition controlling the location and display of category links (a Microcosm specific link, as described in section 2.4).

#### *7.2.3.6 Associated eprint archive*

Given the number of papers linked to personal servers in PeP, an idea being considered is an eprint archive associated with PeP. This would be based on EPrints.org software, software developed at Southampton University for building author self-archived collections of papers, and thus would be

OAi-compatible. How attractive this would be to authors is unclear, and would depend ultimately on how the archive was positioned in the community. The scope of the possible archive has not been defined, but it would be wider than PeP. Any such archive needs to ensure the long-term maintenance and preservation of the archive if the service is to become attractive to authors. PeP would not be dependent on such an archive.

### 7.2.4 Hidden from users

In the evaluation version of PeP two items of data were removed from the version of the record displayed to users:

1. URL of the selected paper
2. Source URL

Both are presented to users as *added* links. Given the browser restrictions on delivering these links to users, this has since been rescinded for paper URLs, but these URLs stayed hidden throughout the evaluation to ensure a better chance of the link service being used.

### 7.2 5 Anticipating criticisms

As a new model PeP invites criticism, as the evaluation shows (see chapter 9). Some of the likely criticisms can be anticipated.

#### *7.2.5.1 Lack of peer review*

Selection of papers included in PeP is not based on traditional peer review. Some papers are linked from peer-reviewed sources, so peer review does not need to be duplicated. Instead selection is based on relevance and noteworthiness; the aim is to balance comprehensive coverage as far as possible with these criteria. This leaves open the question of whether to include only the best papers; noteworthiness implies including the bad too, with appropriate comment and qualification.

This leaves open the questions of whether PeP should incorporate all papers that fall within its scope, how that scope is defined, and how papers are judged to be in accordance with the scope. PeP was conceived as an editorially moderated rather than an automated service and so is unlike ResearchIndex, for example. If PeP were to be relaunched, the selection criteria would be the first issue on the agenda.

Compared with a traditional journal model, in an open information environment the rules on selection must change. Unlike a peer-reviewed journal, PeP is not deciding whether a work should be exposed

publicly or not – it is already – but whether additional exposure is warranted. With multiple sources and versions and the erosion of exclusivity in journal publishing, this is an issue that will extend beyond PeP. Preprints exposed before peer review, works that may be both good and bad, are being assimilated into the literature faster than traditional peer-reviewed publication can manage. Regardless of the perceived pre-eminence of peer review, this demands that the process of selection and comment becomes more public and adopts new, clearer terms of reference.

#### *7.2.5.2 Originality: is PeP parasitic?*

PeP could be accused of being parasitic on the content of other services, but its primary contribution, like many bibliographic services, is the collection and selection from multiple sources. Yet this purpose alone is insufficient. The collection has to be informed, and this is done with original comments on papers and the development of papers original to PeP. Natural formats for these papers are review and commentary, broad narratives that can exploit the features of PeP. Two exploratory examples, one for each format, have been written (Hitchcock 2001a and b) and can be accessed, with *added* PeP links, exclusively from PeP.

#### *7.2.5.3 Does PeP infringe copyright in selected works?*

Associated with the question of originality, PeP raises the issue of copyright infringement. At one level this is easy to dismiss. PeP links to, but does not copy, original works.

The addition of links to those works could be argued to infringe the author's moral rights (Oppenheim 1996) rather than copyright. Even here the issue is confused because the links are not added unalterably to the original source copy of the selected paper, but are added at an intermediate stage, in the user's browser, *at the user's request*. At the level of the user interface, the means of adding links may be criticised, but this serves to emphasise that the user makes a conscious choice to view the *added* links.

More credible would be the accusation of infringement over copying of abstracts and extracts in the PeP record. For this reason these sections are partly hidden in the record, requiring the user to request each one separately with an additional mouse click, thereby adding the requested part to the displayed record. Clearly, this manoeuvre is tentative rather than legally watertight, nor does it accord with the objectives of the model – better access to more information – but is consistent with the spirit of the model, which is to assist users to access and understand works which authors have chosen to make freely available. In this sense the implementation anticipates author sensitivities rather than copyright concerns. In terms of mouse clicks the abstract, extracts and linked full-text are each equidistant from

the PeP record, presenting the user with a reasonable choice between the author's presentation or the PeP presentation. Author contact details are similarly partially hidden to assuage possible author concerns. If any of these concerns are shown to be misplaced each of these items can be added to the record at first download without the need for additional clicks.

## 7.3 What is PeP: catalogue, portal or journal?

PeP has already been compared with a library catalogue. By adding links, what has PeP been transformed to? Now it has features in common with abstracting and indexing services such as ISI's Web of Science, and with resource discovery services such as portals or subject gateways. Compared with these services, PeP has one major distinction: the user has immediate access to the full text of all papers listed without the authorization to use any other service.

PeP was designed to be a journal in a very traditional sense:

> *Defn.* **journal**
> *a record of current transactions; an account of day-to-day events: a record of experiences, ideas, or reflections kept regularly for private use: a record of transactions kept by a deliberative or legislative body.*

The journal analogy echoes remarks by Okerson (1991), reproduced in section 2.11. PeP has a number of features that reinforce the claim that it is a journal:

- PeP gives access to full-text papers
- All papers linked from PeP have been *critically evaluated*
- PeP has original content
- PeP expresses a consistent editorial viewpoint in its selection and commentary
- PeP is bound by an index and links

There are obvious differences with conventional journals:

1. PeP does not control or own content
2. It has no contact with authors prior to a work first appearing somewhere on the Web and then being included in PeP

In this context PeP is perhaps most closely aligned with review journals, but even where such journals link directly to the papers reviewed, there is no guarantee the user has access permissions.

### 7.3.1 What PeP is not

The two principal components of an open scholarly publishing system elaborated by the September98-Forum are free access managed by eprint archives, and peer review organised by recognised journals. PeP is not an eprint archive, nor is it a peer-reviewed journal.

In practice there are too few eprint archives, and none in the area covered by PeP. While there are some good e-journals covered by PeP, most of the relevant journals established in print form continue to restrict access. So PeP fills a role as an intermediate model. Unlike the September98-Forum model, PeP does not depend on the participation of other journals but on the willingness of authors to make their work freely available. It is evident that while authors continue to want peer-reviewed publication, an increasing number are prepared to act individually, making works available from personal servers.

For many users the lack of peer review may disqualify PeP as a journal. Okerson alluded to the subversion of today's 'fancy' scholarly journal, which is held hostage by an instinctive but inflexible emphasis on peer review.

As with many new Web applications, what PeP was designed to be and what it proves to be useful for may be different. The evaluation of PeP examines this issue and tries to determine whether it matters what label is attached to the service (chapter 9).

## 7.4 Separating model from application

PeP is the application, not the model. The two criteria that distinguish the model on which PeP is based are subject focus, and access. Specific features of PeP – the focus on electronic publishing, the condition of freely accessible papers, papers reviewed but not formally refereed – are not mandated in the model. The model does not preclude peer review, for example, but this implementation does not use it.

Another application could be based on commercial journal papers, with appropriate agreements, including only refereed papers, say. Yet this application would be intolerably compromised. Despite the best efforts of journal publishers – site licences, virtual journals, journal aggregators, CrossRef,

etc. (see chapter 4) – the journals industry is manifestly unable to achieve breadth and access together, because the fee-based journal structure fragments access (as discussed in chapter 5). Ultimately, for this application to be effective the researcher would need the combined subscriptions to cover all sources that he or she may conceivably need to access.

The ideal platform for the PeP model is distributed Open Archives-compliant eprint archives. OAi aspires to encompass broader types of materials that might be found in digital libraries, but eprint archives are the ideal foundation to encourage competition for compelling PeP clones.

## 7.5 PeP sources: towards a coherent literature on electronic publishing

PeP could equally well cover other topics, but focusing on electronic publishing works well in this framework because it is a discursive topic, amenable to some informality and not dependent on formal peer review. It is also interdisciplinary, appealing to a broad constituency with widely varying degrees of commitment. Every researcher has a stake in the effectiveness of publishing as a communication channel for his or her work, more so at times of change like the present. So while relatively few researchers would see electronic publishing as their primary interest, many will contribute intermittently, and when they do they will usually address their own peers, not always the broader community. In other words it is a highly fragmented literature that PeP can act on to the benefit of researchers.

PeP fills a curious role, both cooperative and competitive, within a hybrid online literature on electronic publishing, which includes open access journals, author-posted papers, Web catalogues of papers from print journals, alerting services and discussion lists.

Open access journals that cover e-publishing range from *D-Lib* and *Issues in Science and Technology Librarianship* about the role of the library, to general titles such as *The Journal of Electronic Publishing* (JEP) and *First Monday* [1]. While open distribution is innovative, the publishing strategy demonstrated by these e-journals fundamentally mimics print journals – a finite, periodically

[1] URLs for source e-journals well covered in PeP:

*D-Lib* http://www.dlib.org/

*The Journal of Electronic Publishing* http://www.press.umich.edu/jep/

*First Monday* http://www.firstmonday.dk/

*Issues in Science and Technology Librarianship* http://www.library.ucsb.edu/istl/

published collection of papers – rather than reaching out into the available literature. Integration between these and related titles and author-posted papers is as limited as that between print titles.

In other words, even on the Web with its facility for cross-linking, a journal can apparently only bestow recognition on collected works maintained on its servers, while other papers distributed elsewhere on the Web, remain unexploited. This omission is exploited by PeP, however, which maps the content of these journals and many other sources into a more coherent collection.

In its attempt to provide this coherence to the literature on scholarly publishing, PeP effectively competes with Web journal 'catalogues' [2] (e.g. Library-Oriented Lists and Electronic Serials; and in the UK, BUBL at the University of Bath) and selective dissemination of information (SDI) alerting services. The experimental *Index Morganagus* seeks to improve access to a selected array of freely available and relevant journal titles by collecting content lists and providing search facilities at a single site. While not directly comparable with these and other more general bibliographic services, it is against these services that PeP must justify its contribution.

At the article level [3], traditional indexing and abstracting services typically offer delivery of full-text papers from e-journals (e.g. BIDS provides access to the ingentaJournals, a full text service), although indexing of papers in electronic-only journals themselves can be patchy or non-existent (Cameron 1997, Crawford 2001, Jacsó 2001). For papers relating specifically to scholarly publishing there is an alerting service (*Current Cites*). These services do not typically provide direct access to the full papers, however. Another approach is Bailey's comprehensive and regularly updated bibliography on scholarly electronic publishing, supplemented with a daily news Weblog, both of which link to papers available online, but also cover print sources. Useful though these services are, they must still be viewed as potentially cooperating fragments of the whole online picture.

---

[2] URLs for journal 'catalogues' and services

Library-Oriented Lists and Electronic Serials http://www.wrlc.org/liblists/

BUBL http://bubl.ac.uk/journals/

*Index Morganagus* http://sunsite.berkeley.edu/%7eemorgan/morganagus/

[3] URLs for article services

BIDS Web search services http://www.bids.ac.uk/

*Current Cites* http://sunsite.berkeley.edu/CurrentCites/

Bailey's *Scholarly Electronic Publishing Bibliography* http://info.lib.uh.edu/sepb/sepb.html

Less formally, scholarly publishing is discussed on numerous email discussion lists (e.g. arl-ejournal, lis-e-journals, VPIEJ) most of which are available as threaded (by date, subject, author), searchable Web archives. Kovacs maintains a directory of scholarly and professional ‘e-conferences’. [4]

## 7.6 Comparison with other link publishing models: PeP's Southampton ancestry

PeP was inspired by more visionary systems, such as Memex (Bush 1945) and Xanadu (Nelson 1987), which were never fully realised. The real model on which PeP was based is Microcosm, a multimedia linking system also built at Southampton University (Davis *et al*. 1992). This is ironic because Microcosm was never a Web application. Its approach to linking has been adapted for the Web in various guises, notably as the Distributed Link Service (Carr *et al*. 1995), but these were effectively linking engines that discarded the editor and user interface in favour of a programming based approach. PeP re-invents the Microcosm interface for the Web and, although PeP also adopts the Microcosm 'generic' (link everywhere) link type, it does not use any part of Microcosm.

Microcosm was among the first systems to show that within digital information systems links can be managed as separate entities from other information content, referred to as ‘open’ linking. Long heralded by hypertext developers, this provided the capability to interconnect materials created independently and in different media. Demonstrating this, Microcosm was used to present two self-contained, intra-linked text collections that can be seen as ancestors of PeP:

- *Caerdroia*: the journal of mazes and labyrinths (Warren 1995)
- *Software Teaching of Modular Physics* (StoMP): a text book on a CD (Bacon 1994)

The potential for open linking has become more evident with the emergence of networked information services such as the Web where sites are frequently presented as self-contained 'islands' of information rather than interconnected webs. The first trial implementation of the pre-PeP model used Microcosm in conjunction with the Universal Viewer (UV) (Davis *et al*. 1994). The UV used the mechanism of the Windows clipboard to apply Microcosm linking actions to viewers, such as

[4] Addresses for discussion lists and URLs for list archives
arl-ejournal@cni.org, http://www.cni.org/Hforums/arl-ejournal/
lis-e-journals@jiscmail.ac.uk, http://www.jiscmail.ac.uk/lists/lis-e-journals.html
vpiej-l@vtvm1.cc.vt.edu, http://vega.lib.vt.edu/ejournals/vpiej-l.html
*Directory of Scholarly and Professional E-Conferences* http://www.kovacs.com/directory.html

Web pages in browsers, which were not supported in Microcosm. This Windows-only approach did not survive the switch to newer versions of the operating system.

An alternative, platform-independent approach to linking was attempted for another of PeP's predecessors, the Open Journal project. Based on the Distributed Link Service (DLS), Open Journal links were added by a Web proxy service. While effective, and a simpler way to maintain the link service, the proxy offered little support for editor or user control of links and could only be accessed by resetting browser preferences, a serious limitation in shared and managed computing networks such as within libraries and large corporations (Hitchcock *et al*. 1998b). PeP's commitment to controlled open linking continues to constrain it to platform-dependent solutions.

Table 7.2 compares PeP with the Open Journal project and other Southampton link publishing projects. With its ambitions to remodel the journal PeP has more in common with the Open Journal project than the earlier Microcosm-linked publications, and was similarly based on distributed resources. Open Journals were effectively unbounded in scope, however, and informed by the difficulties this created it was determined that PeP would be more focused.

There is another difference. Open Journals used journal contents supplied by publishers. PeP may be platform dependent but, in terms of the content it points to, PeP is publisher-independent.

The most successful application of Open Journals was reference linking, The successor to that project, the Open Citation (OpCit) project, takes reference linking to a new scale in terms of the

**Table 7.2 PeP in the timeline and context of Southampton link publishing projects**

| **Timeline** | **Project** | **Linking tool** | **Links documents from distributed independent Web sites?** | **Model/service** | **Scope** |
|---|---|---|---|---|---|
| 1994 | *StoMP* | Microcosm | No | Textbook | Physics |
| 1995 | *Caerdroia* | Microcosm | No | Journal | Specialised |
| 1995-98 | Open Journal project | DLS proxy | Yes | Journal | By discipline |
| 1999-2002 | Open Citation project | Citation database | Yes | Reference linking | Disciplinary (multi-disciplinary) |
| **2001** | ***Perspectives in Electronic Publishing*** | **DLS-like client** | **Yes** | **Journal** | **Focussed** |
| 2001 | OntoPortal | Ontology linking | Yes | Portal | Focussed |

number of linked documents (Hitchcock *et al*. 2000). Unlike PeP, OpCit doesn’t define scope, or seek to emulate journal-like functions; it is simply a service. OpCit supports reference linking and citation analysis for eprint archives, arXiv in particular. OpCit demonstrators have so far avoided platform dependence because all the linked content is managed locally, which is not an option for PeP.

Even though OpCit is targetted at much larger document collections, it would be interesting to build an OpCit service for a focused collection such as PeP, which should in principle be quite strongly self-citing, i.e. many documents linked by PeP are likely to cite other documents linked by PeP.

A new approach to linking is apparent in the OntoPortal project, which uses an ontology to model the concepts and complex relations used within a particular community to inform the linking between resources (Kampa *et al*. 2001). Like PeP, OntoPortal projects links over related but unlinked Web resources, in this case including formal papers as well as less formal resources such as the home pages of projects, institutions and individual researchers – hence the portal description.

## 7.7 A dynamic and integrated e-journal

It is now possible to assess whether PeP realises the features, proclaimed in section 1.1, of a new system for electronic scholarly communication (Roberts *et al*. 2001): is it dynamic and integrated?

Links and search are the computational tools provided to support the online user. Raney (1998) argues that citation and search should be seen as tools, not alternatives to journals, but in the design of PeP they are integral.

PeP is a primary example of decoupling journal processing tasks. As currently structured, it can link and present papers in an editorially-informed context with much lower cost overhead than conventional journals, because it doesn't perform all the functions – refereeing, editing and layout, etc. – of those journals. PeP supplements linked papers with its own original content in the form of review articles and commentary, and adds additional services for the user that aren't available elsewhere, the link 'perspectives' of the title. This is a manifestation of Hellman's (2001a) call to allow layers of functionality to be added to journals.

PeP has limited scope but unlimited space, which it can in principle fill faster than a conventional peer-reviewed journal. This reinforces integration – making something accessible from something else – which becomes more powerful as the number of selected papers grows.

By linking content from over 150 sources, PeP exploits the defragmentation of access to many diverse sources. While PeP is selective, thereby recreating a fragment from the whole of the freely accessible literature, it does not own the selected works and therefore does not compromise the benefits of defragmentation for other services.

PeP adapts to the dynamic of the online environment without relinquishing the editorial coherence of the conventional journal in favour of wholesale automation.

## 7.8 Implementing a new scholarly publishing model: summary

The lesson of the Web is that scholarly papers do not need to be owned or copied to benefit from the emerging scholarly publishing framework as long as they can be accessed easily. Authors, publishers, service providers and librarians need to re-think how collectively they can add to the value of these resources for users. PeP's contribution is a new approach that filters resources by plain editorial judgement backed by links and support for navigation.

Although relatively small in terms of content and data (Table 4.1), PeP is an original and complex amalgam of features and information that, when viewed by the user, may prove to be the right or wrong balance of features, and the right or wrong service for the user. User reactions will be determined by prior expectations, led by what PeP is perceived to be, but this is not straightforward. PeP was designed to be used, but also to be critically examined to determine which features might be valuable for future models and implementations. These features will be discovered in chapter 9, after more about another key feature, the PeP link service (chapter 8).

With no commercial interests to serve, PeP is adaptable if almost certainly not a universal model. In analysing the main features of PeP, this chapter has shown what to look for in topics that might be amenable to similar treatment.

PeP anticipates improved access to papers in many new and possibly confusing contexts, and that may be contentious, as might its viewpoint, and its relationship with authors and users. Otherwise PeP has a simple agenda against which to be judged: does it inform and improve research?

# 8 Links in PeP

*Added* PeP links were specified to be simple but editorially controlled, recognising that integration based on links needs to be focussed on content. An early motivation for this approach was O'Reilly's (1996) 'information interfaces'. O'Reilly urges publishers to reinforce the fundamentals of the Internet: 'participation, access, communication'.

Link services, which add links to documents when viewed by users, can be considered a special case of annotations services. Services that enable annotations – personal notes, links, etc. – of electronic documents by users or other third parties have a history that pre-dates the Web. The Web has made it possible to deliver such services to large numbers of users. The evidence suggests, however, that the nearer these services get to large-scale public or commercial use, the more controversial they become. The most striking examples have been Microsoft's planned inclusion of Smart Tags, a linking service, in the Windows operating system, and Third Voice, a true annotations service.

There seem to be two points of contention with these services:

- Increased complexity of the user interface
- The appearance of interfering with a document produced by another author

Most outrage is reserved for the latter feature, especially by authors, content producers and publishers. This is a curiously uninformed reaction, since the whole point of electronic communication is that it is interactive; to put it less glibly, it allows users more choices about how they manage and respond to received data. The Web browser itself allows users to alter presentation, from window sizing and text flow to text formatting. Producers use style sheets, or in some cases page-like formats, to reassert a degree of control over presentation, but users can ignore these too (see section 2.7).

The model developed and evaluated in this thesis uses a link service, similar in principle to Smart Tags but quite different in motivation. Even before it was implemented, Smart Tags attracted condemnation from mainstream media (Mossberg 2001), most notably from producers who felt that superimposing links would violate their rights. Users were concerned that tags would be imposed on their browsing experience. It was announced that the tags would by default be switched on in newly installed systems; the option to switch off would no doubt be well hidden, it could be speculated. Microsoft has so far postponed introducing the service.

Experience of applying a link service to a journal- or textbook-scale collection of documents has produced some promising results (Hitchcock *et al*. 1998b, Bacon 1994), but beyond this scale results have been mixed, as link management becomes more complex and the expectations of larger user groups become more difficult to anticipate. Smart Tags are envisaged on the scale of the Web. Indiscriminate linking on such a large scale has just one motivation: advertising. In contrast to Smart Tags, PeP returns the focus of linking to a narrow speciality, to a scale compatible with editorial control and integrity and audience definition. PeP links are based on highly specified terms: linkwords, tailored for the application rather than published keywords designed for other thesaurus-like bibliographic classification systems.

Third Voice offered a user-driven service, enabling Post-it note-style annotation of Web documents. These annotations could be created and viewed only by subscribers to the service, who had to install software on their machine to do so. Third Voice suffered critical reaction at launch, adapted and then died. The service was caricatured as an unregulated forum for libel. As with many commercial Internet services, the business model was not sustainable, but there is no doubt the initial reception damaged the service and the principle (Margolis and Resnick 1999).

Ironically, although not implemented in PeP, a Third Voice-like annotations service has been advocated for one feature – extracts. The PeP editor selects extracts, but a better approach might be to allow users to annotate documents they read, as they might typically annotate a printed copy. Such annotations could be stored for personal use, or shared. The lesson of Third Voice is that implementation and access to such data has to be managed carefully within a defined community.

Even without user annotations, by adopting a link service PeP is likely to prove controversial. Before discovering user reactions to the link service, this chapter explores the motivations for including a link service in the proposed model in the context of other Web-based linking initiatives. The specification is described, but not the detailed implementation and design.

## 8.1 Why a link service in PeP?

It would have been simpler and less contentious, but much less interesting, to have built PeP without a link service. Despite the evidence of setbacks for other applications, there is a conviction in this thesis that more effective ways of interconnecting the literature, especially the research literature, are not just desirable but necessary.

That conviction is founded on the visionary systems described by Bush (1945), Nelson and, most particularly because of its direct comparisons with a journal system, by Engelbart (1975). It is also based on personal experience of the purpose of reading for research and how the process of reading changes in an online environment.

There are typically two types of online reader: those seeking knowledge, awareness or pleasure, and those looking for specific answers, browsing readers and directed readers, respectively. In an online environment both types of reader become more agitated, less patient, less processional, even with linear documents (Nielsen 1997, Ubell 1997). Users are immediately more aware of the breadth of the online information space, as the boundaries of individual documents disappear. Papers are printed not just for the comfort of reading but to rebuild those boundaries. In the electronic space, however, the user is compelled to explore more widely. The directed reader is greedy for links that might lead to the desired conclusion. The browsing reader is grateful for the scope to explore, but both users will eventually tire, frustrated if they have not reached an end point or a satisfying boundary. Links establish relationships between texts; added links from a link service should also establish a boundary for these readers.

Researchers read a paper to gain an appreciation of the content of that paper, hoping to understand the views of the author and discover things that are new, to them at least. Reading alone does not create anything new, but creates a wider awareness so that what is truly new can be recognised. New insights come from connections we make with other works. Improved access to all scholarly papers will enable researchers to mine these connections more thoroughly, to return to papers looking for specific features they may have been alerted to elsewhere, or to find evidence to support new theories. There is nothing new in this, except when open access makes all papers available on demand the real nuggets of discovery will require the researcher digs more deeply, using appropriate tools to examine details and connections more forensically. Link services enhance the information environment when reading, knowledge and appreciation are not enough (Carr *et al*. 2000).

## 8.2 Specification of the PeP link service

The aim of the PeP link service became not to author new links but to try and expose those implicit in the selected texts. It was not the objective of the project to re-invent the link service, but to use tools already available. What was to be new was the specification of the information environment in which the link service was intended to work, and ultimately to give both editor and user some control over the presentation and number of links. In practice, this did not succeed. Two tools immediately available in the laboratory at Southampton were the Distributed Link Service (Carr *et al*. 1995) and a

commercial derivative, Webcosm. Neither could fulfil the requirements of the PeP application, nor could they be easily adapted. This was surprising because PeP was modelled conceptually on earlier linking applications using these tools and appeared to have simpler, if more specific, requirements.

Instead Tim Miles-Board, directed by Les Carr, built a link service for PeP. In the implementation some editorial control was re-introduced compared with the earlier link services, but not yet user control of links. The design is not described in detail here, just the requirements specified for linking within PeP, which were to:

1. Collect link data automatically from records in the PeP Notes database;
2. Add links to documents requested by the user, from wherever they may be delivered (the classic link service);
3. Support links pointing to external documents using specified URLs (direct links) or links that would query the PeP database and format the results (indirect links);
4. Support discrimination in link presentation based on location (by type of document, case sensitivity, and the number of instances of a term to be linked in a single document);
5. Create an interface for editor control of the link service.

Figures 7.1–7.3 illustrate the implementation of features 2 and 3 above. The resulting editor interface, demonstrating the inclusion of feature 4, is shown in Figure 8.1. Activating the ‘create’ button instructs the link service to write a structured XML document containing data from the specified PeP fields, with markup describing the fields and the conditions set in the editor interface (Fig. 8.2). This XML document is the link database, or linkbase.

The operation, limitations (browser-specific) and apparent idiosyncrasies (multiple windows) of the PeP link service, as remarked in chapter 7, can now be explained. According to Carr: “We have experimented with various ways, including (for a while) proxy-based rewriting of the HTML. The method which we prefer at the moment is to use the browser's own copy of the document object model (DOM), traversing it and altering it where necessary.”

In the PeP implementation “a control panel starts up an IE browser, and gets the innerHTML of the entire document once it has finished loading. It then adds links to the HTML text and resets the innerHTML when it has finished. You can generalise the approach, or do the controlling from different environments. We have used java or javascript in separate browser frames as an alternative. All have their own drawbacks: usually altering the DOM from within the browser falls foul of one security model or another. That is why it is better to have an external control panel.”

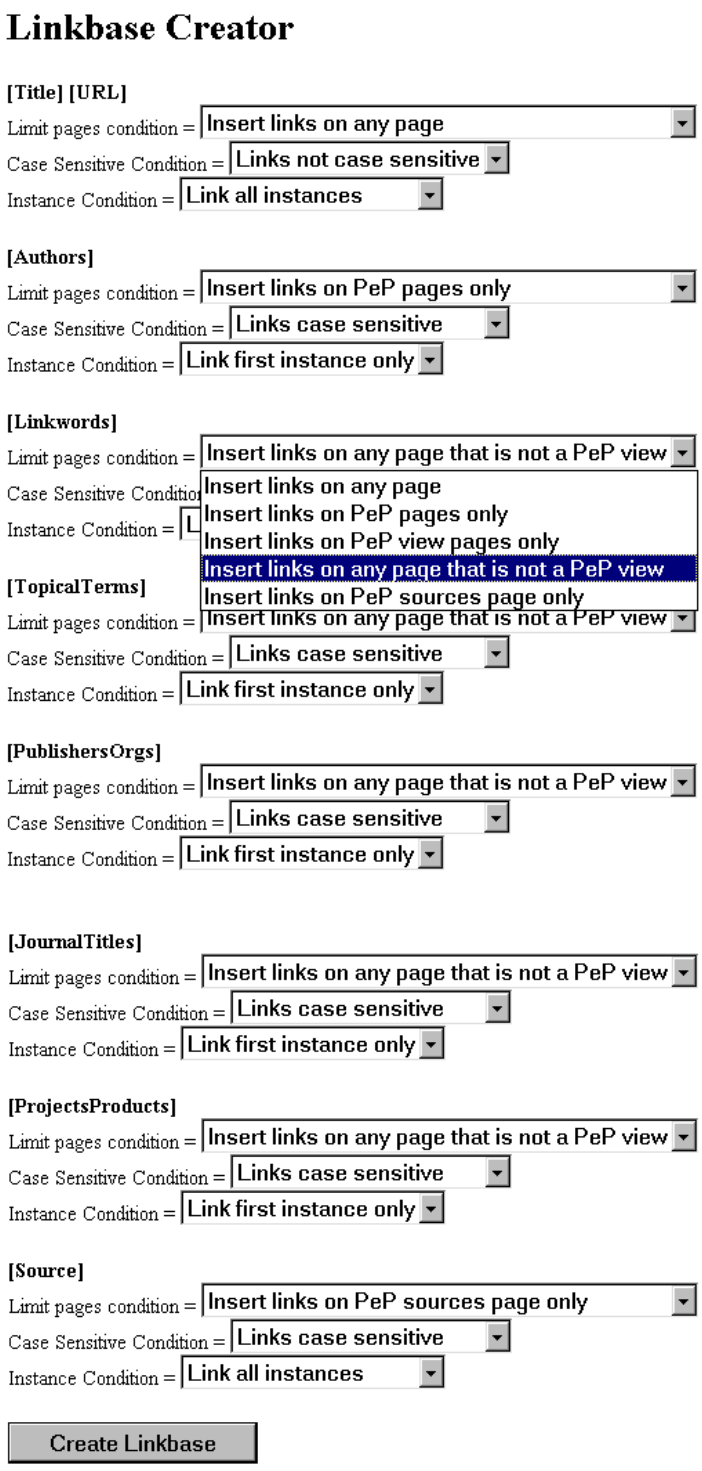
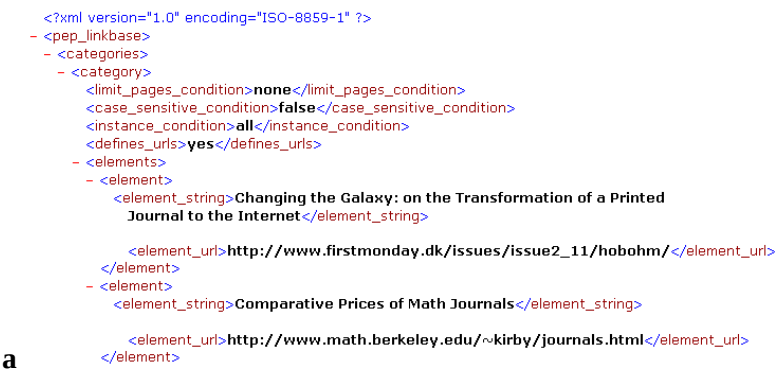



**Figure 8.1 Editor's control interface for the PeP linkbase creator, showing control parameters for data extracted from the [Title], [Authors], etc., fields in the PeP database record**

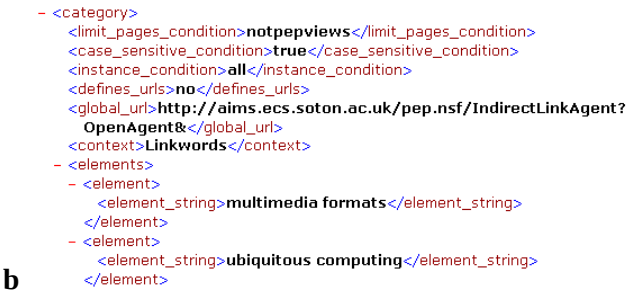
## 8.3 How the linkbase works

Each document described by a PeP record (Table 7.1) contains locations for the document (**url**) and its source (**source url**). The [Title] and [Source] fields indicated in Figure 8.1 are read by the linkbase creator and associated with these locations, respectively. These are the direct links of the specification. All other terms from PeP fields indicated in the linkbase interface are associated with a global URL, a 'link agent' query that searches the PeP database for instances of a term within the specified field. A link agent query is much like a user search query, but more specific. Thus a search for the author 'Hitchcock' will produce many results, but a query link on Hitchcock will return far fewer papers because it only searches the author field.

Fragments from the linkbase are shown for the extracted [Title] and [Linkword] fields in Figure 8.2. This shows how each <element_string> for a [Title] has an <element_url>; each [Linkword] is referred to a <global_url>. In addition, Figure 8.2 shows how each link term is associated with the three conditions set by the editor using the interface above. [Source] links are structured in the same way as [Title] links, but with different conditions; all other fields – [TopicalTrems], [PublishersOrgs], etc. – are specific instances of linkwords and are typically treated as [Linkwords].




**a**




**b**

**Figure 8.2 Fragments from PeP linkbase showing the structure of: a, [Title] links; b, [Linkword] links**

When a Web page requested by the user is rewritten, as described by Carr above, the link service matches text strings in the page with the linkbase data to find points for linking. A typical linkbase is set to display [Title] links on any page. To rationalise the number and accuracy of linking on common terms all links are set to be case sensitive except [Title] links, which are case insensitive because the longer title strings are unlikely to be confused with other terms. The option to set links for all instances of a term or the first instance only is another rationalising condition. An optimum setting for these conditions has not yet been determined, but ultimately should be set by each individual user.

[Source] links are displayed on a single PeP view listing all PeP sources. All other [Linkwords] are set to display only on pages that are external to PeP, i.e. the original documents. [Author] links display on PeP views only. Page types are identified with reference to a generic URL.

To initiate a link session a user is invited to download a WebLink applet. As shown in Figure 7.1, this action opens a small control panel in a new browser window, which loads the applet. Then, subject to accepting a browser security warning, the requested document (or the PeP front page on first opening a link session) is downloaded to a second new browser window. As this page is downloaded the applet scans the page to identify text fragments that match any link terms with the linkbase, and rewrites the HTML of the page to display the appropriate link icons where a match is found.

The applet downloaded by the user contains the structured linkbase and conditional settings together with java control scripts and scripts for text matching and link insertion.

## 8.4 Known problems with the PeP link service

In operation some problems caused by the interaction of the link service and the browser presentation have become apparent and are to be resolved:

- Framed pages: in some cases the content of a frame is deleted by the action of the link service
- Animated pages: in pages that refresh periodically using JavaScript, commonly with adverts, added links are duplicated with each refresh
- Special characters: individual acute characters are overwritten by the link service; names beginning with O’ are not linked.

## 8.5 Other Web-based annotation and link services

There is no annotation or link service that can avoid adding complexity to the user interface. This is because such services have yet to be accepted as primary elements of common user environments, particularly the Web. There may be some hope this could change if open, non-proprietary annotation standards can be agreed. With increased usage and familiarity the respectability of such services might improve.

One such open standard for a Web-based shared annotation system, Annotea (Kahan *et al*. 2001), is being developed under the auspices of the World Wide Web Consortium (W3C). The paper immediately seeks to justify an acceptable basis for development, "the association of metadata to content", but soon reveals that the services it expects to support are not so arcane: "The annotations that we handle are collections of various statements about a document. They may be comments, typographical corrections, hypothesis or ratings, but there is always an author that makes a statement about the document or some part of it at a certain time." Critically, as an open technology infrastructure combining RDF with XPointer, XLink, and HTTP, the approach does not specify a user interface and is client-less, but concedes that a proxy or browser based solution is necessary, as in the case of PeP. The paper describes as "an interesting possibility" for presenting annotations on a Web page the use of "internal DOM events without actually changing the mark-up of the page", again as in the PeP link service approach.

Analysis of other commercial Web-based annotation or linking systems (Table 8.1) shows clearly how all are compromised by browser or system requirements. The venerable Alexa Internet is the most successful service and is now integrated as standard buttons on the major browsers. Other services require the installation of software on the user's machine (Atomica, Babylon, Flyswat), which was not the preferred approach for PeP's link service. Installable toolbars limited to MIE/Windows systems are offered by Alexa Internet and Flyswat. Only Deapleap's installable toolbar, according to user reviews, appears to have been browser independent, but since it has been discontinued this is hard to verify.

The growing number of discontinued services of this type is discouraging, but in a commercial environment it is difficult to reconcile the level of innovation with perceived user discomfort, either with the interface to the service or with underlying concerns about the ethics of the service, which is pertinent to PeP.

**Table 8.1 Commercial Web-based annotation and linking systems**

| Service | Description | Interface | System requirements |
|---|---|---|---|
| Alexa Internet http://www.alexa.com/ | Information and statistics for page in view; archived versions where pages unavailable; related pages | 1 Integrated into Netscape, MIE as the * Related button<br>2 Installable toolbar for Microsoft Internet Explorer (MIE) | MIE on Windows 95 and above for toolbar |
| Atomica Personal (formerly GuruNet) http://www.atomica.com/ | Word analysis | When the tray-based Atomica application is active, hold down the Alt key and click on a word in any Windows document. A window pops up that analyses the word in context. | Install software for MIE and Netscape browsers. Windows 98, NT, or 2000 |
| Babylon http://www.babylon.com/ | Word translation | A single mouse click on a word displays the translation, with the part of speech shown just above the word. Text-to-speech conversion possible | Download software and dictionary for Windows 95 or above |
| *Deepleap | 'Browser companion' with 40-plus tools, from search engines to language translators | Installable toolbar | MIE and Netscape browsers |
| Flyswat http://www.flyswat.com/ | On-the-fly, classified hyperlinks on recognised words and phrases | Toolbar buttons added to MIE. Click a designated link and up pops a menu of categorized related information. Can be customized via toolbar buttons added to MIE. Also works with other Windows applications without adding links. Instead, Alt-click any word to display links. | Install software for Internet Explorer 4.0 or later and Windows 95, 98, or NT |
| *Third Voice | Search and annotation | | Install software |

* withdrawn

Not included in the table

- QuickClick (NCBi) discontinued, believed to run on Flyswat technology
- Microsoft Smart Tags, not yet released

## 8.6 PeP links, reference links and OpenURL

Reference linking can, depending on the implementation, be viewed as a special case of a link service. For scholarly papers reference linking is becoming increasingly important because it exploits authored connections between papers, the references. Depending on the quality of the data extracted from the linked documents and the resulting link database, the links are unambiguous – in most cases the user can anticipate where the reference link leads.

In contrast, PeP's edited links are less familiar and less intuitive, but if designed effectively can supplement reference links. Despite appearances – instances of PeP [Title] linking within reference lists may look like reference linking, but is not because it is too rudimentary and inflexible – it was never the intent of this project to build a reference linking service for PeP. Such services are being developed and evaluated elsewhere, but it is envisaged that a PeP-like model in the future will include reference links, and this prospect reveals some interesting possibilities.

Currently reference linking and support services are being developed to serve three broad and overlapping communities:

- Journal publishers: CrossRef (Pentz 2001a)
- Library services: SFX, LinkOpenly (Van de Sompel and Hochstenbach 1999; Hellman 1999)
- Eprint and Open Archives: Open Citation project (Hitchcock *et al*. 2000)

Underpinning these is the proposed standard for OpenURL (Van de Sompel and Beit-Arie 2001a).

In terms of the current implementation – free access to papers; HTML and PDF are the principal formats – PeP could adopt the Open Citation approach and tools for reference linking. Other communities, however, could adapt the PeP model and use other reference linking services.

If OpenURL succeeds in establishing a framework in which these reference linking services can be used interoperably, some more interesting ideas emerge with special significance for the PeP model.

Based on the commercial SFX reference linking model, OpenURL is a 'context-sensitive' approach to linking, i.e. the linking service has some knowledge of the user – of the location, preferences, and services available to that user. The origin of this is that academic libraries subscribe to many journal and information services that may contain duplicate or different versions of papers; some papers may not be accessible at all to the user of a particular library (section 4.1.3). In this environment a fixed reference link to a paper is of little use if the user does not have rights to access that paper. The idea behind SFX was to recognise that such links need not be fixed but can instead resolve to services that are available to the user (Van de Sompel and Hochstenbach 1999).

OpenURL models a general framework and syntax for SFX-like context-sensitive services. OpenURLs contain three types of entities:

- the referenced item (OBJECT-DESCRIPTION)
- the information service in which the item is referenced (ORIGIN-DESCRIPTION)
- the service component that will deliver the extended services (BASE-URL).

An extension of this approach to allow third parties to deliver alternative services that relate to items referenced on the Web, the Bison-Futé model (Bison-Futé refers to the name given in France to alternative roads; in this analogy the author-embedded, default links provided on the Web are the parallel of main highways), adds the following entities (Van de Sompel and Beit-Arie 2001b):

- the user requesting the services (the requester)
- the type of service that is requested (serviceType)
- the information entity that actually makes the reference to the item (the referring-entity).

All entities are contained in the *ContextObject* and turned into an HTTP request, which is called an *OpenResolutionLink*. These are the essential components of OpenURL.

In correspondence with one of the principal authors of OpenURL, Van de Sompel proposed how open linking services such as PeP can participate in extended OpenURL services, in which it is planned "to allow description of 'objects' that belong to other spaces than the scholarly communication space. In such a generalization, concepts (such as – say – 'open linking') could be objects too." Because Web authors could not be relied on to insert OpenURLs for these objects "Software like yours would find occurrences of words that are contained in that vocabulary in Web-pages (like you do now). The software would then insert OpenURLs that contain those words (and their 'namespace/vocabulary-identifier'). These OpenURLs could be resolved by different linking servers that provide services for terms from the vocabulary. Your linkbase would be a default."

Yet there appears to be a fundamental contradiction, for in its detail the Bison-Futé paper implicitly deprecates elements of the PeP model, referring to the client-side approach as 'screen scraping', and appears to deviate from its own user-centric philosophy, advocating instead more author control: "a collaborative approach may be more appealing to authors of web-documents, who may be concerned about the intrusive screen-scraping approaches which blur the authorship of documents. They might feel more comfortable with a model in which the decision regarding which references are subject to the delivery of overlay services remains under their control."

How this can be reconciled with user choice or with implementation issues is not clear, because as this chapter has shown there are limited alternatives to browser or proxy solutions for open linking-based overlay services. Even if well formed, the elements of extended OpenURLs suggest complexities for author and user alike. As we will discover in chapter 9 and in the correspondence in Appendix 8, it can be hard to convince users of the benefits of a new service.

What is apparent is that there is an opportunity for the PeP model and implementation and the extended OpenURL model to be mutually informing.

## 8.7 Links in PeP: summary

The PeP link service designed by Miles-Board and Carr is a brilliant application that appears to be operating in diminishing developer space. It is hoped that further development can identify a browser-independent way to present PeP links to the user. In the mean time it remains to be seen how stable the service is with new versions of Microsoft's browsers and operating systems, which are notorious for breaking legacy applications, and how willing PeP users with future Java-less versions of Microsoft browsers will be to download a Java virtual machine to open the Java-based WebLink applet. Then there is the question of how PeP links might compete with Smart Tag links, and whether Microsoft will rewrite the browser's programming interface, which is used by the PeP link service, to accommodate Smart Tags.

An alternative to link services could be intelligent agent-based approaches. In this respect there are some similarities with emerging ideas for the 'semantic Web' initiative, as noted in the Bison-Futé proposals: "The notion of the annotation of informal documents by formal concept descriptors for the Semantic web is a relationship that is of special relevance to the ideas described here, even considering that the focus of that work is on querying, not linking." While there seems to be an infinite capacity for the technology and communications network to process these more sophisticated approaches, there is a danger that, from a user perspective, the complexities and limitations of annotations and link services will simply be re-introduced unless support for the user interface receives serious consideration at the level of the Web, or its successor.

If the Web achieved massive popularity and displaced other hypertext systems by being simpler and more user-friendly, chapter 9 shows there is limited appetite among users to sacrifice simplicity and familiarity in return for advanced services.

# 9 Evaluating PeP: results and analysis

## 9.1 Objectives and methods

*Perspectives in Electronic Publishing* has been designed as an exemplar and test model for the broader thesis:

> *Free access to electronic versions of all scholarly papers will unleash innovative journals and services that will serve users better than a journal system that relies on exclusive ownership of the original content.*

The features of the test model can be summarised:

> *Predicated on selected papers that are free to access, PeP is designed to assist browsing and, by adding links, to assist discovery.*

Thus the purpose of the evaluation is two-fold:

1. To measure the effectiveness of the implementation of the model described in the thesis.
2. To assess the viability of future versions of the model in the digital library environment

The evaluation must examine these key aspects of the PeP model:

1. Usability: operational aspects; user satisfaction. Does it work? Is it easy to use?
2. Perspectives (i.e. the added links): do the added links improve access to cited works, inform browsing, enhance discovery and make retrieval faster?
3. Is PeP a valid journal model?

The principal method adopted was user-based heuristic evaluation as advocated by Nielsen (1994a and b). In accordance with Nielsen's recommendations, and consistent with the small scale of the project, the work initially involved observed evaluation by four *expert* evaluators, all based locally in the Intelligence, Agents and Multimedia (IAM) group at Southampton University. These evaluators are experts on the features applied in PeP, notably the link service. Another local user, who was not involved in the observed evaluation, pre-tested the Web forms prior to release to external evaluators.

The results of the expert evaluation informed modifications of PeP and the design of the evaluation aimed at a wider group, in this case mediated by email correspondence and Web forms. These evaluators, the *target* group, are specialists with knowledge of the subject coverage of the PeP model, i.e. they might be typical users, but they may be 'naive' users as far as the link interface is concerned. The target group participated in a comprehensive three-stage evaluation covering every aspect of PeP.

Since the PeP model reviews and links content from other sources, it was felt prudent to invite the participation of authors of the linked content (the *authors* evaluation group) and representatives of some of the sources. This latter user group also included correspondents who received a personal invitation to review PeP (the *invited* user group). In addition, release of PeP for evaluation was announced on two selected email discussiuon lists (the *open* evaluation group). These three user groups were directed to a single-stage, Web form-based evaluation similar to the final stage of the target group evaluation.

Although it was planned to monitor usage passively, using server log statistics for example, as another means of assessing impact, this was not done. Usage was too low over too short a period to draw meaningful conclusions based on accepted Web usage metrics.

## 9.2 Plan of the evaluation

### 9.2.1 Focus

The most objective and straightforward basis for evaluating the implementation is to view it as a user interface. This aspect of the evaluation was based on Nielsen's (1994b) ten usability heuristics for user interface design, which judge issues such as design, user control, consistency, anticipating user actions, recovery from errors and user help. Instone (1997) suggests how the heuristics can be generalised to Web applications, and in the context of the PeP model it can be seen that the heuristics implicitly embrace most components of the implementation, such as links and metadata. As Instone says: “The overriding theme for applying these heuristics to the Web is to use links effectively.”

Nielsen's approach encourages the use of category-specific heuristics in addition to the general heuristics, and such heuristics were needed to evaluate certain features of this work such as content and link effects. In terms of the PeP model, the user interface approach is reasonable but strictly it is incomplete because it does not take full account of the social effects of an uncompromising switch to a new medium that the model implies. The aims and objectives of the PeP model suggest that the broader criteria to be evaluated should include:

- preferences/compatibility with existing sources of print, 'parallel' e-journals
- social acceptability, as a valid information product
- value of links

It is recognised that these criteria are subjective and difficult to evaluate. It is fair, however, to extrapolate results from the user feedback on the interface to the social impact because the two facets are closely related, i.e. you cannot motivate change without proven usability. Thus the goal was to tailor the heuristic methods to elaborate responses that, as far as was possible, would produce meaningful results on these wider issues.

A summary of the evaluation must seek to judge the answer two key questions:

1. Can an editorial viewpoint superimposed by a link database provide the level of control and integrity necessary to give users the same confidence that traditional publications can confer?
2. Can a linkbase acquire a value consistent with its value-adding role?

In terms of question 2, it is not intended to attach a financial value to the product or to put the work in a commercial or market context.

### 9.2.2 Evaluation methodology

A Goal-Question-Metrics approach was used to design the investigation.

#### *Goal 1: Ensure ease of use*

Questions:

1. Is the user able to navigate the information space?
2. Can the user navigate the database via the Web interface and use the tools provided?
3. What is the user's reaction to the presentation and the navigation tools?

Metrics:

**Number of specified tasks completed**, measured by number of correct answers supplied. Other measures: time taken to complete tasks; number of pages visited; whether search function was used.

**User satisfaction form**: measures positive and negative sentiment about the system; repeating the form later measures change in sentiment over time and with increasing familiarity with the system.

*Goal 2: Inform browsing, enhance discovery and improve retrieval*

Questions:

1. Can the information be found without added link perspectives?
2. Can the user set-up a link session and follow links effectively?
3. Is information easier to find with added link perspectives?

Metrics:

**Number of specified tasks completed**. Breadth of responses to less specific tasks. Monitor users' preferred approaches: PeP with links; PeP without links; PeP with other services; not PeP. Limitations of IE5.

*Goal 3: Establish PeP as a working journal*

Questions:

1. Is PeP a valid journal model?
2. What are the most useful features?
3. What factors will influence acceptance in the user community?

Metrics:

**Subjective questionnaire**

**Other input**: author submissions, author updates, user comments, user recommendations.

### 9.2.3 Requirements of the evaluation

The evaluation required the following activities:

- Individual meetings (local evaluators)
- Formal questionnaires (all evaluators)
- Web forms (wider user groups)

Users were asked to act on the forms-based questionnaires while viewing the link journal.

The form used by the *expert* evaluators involved two elements: a task-based activity, which became the principal activity in stage 1a for *target* evaluators, and a section concluding their views.

For *target* users these procedural elements were split between stages 1a and 2a. In addition, in stage 1a *target* users began by providing some background information. Stage 1b for *target* evaluators presented a series of multiple-choice questions to determine user responses to the object of stage 1a.

This stage was linked from the automatically generated response to the submission of stage 1a, and *target* evaluators were encouraged to complete this form immediately after stage 1a.

Stage 2a opened by asking for further user information, and then introduced the second procedure and a new task-based activity. Stage 2b, a longer version of stage 1b, was again presented for immediate completion.

Stage 3 for *target* evaluators invited further opinions and conclusions on the project in the context of other services, and did not require any further activities involving the model.

The form for stage 3 was presented to other evaluator groups, with a modified introductory section and a request for background information similar to that requested of *target* users in stages 1a and 2a. Although less familiarity with the model could be assumed for these evaluators, links were offered to versions of the procedures and tasks from stages 1 and 2.

Based on the submitted forms, opportunities were taken to extend the dialogue with evaluators. Some non-target evaluators preferred to mail a brief response rather than complete the form. An edited version of this correspondence is included in Appendix 8.

### 9.2.4 Summary methods and timetable

The timetable of the evaluation gives some indication of the process and the constraints (Table 9.1). To maintain deadlines the target evaluations had to be iterated over three months, when a longer period between stages might have produced different results. Timing of the announcements to the *invited*, *authors* and *open* groups missed many intended users during the summer break. One user's idea to invite student evaluation also had to be abandoned as no classes were then in session.

Web forms were created in Lotus Notes, the database used to build the PeP model, so copies of the forms returned by users were stored and viewed directly from the database.

Analysis of the results of the evaluation is reported in the remainder of this chapter. Detailed results are presented alongside reproduced copies of the Web forms in Appendices 1–7.

**Table 9.1 Summary methods for all stages of the evaluation of PeP**

| **User group** | **Usability/design criteria** | **Tasks** | **Methods** | **Inputs** | **Date of evaluation** |
|---|---|---|---|---|---|
| *Expert* evaluators, Southampton | Usability: operational aspects | Evaluator background questionnaire; GQM-based scenarios-browse and links | Printed hand-out guidance and form; observation; audio recording | Completed questionnaire; responses to scenarios; user and observer notes; audio tape | 2000 November |
| *Target* evaluators Stage 1 | Browsing and retrieval<br>Usability: user aspects | 1a User background; GQM scenarios-set-up and browse<br>1b Usability satisfaction (short) | Personal correspondence; Web forms | Two completed questionnaires; responses to scenarios | 2001 22 February |
| *Target* evaluators Stage 2 | Discovery: link perspectives<br>Usability: user aspects | 2a User background GQM scenarios-link set-up and discovery<br>2b Usability satisfaction (long) | Personal correspondence; Web forms | Two completed questionnaires; responses to scenarios | 5, 9, 25, 30 April |
| *Target* evaluators Stage 3 | Model Validity | Viewpoint GQM questionnaire<br>a Validity<br>b Model preferences | Personal correspondence; Web form | Completed questionnaire | 23 May |
| *Invited* evaluators | Model Validity<br>*Options*<br>Browsing and retrieval<br>Discovery: link perspectives | GQM questionnaire<br>a Validity<br>b Model preferences<br>*Options*<br>1 GQM scenarios-set-up and browsing<br>2 GQM scenarios-link set-up and discovery | Personal email; Web form; follow-up correspondence | Completed questionnaire | 25-28 June |
| *Authors* group | Model Validity<br>*Options* as above | Author forms<br>1 Submissions<br>2 Updates<br>GQM questionnaire, *Options* as above | Group email; Web form; follow-up correspondence | Submitted input forms<br>Completed questionnaire | 2 July |
| *Open* evaluation | Model Validity<br>*Options* as above | User forms:<br>1 Comments<br>2 Recommend<br>GQM questionnaire, *Options* as above | Open ‘call’; Web form; follow-up correspondence | Submitted input forms<br>Completed questionnaire | 17 July |

## 9.3 Expert user evaluation

*Appendix 1 shows a copy of the form used for this evaluation, with scores and comment.*

Number of evaluators invited to take part 4
Agreed to be an expert evaluator 4
Evaluation performed in two teams of two evaluators

The primary purpose of this exercise was to test the usability of the advanced features of the model. These users are familiar with the type of link service applied in the PeP model, but not this particular implementation, nor had they previously used PeP. Results are based on written notes made by the evaluators during the session, along with notes by the observer.

### 9.3.1 Set-up

All three tasks

- opening a browser window at the PeP home page
- opening a link session
- opening the remote index

were completed by both teams without problem, although it was noted that all evaluators proceeded carefully, trying to understand the model. Generally, the feedback was positive at this stage. The main queries were on terminology, e.g. what is a link ‘icon’, what is a ‘link session’? The first signs that the number of Web browser windows could be excessive were noted.

### 9.3.2 Exploring PeP: browsing

For each question in this section the following points were noted:

- answers to the questions
- number of clicks to reach each answer, compared with a pre-determined optimum
- whether search was used

Direct comparison between evaluator groups is not possible as some questions were modified between the sessions.

- 19 questions completed correctly: 4 not completed
- 15 questions answered within one click of optimum; 8 non-optimum
- 15 questions answered without search; 8 with search

When using search

- 1 question answered optimally; 7 non-optimally
- 5 questions completed; 3 not completed

It can be seen that most questions were completed successfully. Search was most often used as a last resort, and although it assisted completion in more cases than not, this was invariably at the expense of significant extra clicks. In other words, these results suggest users were more comfortable using links to navigate the database rather than search. The accompanying notes and comments (Appendix 1) do not entirely support this.

Questions that caused the most difficulty (i.e. most overclicked):

- Which e-journal reviews its fifth anniversary in a paper included in PeP?
- Why did the first Web-based journal to name a biological species have to be produced on CD-ROM?

A number of questions included in this exercise were designed for these evaluators, an attempt to build interest. These questions were subsequently replaced in later versions. The 'fifth anniversary' question above was omitted as it was shown not to lead to an intuitive procedure.

### 9.3.3 Link presentation

There was support for the *added* PeP links, but also informed comments about improvements. The use of graphics to indicate *added* links was approved, although the design of the graphics was questioned. This still has to be addressed in the implementation. The *added* links were found to assist users, justifying the overhead associated with the set-up procedure, although there were concerns about browser dependence, which may have been stronger had these evaluators not been provided with an MIE5-enabled PC. There was less unanimity about the number of links added and how this might be controlled, perhaps because the evaluators were not offered any means of control.

### 9.3.4 Significant changes to PeP based on findings of expert evaluation

1. Changes to the arrangement for the remote index, including:
    a. location of the opener buttons on the PeP front page and menus
    b. title of the Web window in which it is displayed
    c. size: the dimensions of remote window, previously fixed, can be adjusted by users

2. An A–Z link bar was added to Author and Title views (i.e. the longest lists) so that users would not have to click through a series of pages alphabetically to reach the required point in the list. In Web display, sorted database lists are limited by the number of records that can be displayed in a page. Without the A–Z links, up to five Web pages might have to be downloaded to reach a mid-alphabetical point in these views.
3. Fast search added to the remote (binder) index; presentation of search results improved.
4. Category links added to the remote (binder) index.
5. A *Publish-Ref* field added to the PeP bibliographic records for papers linked from PeP. In some cases the source of a paper is not the formal publication. Splitting this source from the published reference should reduce confusion between the two.

Comments and recorded timings indicated that this evaluation would be too long for Web users. So the number of questions was reduced, and the link session set-up was omitted from what became the 'browsing' evaluation, stage 1a of the *target* evaluation, which follows.

## 9.4 Target evaluation stage 1a: set-up and browse

*A copy of the form used for this stage can be found in Appendix 2. Summary scores for part 1 of this stage, about the target users (Qs 1.3b, 1.4, 1.5), are combined with results for other users, in Appendix 7 – all-user evaluation: validity. For reference with the Appendix, question numbers are shown in brackets.*

Number of evaluators invited to take part 26
Agreed to be a target evaluator 22
Completed forms 21 (incomplete 1)

This evaluation asked *target* users to test the usability of features designed to assist browsing of the PeP database. The exercise – tasks and responses – was performed entirely on the Web.

### 9.4.1 About the target evaluators

Most *target* evaluators are interested in electronic publishing (18), fewest in technical development (7). Four evaluators were interested in all (Q1.1). Other interests are also represented: Library and Information Science (15); Journal publishing (12); Web development (12); None (1).

All users were from a non-commercial background (Q1.2). This is not a very balanced group from this viewpoint, but one evaluator questioned whether universities are commercial or not.

Every evaluator uses print journals – although 16 users also say they access more papers in electronic formats than in print (Q1.4) – and 10 use all sources specified (Q1.3a). Nearly all use free e-journals (20), e-journals via library subscription (19), and other free Web sources, e.g. author sites (19), but numbers fell away for eprint archives (10).

When asked which source was used most, surprisingly ‘other free Web sources’ (7) came out ahead of print journals (5). Eprint archives were most used by none (Q1.3b).

All users have some experience of e-journals, eight in managing and producing e-journals. (Q1.5)

Although the number of evaluators is small, **these findings indicate an appropriate target group**, and suggest a progressive and experienced group of journal users prepared to adapt to developments in the electronic space but less interested in radical changes. There are some weakly favourable and negative features for the model promoted by PeP, but nothing conclusive. Although PeP has been built to appeal, in principle, to the chosen target group, it is not preaching to the already converted.

Less favourable for PeP, only seven users began the exercise using the MIE5.x browser (Q1.6a), although more (14) said they had access to such a browser and would use it if needed (Q1.6b). This still leaves a significant group (33 per cent) unable or unwilling to use the latest Microsoft browser.

### 9.4.2 Starting PeP: opening the remote

The following questions were asked to determine the success, or otherwise, of starting PeP and opening the remote binder (index):

- How many papers have been selected for PeP? 19 (90 per cent) correct
- How many browser windows do you now have open? 14 correct
- How many ‘browse’ and ‘browse about’ indexes are listed in the remote binder? 17 correct
- Who is the first author to be listed in ALL authors? 19 correct

Although this was a simple and, one said, 'straightforward' exercise, there are a number of things that can go wrong, so at one level usability and reliability have been shown. Most users were alert to the issues this process illustrates, as indicated in the comments. One felt it was distracting: “Too much

focus on browser issues, distracts from core functions of PeP." Some users reported temporary system crashes, and problems using certain networks (firewall problems, slow modems), but none was attributed directly to PeP. Although no question was correctly answered by all users, there were no critical failures as the identical scores for the first and last questions, marking the first and last stages of the procedure, indicate.

*Comments on starting PeP (with responses in italics)*

- A distracting amount of information on the home page (2 users) (*the home page was revised*)
- Can't alter dimensions of remote (*now resizable*)
- Why not use frames? (*to be addressed in next stage of the evaluation, using the link service*)
- Binder window hidden (2 users) (*classic usability problem of multiple browser windows*)
- Can't see link details in remote (*status bar hidden by sizing*)
- Needs Javascript (ugh)
- Not clear that remote window updates main PeP window (3 users)
- Term 'remote binder' won't mean much to most users; 'index and search' better (2 users)
- What is a 'link session'?
- I have a 15in screen so I needed to move windows about sometimes to see what I was doing.

### 9.4.3 Exploring PeP: browsing

For each question in this section three items of information were relayed to indicate the efficiency with which it was solved: the answer to the question, the number of clicks to reach the answer, and whether search was used. All evaluators were circumspect about providing each item of information.

*Answers*

- Question with most correct answers: ALL evaluators identified the authors, Paul Miller and Peter Boyce, responsible respectively for the 'name and a number' and 'Urania' papers.
- Question with fewest correct answers: only eight evaluators discovered precisely why the first biological species paper on the Web had also to be produced on CD-ROM. It was for durability.
- Overall, 83 per cent of answers were correct or part correct

*Overclicking and underclicking*

The optimum number of clicks to find any answer without using search was determined, and evaluator clicks were compared with that optimum (to within +/- 1 click):

- The most overclicked (more clicks than optimum) question, both in terms of the number of evaluators overclicking and the total number of clicks beyond optimum, was the biological species question.
- The least overclicked was the Urania question.
- Two questions that could unambiguously be underclicked (fewer clicks than optimum) using search were the Urania and the *Conservation Ecology* questions. Three evaluators and one evaluator, respectively, used search efficiently to answer these questions by underclicking.
- Overall, 24 per cent of questions were overclicked.

### *Search*

On the evaluation form, and on the main search page within PeP, users are encouraged to ‘browse not search’. Despite this:

- Search was used most (by nine users) for the *Conservation Ecology* question.
- Search was heavily used (by eight users) for the Miller and 'Association of ideas' questions, and also (by seven users) for the Boyce question.
- Overall search was used for 27 per cent of questions.
- Thus, search was mostly used moderately and appropriately, although if used indiscriminately in preference to browse and navigation tools it was shown to lead to significant overclicking.

### *Use of the remote index (binder)*

All users opened the remote index (called the ‘binder’ at the time of the evaluation) successfully, but there are indications that its purpose of assisting navigation of the database was not fully recognised. This can be seen most clearly from the low scores for the ‘biological species’ question, which could have been answered effectively by using the remote to identify views that might reveal the paper. There could be a number of reasons for limited use of the remote: it has too many choices, users were still unfamiliar with it at this stage in the evaluation, it is not integrated with the main browser window, and users were distracted by the need to manage multiple windows, a feature of the evaluation process.

*Time taken*

Users were asked to record how long it took to perform two sections of the evaluation. Times reported varied between 100 and 10 minutes, and the average time for these parts of the evaluation was 43 minutes. The average is in line with times taken for similar sections by the *expert* evaluators.

### 9.4.4 The perfect user

The perfect user would get no questions wrong, would not overclick any questions and would not use search. There was one perfect *target* user, and one who got close to perfect. In addition, one evaluator who used search for most (8) questions still managed to get all answers correct, overclicking just once. In fact, the evaluators who used search most scored highly on other counts.

The times taken by these three evaluators were 25, 18 and 40 minutes, respectively, which suggests that **users who are comfortable with the tools provided by PeP are also among the quickest to assess effective use**. Clearly, they were not the fastest users, but only two were faster.

This is the fun result, but **it shows that PeP can be used for browsing effectively as designed**.

## 9.5 Target evaluation Stage 1b: usability measurement (user satisfaction) – short version

*This measurement was performed as a supplement to stage 1a target evaluation: set-up and browse. An extended version of this form is shown in Appendix 3, A table of all the user scores by question is given in Appendix 4.*

Total response 21 completed forms returned – 16 questions

The purpose of this measurement, which *target* users were invited to consider immediately after stage 1a, was to assess personal reactions to the model first experienced and tested in the task-based exercises just completed.

### 9.5.1 Summary of results

Questions were scored from +2 to -2 according to how favourable the response was to the system being evaluated. Users were not offered the chance to make a neutral response (i.e. a zero-rated answer), although blanks scored zero (only 5 blanks returned in 336 replies suggests that users rarely considered this as an option).

Table 9.2 summarizes the results for this stage. A more detailed analysis can be found in section 9.8, after the results for stage 2b: usability measurement (user satisfaction) – long version.

## 9.6 Target evaluation stage 2a: discovery

*The form used for this stage, with summary scores, can be found in Appendix 6.*

Number of forms sent 21
Completed forms 17 (incomplete 0)

In this evaluation users were asked to perform a series of tasks, or scenarios, designed to mimic the typical actions and queries of a user, to examine PeP's support for discovery, that is, finding unanticipated information in response to less well specified queries. For the first time *target* users were invited to try the PeP link service. The exercise was conducted on the Web.

### 9.6.1 Resource discovery services used by evaluators

It is intriguing to find out before users investigate a new discovery tool what discovery services they typically use. Later, the evaluation tries to find out how inclined users are to revert to their chosen services after experience with PeP.

**Web-based discovery services are becoming more important to users than print and disc based services (Q1.2a)**. Use of references for discovering new works indicates possible support for reference linking (a rudimentary feature in PeP), which is affirmed (Q1.3). The zero preference for subject gateways among most-used services probably indicates the limited availability of such services in this field rather than any intrinsic dislike of gateways.

**Table 9.2 Summary of results, Stage 1b: usability measurement (user satisfaction) – short version**

| **By evaluator** Range possible 32 to -32 |
|---|
| Total 329 (Max. possible 672; min. possible -672)<br>Average evaluator score 15.7<br><br>Highest scoring evaluators<br>SCORE 27. 25, 21<br>Lowest scoring evaluators<br>SCORE 2, 8, 9 (x2 users) (note. scores not negative) |
| **By question** Range possible 42 to -42 |
| Total 329 (Max. possible 672; min. possible -672)<br>Average score 20.6<br><br>**Highest scoring questions**<br>SCORE 29 for question 10, i.e. Using the system did not hinder the task I was undertaking<br>SCORE 29 for question 11, i.e. I was able to find the information I required<br>SCORE 26 for question 3, i.e. The system was not frustrating to use<br><br>**Lowest scoring questions**<br>SCORE 9 for question 5, i.e. It was not very easy to make this system do exactly as I wanted (note. score was not negative)<br>SCORE 11 for question 14, i.e. There were not so many ways to find the information I needed<br>SCORE 15 for question 9, i.e. To get the information I needed was not very straightforward<br><br>Question with most (3) blanks (zero scores)<br>8 If the system stopped working it was not easy to restart it |
| **By section** |
| Highest scoring section<br>**Effectiveness** - the degree to which you feel you can complete the task while using the system SCORE 95<br>Lowest scoring section<br>**Navigability** - the degree to which you can move around the application SCORE 70<br><br>**Impression** - your feelings when using the system SCORE 88, average per user 4.19<br>1 The system is one that I would want to use on a regular basis SCORE 22, average per user 1.05<br>2 The system was demanding and not enjoyable to use SCORE 18, average 0.86<br>3 The system was frustrating to use SCORE 27, average 1.29<br>4 I would recommend this system to my colleagues SCORE 21, average 1.00<br><br>**Command** - the measure to which you feel in control SCORE 76, average 3.62<br>5 It was easy to make this system do exactly as I wanted SCORE 9, average 0.43<br>6 I did not find it easy to start the application SCORE 24, average 1.14<br>7 The system did not respond quickly enough to my selections SCORE 22, average 1.05<br>8 If the system stopped working it was not easy to restart it SCORE 21, average 1.00<br><br>**Effectiveness** - the degree to which you feel you can complete the task while using the system<br>SCORE 95, average 4.52<br>9 To get the information I needed was straightforward SCORE 15, average 0.71<br>10 Using the system hindered the task I was undertaking SCORE 29, average 1.38<br>11 I was not able to find the information I required SCORE 29, average 1.38<br>12 There were too many steps required to get the information SCORE 22, average 1.05<br><br>**Navigability** - the degree to which you can move around the application SCORE 70, average 3.33<br>13 It was easy to move around the information SCORE 25, average 1.19<br>14 There were plenty of ways to find the information I needed SCORE 11, average 0.52<br>15 I could find my way around the information using the remote binder and content pages SCORE 19 average 0.90<br>16 It is easy to become disoriented when using the system SCORE 15. average 0.71 |

**Do the majority of users who profess to want 'free access' to papers (Q1.3)** really mean 'free', or is the appearance of free access, i.e. library paid subscription, sufficient? This will be considered further in a later evaluation (see section 9.9.5). If taken in conjunction with the preference for stronger integration, then really free may be necessary, or users can rapidly run out of free sources on their research path. A promising result for PeP.

*Target user comments on discovery services*

- I use a variety of indexes such as Proquest, WilsonWeb, ACM DL, IEEE DL, that are available through (the) library. I find it annoying that I have to search each of those indexes separately for the same search terms

### 9.6.2 Using PeP with the link service

The link service is a critical feature in the PeP model, so this is a vital stage in the evaluation. **Only 11 from 17 users managed to open a link session due to browser restrictions (Q2.1)**. It emerged at this stage that there is also a machine limitation on opening a link session, as one MIE5.x user on an Apple machine tried unsuccessfully to use the link service. According to Tim Miles-Board, the developer of the link service: "The WebLink applet uses COM technology – not part of the MacOS as far as I know."

For Windows users the results (Q3.1–3.3) suggest all users opened the link service successfully, correctly identifying the number of added link types (Q3.3), although some remarks indicate intermediate problems, most notably "loads slowly" and "too many windows" (the evaluation form is in an additional window). It is not obvious to users, even to those who may have read the instructions, why two additional windows – the WebLink applet and the new PeP content window (see section 8.2) – are launched when opening a link session (e.g. one user wondered "why not overwrite the 'How to use PeP' page with the newly opened 'Welcome!' page?").

The browser security warning that users receive when opening the link service, indicating they are opening an application that will act on the data displayed, provoked surprisingly little concern but introduced practical problems: "My PC firewall blocked Java applet; had to enable. My IE setting blocked both ActiveX and Java; had to enable."

*Comments on opening a link session*

- It may be better if the number of windows opened (4 altogether now) can be reduced

- My browser showed a certification warning: "Test root not enabled as a trusted root"
- link session step guide did not give a security warning, I used fast start instead
- took a little time for the applet to run
- At first I didn't accept the security warning, as I thought it was trying to install some additional software. I then had to close down MSIE altogether in order to regenerate this stage. The 2 (?) kinds of link symbol look pretty alike
- two main windows (start and end) awkward

**All but one of the link service users appear to have opened the remote index correctly (Q3.4–3.5),** allowing +/- 1 window (Q3.4) to differentiate PeP and evaluation windows. With an extra window on the screen some users commented on the management of screen space ("I find myself resizing all my browser windows a lot to organize them most conveniently on the screen"). Most users were alert to the mistake in the form – the remote binder had been re-titled remote 'index' in response to stage 1 of the evaluation.

*Comments on opening remote index*

- opens fast
- I assumed "remote index" was what was meant by "remote binder" (8 users echoed this)
- Yikes, too many windows (2 users)
- New window is bigger than the screen so I can't see the bottom. No scroll. This worries me although it seems clear there is nothing below "Modified services"

### 9.6.3 Guided tour

Only MIE 5.x-enabled link service users could participate in the topical guided tour on eprints and knowledge networks. To reduce the time required to complete the exercise satisfactorily, users were offered an optional short-cut. **Only one of the 11 users able to participate shorted the tour. All but three users completed the tour without at any point losing the trail**, which included added links to external resources and returning to PeP data. Two users, coincidentally it seems, lost the trail in part 2 of the tour, missing one item; more seriously, one user missed four items in part 3. Overall, this is encouraging after the effort of opening the link session.

For one user the tour was "Fascinating – a lot of sources", while for another it was "A bit tedious". The number of different link types was found to be confusing. One user's browser crashed mid-way through the tour and was re-started.

Some difficulty was introduced by referring in the tour guidance to ‘PeP/New’ links. At the time the evaluations were performed all ‘New’ links had timed out (no new records had been added to PeP in the previous 30 days) and so these reverted to ‘PeP’ links. Some evaluators spotted this (“No such thing as a ‘PeP/New link’; it just says PeP”); others were puzzled (“Perhaps because I completed this so late that they aren't new any more!”).

*Comments on the guided tour*

- Not difficult to follow, useful for learning what PeP can provide
- I only browsed quickly through the material that was reached (but was tempted to read it all!)
- Fascinating. You have a lot of sources here I never thought of
- I didn't take time to read the articles as I went along, but for the most part the instructions were easy to follow.
- I confused PeP link with added link/added link with PeP link and misread 'record' as 'full-text'
- Went down one blind alley, clicking on an added link instead of a PeP link
- MIE partly crashed halfway – my remote index has closed as has the window where the links are displayed. I can't open a link session, so will have to close all windows and restart
- Very interesting. I'd heard vaguely about the earlier NIH effort, but not seen anything in detail about them.
- Intriguing, if a bit slow

### 9.6.4 Response to the added links

**The 11 users who followed the guide tour were unanimous in finding the *added* links to be useful (Q4.1)**. Satisfaction lessened when the system overheads associated with adding links were considered (Q4.3), especially the browser limitations and, to a lesser extent, multiple windows (Q4.2). Again, the browser security alert associated with downloading the WebLink applet seems to be of little concern (Q4.3). The style of the *added* link icons was queried (Q4.4): “I find the icons a little clumsy”.

Users were split on how to control link presentation in terms of the number of links presented (Q4.5) and who should control presentation (Q4.6). These are speculative, and in one respect unfair, questions because users do not know how PeP’s *added* links are served or what the control interface is like, but this perceptive comment recognises the way in which terms to be linked are selected: “I

value the fact that the added links are not to every possible occurrence of that term in all other papers, but to papers that have that term as a major focus".

The number of users for this section was small, but with most users completing the tour and commenting on the benefit of the added links, the results are encouraging as the link service is the trickiest part of using PeP.

*Comments on added links*

- Really great, but on a practical level I would need to permanently change my IE security settings to fully use it
- The number of links is fine. Does 'link presentation' mean 'WHERE the links should be added'? If so, then the 'user' should have the control, I think
- Even though I said 'too many' in (Q4.5), I wasn't really overwhelmed or anything, but I felt that was more appropriate than too few, which there certainly weren't
- Potential number could be overwhelming if not filtered
- It's difficult to evaluate the usefulness of the number of links without more extensive experience with the system. In general, I prefer more links than fewer links. I recommend that the system permit the user to control the number/types[?] of links
- I find the icons a little clumsy: how about shading? Ideally, user should be able to switch off some, or all
- I can see them, but I am using large font
- give the user a choice of edit controlled or user controlled

### 9.6.5 Topic selection

This is perhaps the most revealing section of this evaluation. It confronts *all* users, not just MIE 5.x users, with a simple version of a familiar exercise – discovering references to support a viewpoint – but one that requires more thought than other sections of this form. It presents an interesting comparison between those users who could access all PeP functions, including links and full-texts, and those who couldn't.

There are potential problems. One evaluator said: "the goal of the exercise is a bit unclear". Also, it's a toy exercise, so users may not tackle it as they would real research. Such an exercise could frustrate users who, at this late point in stage 2a, probably want to finish as quickly as possible, but this is a critical aspect of this section. An obvious escape route was offered for frustrated users: select the 'life

scientists' question given in the examples and copy three references from the tour summary (Appendix 6, section 4b).

**Every user but one submitted a valid viewpoint supported with relevant references**, and none took the easy option of copying from the tour summary (see Table 9.3).

Which services did they use? **All but two evaluators used PeP for the topic exercise (Q4.9)**, mostly using the lists and records of papers (Q4.8) rather than going to full texts (Q4.7), which wouldn't be necessary for an exercise of this type although it would be for real research. Even so, nine users still accessed full texts to build their lists. **Only three used other services – apparently none of these was a library-provided service (Q4.10).** One user admitted to not using PeP for the topic test because: "I was too lazy".

Question (4.10) is important. It was set to try and discover what users do rather than what they say they do. Those prepared to complete the topic selection would surely switch to services they were more familiar with if PeP was unusable. The need to save time would be acute by this stage.

The alternative view might be that however much users liked or disliked PeP, they would continue to use it for this exercise, either because they were already in this environment (which is an argument for an integrated service like PeP), or they thought they were expected to use PeP. One said: "used because asked to" (even though it is stated they may use PeP or 'any other resources of your choice').

*Comments on topic selection exercise*

- took longer because I had to actually read stuff
- I'd like to pursue the topic when I have more time
- A history trail/reading list/shopping cart would be useful.
- Searching on 'author' brings up all records because they all contain this word!
- Used 'Search' option; strategy: serial AND crisis AND origin
- It is not difficult to find an entry point for the information searching (e.g. using the PeP categories, or Search facility), but once presented with a list of papers it was not very easy to find which papers have the answers to my query. I used the Comment or occasionally the Abstract in the PeP (record) to make the selection, then go to the full-text if interested.
- PeP worked well, even with inexpert use

**Table 9.3 User selections and viewpoint/reference submissions for the topic selection exercise (this information is presented to assist interpretation of the results of the evaluation, not as authoritative research, which users would not have intended)**

| **What is the origin of the serials crisis?** |
|---|
| *The origin of the serials crisis is the "rapid increasing of the prices".*<br>1. Cummings, Anthony. University Libraries and Scholarly Communication, Chapter 6: Book And Serial Pricing<br>2. Okerson, Ann. University Libraries and Scholarly Communication, Synopsis<br>3. King, Donald W. Scholarly Journal and Digital Database Pricing: Threat or Opportunity? |
| *too much publishing, due to growth number of scholars and pressure to publish, combined with rise of commercial publishing since world war II*<br>1. Okerson, Ann / Periodical Prices: a History and Discussion<br>2. Grimwade, Alexander M / Why Science Journals Are So Expensive<br>3. Branin, J J / Reforming Scholarly Publishing in the Sciences: a Librarian Perspective |
| *increased volume and price of scholarly journal articles coincident with budgetary restrictions in research libraries*<br>1. Okerson A, University libraries and scholarly communication, synopsis<br>2. Cummings, A. M. et al., University libraries and scholarly communication, Ch 6<br>3. Henderson, A., et al., Decade long legal battle ... |
| *Increase in subscription rates by commercial publishers to raise profit margin / decrease in funding provided by university administration / cancellation of personal subscriptions*<br>1. University Libraries and Scholarly Communication, Chapter 6: Book And Serial Pricing<br>2. University Libraries and Scholarly Communication, Synopsis |
| *Virtual monopoly control by a small and consolidating group of publishers combined with explosion of article publishing and increased specialization of journals*<br>1. Mark J. McCabe, The Impact of Publisher Mergers on Journal Prices: A Preliminary Report<br>2. Mark J. McCabe, The Impact of Publisher Mergers on Journal Prices: An Update<br>3. Ann Okerson, PERIODICAL PRICES: A HISTORY AND DISCUSSION |
| *Only partial competition in the academic serials market with little incentive for serial publishers to constrain costs leading to price inflation greater than the expansion of library budgets.*<br>1. various at http://www-mathdoc.ujf-grenoble.fr/NSPI/NSPIe.html<br>2. Odlyzko Competition and Cooperation: Libraries and Publishers in the Transition to Electronic Scholarly Journals<br>3. Fishwick, Frances Economic Implications of Different Models of Publishing Scholarly Journals for Professional Societies and other small or Specialist Owners and Publishers Economic Implications of Different Models of Publishing |

| **What is the solution to the serials crisis?** |
|---|
| *The serials crisis will only resolved when the research community takes control of the publication of its own work.*<br>1. Frazier. The librarians' dilemma<br>2. Payne. A revolutionary idea in publishing<br>3. Rambler. A new solution to the journals crisis |
| *More competition in journal publishing*<br>1. Hitchcock - Concise Harnad reader<br>2. Arms - Economic models |

| **What do authors want from e-journals?** |
|---|
| *speedy publication, exposure, recognition*<br>1. Anderson, et.al - JEP , 6(3) March 2001<br>2. Langston,1996, untangling the Web Conference<br>3. Cronin, Bibliometrics, Freedom of Information Conference |
| *Authors still prefer print especially in mathematical disciplines, and are slow to change authoring styles to suit electronic publication.*<br>1. Anderson, Kent. Publishing online only....<br>2. Quinn, Frank. Roadkill on the electronic highway....<br>3. Harmsze, F-A.P. Forma and Content in the electronic age.... |

| |
|---|
| *Are authors ready to take full advantage of new publishing opportunities, or is it just techno-hype at this stage?*<br>Nothing since 1998/99. Couldn't find an answer.<br>1. Baldwin, IFLA paper |
| *Recognition*<br>1. Lederberg, Joshua Options for the Future<br>2. Anderson, Kent Publishing Online |

| **What do readers want from e-journals?** |
|---|
| *all the info in the most concise, easy to absorb format*<br>1. McKnight, Cliff E-Journals: What do users think of them?<br>2. Hazel Woodward -- Cafe Jus: an e-journal's user survey<br>3. Fytton Rowland -- Human and Economic Factors Affecting the Acceptance of e-journals by readers |
| "for every book a reader, and every reader a book" in other words information creation is for the reader, not necessarily for the author. We need to build systems for readers.<br>No supporting references |

| **Can the life scientists achieve the same success with e-prints as the physicists?** |
|---|
| *Biomedical scientists will continue to be late adopters of eprint servers.*<br>1. Laporte RE et al: The Death of Biomedical Journals<br>2. The Editors, Science: Is a Government Archive the Best Option?<br>3. Till JE: Predecessors of Preprint Servers |
| *Will e-prints work in biology? E-prints will work equally well in all fields.*<br>1. Varmus, Biomed<br>2. LaPorte, Death of biomedical...<br>3. Roberts, Genbank |

| **What's happening with ISBNs and linking to books?** |
|---|
| *Not much.*<br>1. Lynch 1997 |

### 9.6.6 Summary of discovery exercises

**A majority of users who used PeP at least in part for the topic selection exercise found that it fulfilled its brief: discovery of new papers and faster access (Q5.1, Q5.3),** confirming the finding of (Q4.10) and indicating that PeP was not typically "used because asked to" but because it worked. Only one user did not use PeP for the topic exercise, preferring more familiar sources (Q5.4).

More needs to be done to enable PeP users to find unexpected relationships between papers (Q5.2). One evaluator noted that it is "difficult to determine if 'unexpected' papers or relationships were 'revealed' without a systematic comparison".

*Comments on Summary questions*

- I don't feel that I've evaluated anything different from the stuff that I encountered in part 1
- too close to my own specialties to be a fair test – faster within its extremely limited domain; much more interesting and varied results found with Google

### 9.6.7 Making PeP more useful

**PeP could be improved most by making the link service independent of the browser, and by adding more papers (Q5.5)** (although how many papers would be needed to satisfy this requirement is unclear – stage 3 of the evaluation may offer some clues).

Ideas from users that may be acted on include:

- Maybe an alphabetical listing of the terms used in the "PeP categories"
- Listing the categories of papers instead of their 'Source' on the e.g. "Latest papers added to PeP" page seems more useful – it gives users a quick idea of what the paper may be about
- I'd prefer the abstract, rather than the comment, to be directly presented
- I found that not all that many full texts got links added
- Fewer open windows or much shorter window titles so that it is possible to easily determine what minimized windows are
- I found that if I scrolled down the page at full speed that I couldn't see the added links. This meant I had to go in short increments.
- I also wanted added links for all the papers in the reference section of the full text papers
- Control links by selecting ‘editor’ perspective/point of view on material?

PeP already has a list of sources (if using the remote index), fulfilling this suggestion: “Not sure if it has anywhere a full list of all its sources, as abstracting and indexing services generally do. That helps its validity in the eyes of researchers.” Nor does the implementation need to tackle “how it will deal with licensed content, where a site may have authorisation to access – or may not”, although solutions such as OpenURL are emerging for systems that deal with licensed content.

### 9.6.8 Target evaluation stage 2a: summary

Numerically, the experience of using PeP for discovery appears to have been positive, although anecdotal reaction is mixed. All but two *target* users completed a simple citation task by using PeP and for which other more familiar services were available; a majority of users noted that the exercise revealed unexpected papers.

It is a major concern that one-third of this select group of *target* users was unable to try the PeP link service, the principal discovery feature differentiating PeP from other models. Those that did use the

link service found it usable and useful, with qualifications. All but three users completed a guided tour on a publishing topic, using all features of PeP, without failures or losing the tour path.

Do these results portend sustained use of PeP beyond the evaluation? Stage 3 of the evaluation will seek to determine if users are likely to continue using a service based on this model.

Stage 3 will also reveal how users see PeP in the wider digital library: “The PeP approach is itself a digital library architecture”, one user suggested. PeP is envisaged as a model that contributes to the emerging digital library, but its integration in such an environment raises new issues.

One criticism was that PeP is “no improvement over a combination Google-like index that would include print sources”. Google is fine if you know exactly what you are looking for, which is not typically the case for browsing and discovery. Admittedly, Google covers a wider range of content! PeP is not a search engine.

In contrast, this comment on the discovery features of PeP accords closely with the Bush-like philosophy that inspired PeP: “I thought it was exceptionally fast at letting me go from paper to paper following my own thought progression.”

## 9.7 Target evaluation stage 2b: usability measurement (user satisfaction) – long version

*This measurement was performed as a supplement to stage 2a target evaluation: discovery. A copy of the form for this exercise is shown in Appendix 3. Appendix 5 shows user scores by question.*

Total response 16 completed forms returned – 28 questions

As in stage 1b, the purpose of this measurement, which target users were invited to consider immediately after stage 2a, was to assess personal reactions to a new feature of the model first experienced and tested in the task-based exercises just completed.

Between the two usability measurements (stages 1b and 2b) two new factors had been introduced:

1. Users were now more familiar with the application
2. A new service, the link service, was introduced in the second evaluation (stage 2a)

Intuitively these should work in opposite directions, the former leading to higher scores, the latter to lower scores because it adds another unfamiliar feature and some additional complexity. The link service is intended to be a helpful user feature, so ultimately it should contribute positive scores. As not all users were able to use the link service due to browser restrictions, it is possible to examine the differences between each type of user to get some measure of the effect of adding links.

### 9.7.1 Summary of results

As before, questions were scored from +2 to -2 according to how favourable the response was to the system being evaluated. Users were not offered the chance to make a neutral response (i.e. a zero-rated answer), although blanks scored zero (18 blanks returned in 448 replies). Table 9.4 summarizes the results.

**Table 9.4 Summary of results, Stage 2b: usability measurement (user satisfaction) – long version**

| **By evaluator** Range possible 56 to -56 |
|---|
| Total 358 (Max. possible 896; min. possible -896)<br>Average evaluator score 22.4<br><br>Highest scoring evaluators<br>SCORE 40 (x2), 34, 30<br>Lowest scoring evaluators<br>SCORE 2, 9, 10, 13 (note. no negative scores)<br><br>**Questions 1–16 only**<br>Long evaluation<br>16 completed forms returned – 16 questions<br>Range possible 32 to -32<br>Total 230 (Max. possible 512; min. possible -512)<br>Average score 14.4<br>Short evaluation average score 15.7 (Table 9.x)<br><br>Higher score for second (long) evaluation: 7 evaluators<br>Lower score for second (long) evaluation: 7 evaluators<br>Equal scores for both exercises: 1 evaluator |
| **By question** Range possible 32 to -32 |
| Total 358 (Max. possible 896; min. possible -896)<br>Average score per question 12.8<br><br>**Highest scoring questions**<br>SCORE 23 for questions 6, i.e. I did find it easy to start the application<br>and 11, i.e. I was able to find the information I required<br>SCORE 20 for question 23, i.e. There was enough information displayed on how to proceed<br><br>**Lowest scoring questions**<br>SCORE -5 for question 16, i.e. It is easy to become disoriented when using the system (with links)<br>SCORE -1 for question 18, i.e. I felt safer when using only the method I was familiar with to find |

the information
SCORE 0 for question 26, i.e. The screens became cluttered and confusing

Most (4) blanks (zero scores)
Question 8 (again) If the system stopped working it was not easy to restart it
Question 22 The system help files provide enough information

**By section**

Highest scoring section
**Effectiveness** - the degree to which you feel you can complete the task while using the system SCORE 72
Lowest scoring section
**Navigability** - the degree to which you can move around the application SCORE 28

**Impression** - your feelings when using the system SCORE 71, average per user 4.44 (short evaluation 4.14)
1 The system is one that I would want to use on a regular basis SCORE 18, average per user 1.13 (1.05)
2 The system was demanding and not enjoyable to use SCORE 14, average 0.88 (0.86)
3 The system was frustrating to use SCORE 22, average 1.38 (1.24)
4 I would recommend this system to my colleagues SCORE 17, average 1.06 (1.00)

**Command** - the measure to which you feel in control SCORE 59, average 3.69 (3.62)
5 It was easy to make this system do exactly as I wanted SCORE 8, average 0.50 (0.43)
6 I did not find it easy to start the application SCORE 23, average 1.44 (1.14)
7 The system did not respond quickly enough to my selections SCORE 19, average 1.19 (1.05)
8 If the system stopped working it was not easy to restart it SCORE 9, average 0.56 (1.00)

**Effectiveness** - the degree to which you feel you can complete the task while using the system SCORE 72, average 4.50 (4.48)
9 To get the information I needed was straightforward SCORE 14, average 0.88 (0.67)
10 Using the system hindered the task I was undertaking SCORE 19, average 1.19 (1.38)
11 I was not able to find the information I required SCORE 23, average 1.44 (1.38)
12 There were too many steps required to get the information SCORE 16, average 1.00 (1.05)

**Navigability** - the degree to which you can move around the application SCORE 28, average 1.75 (3.42)
13 It was easy to move around the information SCORE 16, average 1.00 (1.19)
14 There were plenty of ways to find the information I needed SCORE 3, average 0.19 (0.62)
15 I could find my way around the information using the remote binder and content pages SCORE 14, average 0.88 (0.90)
16 It is easy to become disoriented when using the system SCORE -5, average -0.31 (0.71)

**Learnability** - the degree to which you feel the application is easy to become familiar with SCORE 39, average 2.44
17 Learning to use the system was easy SCORE 10, average 0.63
18 I felt safer when using only the method I was familiar with to find the information SCORE -1, average -0.06
19 The guidance given was enough to use the system SCORE 15, average 0.94
20 I felt safe trying different ways to get the information SCORE 15, average 0.94

**Aidability** - the degree to which the application assists you to resolve a situation SCORE 53, average 3.31
21 The system was helpful in coping with errors SCORE 3, average 0.19
22 The system help files provide enough information SCORE 12, average 0.75
23 There was not enough information displayed on how to proceed SCORE 20, average 1.25
24 I could not understand or act on the information provided by this journal SCORE 18, average 1.13

**Comprehension** - the degree to which you understood the interaction with the application SCORE 36, average 2.25
25 The information was clearly presented and understandable SCORE 16, average 1.00
26 The screens became cluttered and confusing SCORE 0, average 0
27 I understood how to operate the system SCORE 11, average 0.69
28 Too much text was displayed at any one time SCORE 9, average 0.56

## 9.8 Analysis: what the user satisfaction tests revealed

- The first three sections – impression, command, effectiveness – scored higher (per user) in the second evaluation, confirming the hypothesis that familiarity would lead to higher scores.
- The impact of the link service is dramatic. Where scores in the first three sections were marginally higher, the first section where the effect of *added* links becomes evident – navigability – the per-user score falls by almost half.
- This section also produced the first negative score, for a question about disorientation (Q16).
- Scores in the subsequent three sections not encountered in the first evaluation – learnability, aidability, comprehension – were all lower than average scores for the first three sections, and two were noticeably lower. Again, the suspicion must be that these usability features were adversely affected by the addition of the link service, at least on first experience.

Overall, usability results were positive:

- 21 (from 44) questions in both tests scored in the top quartile between 1 and 2
- 21 scores were in the second quartile between 0 and 1
- 2 scores were in the third quartile between 0 and -1
- 0 scores in the last quartile between -1 and -2

The system was accessible at all times, and only one user reported a system crash while taking part in a test. The highest scoring questions show that all users were able to start using the system effectively and to complete the tasks set (except those unable to use the link service), and that enough information was provided to allow this to happen.

On the negative side, the lowest scoring questions show the impact of links, adding to disorientation, introducing a new method of finding information that made users uncomfortable, and adding to screen clutter. A third user satisfaction test, involving the link service for a second time, might have shown higher scores with greater familiarity.

### 9.8.1 Link service users against non-users

Comparing the scores of *target* users for the first 16 questions of each user satisfaction test:

- Of the 16 evaluators in the second test, 9 tried the link service
    - Higher score for second (long) test: 4 evaluators (scored higher by 8+7+6+3=24)

    - Lower score for second (long) test: 4 evaluators (scored lower by -8-1-7-12=-28)
    - Equal score for both exercises: 1 evaluator
- Of the 7 non-users of the link service, 6 participated in both evaluations
    - Higher score for second (long) evaluation: 3 evaluators (3+1+4=8)
    - Lower score for second (long) evaluation: 3 evaluators (-3-6-6=-15)

Neither result is conclusive about whether the link service is a positive or negative additional factor for users or non-users. If anything there is a very slightly negative impact for both groups, based on scoring differences rather than user numbers.

Why might this be? For users there is the additional overhead of using the link service against the benefit of the links providing a new way of finding information. Thus there may be a small indication of the former exerting a slightly stronger effect. For non-users there is no additional overhead but there is the frustration of not being able to participate in the full exercise and missing something. Given the slightly negative return for the second evaluation the latter must be a contributing factor.

Highest scoring users for second user satisfaction test (* link service user)

- First 16 questions only: 24*, 23*, 20*, 20*, 18, 17, 15*, 14, 13, 12*, 11*, 11, 10, 9*, 7*, 6
- All questions: 40*, 40*, 34*, 30*, 27, 26, 25, 25, 22*, 20*, 19, 16*, 13*, 10*, 9, 2

Here a predominance of link service users is evident in the highest scores, and non-users among the lower scores. This is one result suggesting that link service users are getting more satisfaction from the system than non-users.

## 9.9 All-user evaluation stage 3: validity

*The form used for this stage, with summary scores, can be found in Appendix 7.*

This stage was designed to evaluate the features and dependencies of the PeP model.

The *target* evaluators, a small and select group, offered the chance to assess how the views of users might change over time and with greater familiarity with the subject, as was shown in the results for stages 1b and 2b. Stage 3, a single-phase evaluation, afforded the opportunity to widen the number of evaluators.

Table 9.5 shows the level of response from the various invited user groups. The response to personal invitations and from authors of papers described in PeP was especially disappointing, although there are perhaps mitigating circumstances, notably the mid-summer holiday timing. Other interpretations of the level of response would be speculative, although the number of individual correspondences, reported in Appendix 8, suggest that the length and requirements of the evaluation may have been too burdensome; people may have been happy to use PeP and to comment on it, but did not have the time to report in detail.

**Table 9.5 Response levels from different user groups for stage 3: evaluating the validity of PeP**

| *Target users* | *Invited users* | *Authors* | *Open* |
|---|---|---|---|
| Number of forms sent: 21<br>Completed forms: 17<br>Incomplete forms: 0 | Individual invites to c.60 personal contacts<br>Completed forms: 7<br>Incomplete forms: 1 | Common mail sent to c. 200 authors of papers included in PeP<br>Completed forms: 7<br>Incomplete forms: 1 | Open invitation posted on *arl-ejournal@arl.org* and *lis-e-journals@jiscmail.ac.uk* email discussion lists<br>Completed forms: 5<br>Incomplete forms: 1 |

### 9.9.1 Dependencies of the PeP model

The model is **dependent** on

- Broader usage of electronic channels of dissemination
- Increasing willingness of authors to submit papers to open access services
- wider support for open access services, e.g. setting up more eprint archives, more open access peer-reviewed journals
- broader academic acceptance of open access services, e.g. by research review boards (Harnad 2001b), by academic hierarchies (e.g. Glass 2000), by libraries (see section 4.1)
- reliability, stability, integrity

The model is **probably dependent** on

- the integration of peer review at some stage in the open access dissemination process (the 'invisible hand of peer review' (Harnad 1998b))

Unlike a traditional peer reviewed journal, PeP is currently **not dependent** on

- author submissions, simply on open access to an authorised version of the work.

With PeP it is important to separate the model from the specific implementation (section 7.4). Is it likely the model could be generalised to other fields and other applications? The principal features of the model are:

- PeP models an electronic-only service with no print analogue.
- PeP provides access to works but does not store or archive original works.

Some specific features of the implementation are:

- Highly subject focussed
- All selected content must be freely accessible
- PeP produces some original works
- PeP adds critical commentary and presents selected extracts
- PeP selects papers critically for inclusion but does not formally peer review those papers
- PeP *aims* to become a focus for communication between authors and users

Sections 9.9.6–13 examine the PeP model specifically, but stage 3 begins with a series of general questions on the form and changing shape of scholarly electronic communication and publication, reported in sections 9.9.2–5.

### 9.9.2 How the evaluators use access and discovery services

The backgrounds, preferences and needs of the *target* evaluators were assessed in the analysis of evaluation stages 1a and 2a (sections 9.4.1 and 9.6.1). Other evaluators participating only in stage 3 were asked to complete some identical questions, but not all questions were the same. Where comparisons can be made, **there are no discernible differences in background between the original *target* evaluators and those who joined later (Q's 0.1, 0.3, 0.4, 0.5)**, although there were more e-journal producers among the non-target group.

One slightly curious result: no non-target evaluators use eprint archives to retrieve papers, exactly as found for the target group, but some said that more eprint archives would improve access to papers.

Note on design. Two users commented: "You're losing a lot of information by only letting us pick one. Why not rank them? Also: "Single choices will distort outcomes" (Q's 0.5 and 0.6). Ranking

adds complexity for the user. With hindsight, given the small number of responses and the split between results, ranking may have provided more useful information.

*Comments on access and discovery services*

*Invited* users

- Google is king
- Paper is still my most used medium, but only just. Still rely to some extent on browsing the current stacks
- Lots of places to look

*Authors*

- You've given me some very tough choices here. I'm a regular reader of a number of electronic-only publications; however, I'm also the main manager of our institution's e-journal collections. I find papers regularly using journal alerts, listserv bibs and announcements, references and web-based indexes. I could not give up any of these and maintain the effectiveness of my monitoring of this literature. I love *D-Lib, JEP (Journal of Electronic Publishing), ISTL (Issues in Science and Technology Librarianship)*, etc., but the reality is that even more literature comes out in print. I read *JASIS* (*Journal of the American Society for Information Science*) online now but do you consider that an e-journal?

### 9.9.3 Defining a journal

Users are asked to consider what this new model publication they are about to try, PeP, might be. Since user expectations are shaped by what something is called, this is a moot point. Unless this point was in the minds of users as they approached this section, it will have appeared deceptively bland, yet another attempt to characterise scholarly publishing, a popular pastime while journals appear to be mutating, or not, in the electronic environment. In the context of PeP, the responses to this section could be critical and will begin to determine – could even be a predictor of – what follows.

There seems to be a widespread belief that peer review is an essential component of scholarly publication, and this is confirmed to a degree, even though PeP covers a field where peer review can hardly be considered vital. The optimum response supporting the validity of PeP probably would not emphasise peer review, but most of the other functions instead. So there is minimal encouragement that **peer review isn't the exclusive choice of *target* users as the defining function of a journal**; quite a few *target* users believe that dissemination is the defining characteristic (Table 9.6).

**Table 9.6 *Target* users on the defining functions of a scholarly journal**

*This question was put to target users only*
**What are the defining functions of a scholarly journal?** (select ONE OR MORE)
Peer review 15
Dissemination 14
Editing and presentation 7
Established editorial viewpoint 6
ALL 3
Other (specify) 5

- Provides access to novel 'primary' research reports in a defined disciplinary 'niche'; one that attracts able scholars
- Publication of original content. A targeted academic readership–this may be in a single discipline or be cross-disciplinary. For non-peer reviewed journals, editing generally includes article selection
- archival – 'getting it on record'
- establish chronological priority, record peer-reviewed research for posterity, a quality filter for the reader (through peer review)
- Archival
- archiving, assigning priority / time stamp especially in science

**Peer review dominates to a greater extent when the wider group of evaluators is asked to indicate the *most* important function of a journal**, especially if the options of original and archival papers, which are typically associated with peer-reviewed journal publication, are also taken into account (Q1.1).

**The largest number of users believe that peer review will continue to be 'critical' (Q1.2)**, emphasizing with their associated comments the enduring grip of peer review on scholarly publication. If its importance is questionable, the use of peer review clearly cannot be questioned, in the view of some.

Despite this, **a majority of users believe that peer review should not be tied or will be modified**. There is a causal connection between peer review and exclusive publication based on copyright transfer from author to publisher, as argued in section 5.3, which is perhaps beginning to be recognised.

There are a number of myths about the role of peer review – e.g. it is the 'first of line defense' against information overload – but it would be astonishing to believe that scholars read only peer-reviewed papers. To be fair to users, the question is accused by one user of being "Too simplistic in its range of options, too superficial", so perhaps the responses are bound to be too.

*Comments on Defining a Journal*

*Target* users

- How to define an ‘e-knowledge-network’ vs. an ‘e-journal’?
- For (Q1.2), the most important characteristic is expert selection of articles, either by an editor or by peer-review
- I think the peer review is critical *and* it should not be tied to journals that restrict access (unless you mean something else by 'restrict access')
- I'd like to add that peer review may be modified as a new e-journal model emerges but peer review shouldn't be abandoned until any proposed alternative is shown to be as useful to the scholarly community over a period of time
- The Open Peer Review model will become commonplace within five years, if not sooner
- I consider an important additional role of a journal to be to fix particular (versions of) articles in the knowledge archive, allowing them to be used as building blocks (reference targets) for future work.
- Information overload is one of the major problems facing scholars or any information seekers today. At least in theory, peer review is supposed to be our first line of defense in terms of filtering out ideas that are poorly formed or redundant, etc.
- Difficult to choose a single most important function – peer review and dissemination need to go hand-in-hand
- This question assumes journals continue to exist. I don't believe they will.

*Invited* users

- Multimedia means print is not sufficient

### 9.9.4 Transition to e-journals

***Target* evaluators are evenly split on how to improve access to journal papers** (Table 9.7). **There are some promising models and appear to be few pre-conceptions**, but how committed might users be if the established system is threatened? For most users, access to scholarly papers has long been in printed journals through the library. Taken in series, the questions in this section attempt to discover the factors that most concern users.

Broadly, users are not inclined to maintain print as the primary medium for journals. In particular, most users are not concerned about the ability to find materials in electronic sources (Q2.1). This view weakens *very* slightly when users are confronted with concerns about integrity (Q2.3) and stability (Q2.2) of digital data, and weakens further on the issue of preservation (Q2.4). **Users**

**Table 9.7 *Target* users on approaches that will most improve access to journal papers**

*This question to target users only*
**Which of the following approaches do you think will most improve access to journal papers?**
Authors retain more rights in works 4
More open access e-journals (e.g. *D-Lib*) 4
More e-print archives 1+2
New journal models 2
Established journals make papers freely available after a set time (*PubMed Central* (PMC) model) 1+1
Other (specify) 5
*On the basis of the 'Other' responses, added 2 to score for e-print archives and 1 to PMC model*

- For biomedicine: in the short term, the PMC model; in the long term, more e-print archives
- At this time, the SPARC approach is the most promising alternative for short-term real world impact. All of the above are helpful
- Enabling libraries to integrate their article supply services with web hyperlinks
- I use a number of databases frequently such as Proquest, WilsonWeb, Elsevier, and some other sites which carry many articles in full text online form. The main problem is that I frequently don't know where to find a specific journal. Also if I'm doing a general search I have to redo that search on the search engine for each index. One thing I really like about your model is the possibility for overcoming these boundary limitations with time
- Complete change to a system of eprint archives
- I want to qualify my answer again. (Open access) would most improve access to journal papers if it were a realistic scenario. Whether or not this is the case for the whole journal literature is uncertain
- Authors as creators of a work and institutions and agencies as the funding source, it is illogical to completely transfer one's intellectual property to a commercial publisher without some rights of access and use for the individual, institution, or agency

**are most concerned about the long-term maintenance of published works**, and this is the main contributing factor in the caution that users have over the transition to electronic publication, although there remains a clear majority in favour of a faster transition from print to electronic services (Q2.5).

What are the implications for PeP? That PeP is providing services that most users are comfortable with, and does not contribute anything to the issue that concerns them most.

### *Comments on Access to Journals*

*Target* users

- I feel it is better to develop e-journals now rather than waiting for ‘better systems’ to mature
- A major source of frustration for me is the lack of electronic access to older issues of journals
- There are significant problems with licensing electronic information vs. purchasing print information (e.g. first sale doctrine, fair use, ability to ILL (inter-library loan), etc.) that influence these responses because I assume that most journals will continue to be produced by commercial publishers

- The transition to electronic journals is happening slowly but it takes time to establish standards (for good reason). The transition cannot proceed apace until preservation is certain
- I very much appreciate my access to online journals (via our library's subscription)
- The technology is more advanced than five years ago so that we take a leap of faith for each of these controversial issues. These problems will be resolved because they need to be
- I'm uncomfortable with print as the ‘primary’ medium; I think print should be retained as an ‘archival’ medium for all four reasons
- E-journals should not be held to a higher standard than print: you cannot always find them, not always available, etc.
- I live in rural Japan, at least a good four hours from the nearest chance for printed journals in English. Without the internet I could do zero research. However, in the past couple of years I have been very successful doing 100% of my research online–i.e. I'm maintaining a straight A average in my PhD program
- Free access after a set time (from publishers' sites as much as from PubMed Central) has already made a big difference. I think new journal models (e.g. BioMed Central) have the best chance of improving access in future.

*Invited* users

- Given that I don't count PDF as an e-journal format, there are very few e-journals.
- The archival problem is a huge problem as it is not only technically challenging but it is also an organizational problem which extends to definition and acquiescing participation of all interested parties (subscribers)
- Print is next to useless. I don't want to have to rummage around in the bowels of the library, or wait weeks for a copy of an article to be sent to me. People on short-term research contracts, i.e. an increasing proportion of researchers, need IMMEDIATE access to information, online, from their point of work (office, home, out in the field).
- We may want both (print and electronic)!

*Authors*

- Misses the point of portability, the key metric keeping us from digital journals

*Open*

- Early assumptions about the efficiency of electronic publishing compared with print seem not to have led to a rapid transition from the print paradigm largely because of user inertia? Demographic features could play a part here with the slow buildup to e-journal usage becoming more rapid over time.
- Re (Q2.6) – can't always cope with the number of titles we have now – find a model that works before saturating the market with different interfaces

### 9.9.5 Journals and the library

**So ingrained is the idea that scholarly journals are too expensive that it is hardly surprising that most users agree with the statement of this (Q3.1)**. It is revealing that fewer can say they have been affected by this, but most have (Q3.2). Curious then that **most users want to persist with this flawed system of library purchasing**. That is one interpretation of results which shows small majorities of users who do not expect to ignore paid-for journals (Q3.6) and consider subscription and site licences to be the best payment model (Q3.4), and large majorities who do not want to pay directly for e-journals (not one person strongly agrees that pay-per-view will dominate) (Q3.5) and who do not believe there will be 'no role' for the library in delivering e-journals (not a single user strongly agrees with this) (Q3.3).

Another reason for persisting with the current approach to library purchasing might be the lack of consensus on the alternatives (Table 9.8).

**Table 9.8 T*arget* users on what will have most impact on improving library provision of e-journals**

*This question was put to target users only*
**Which of the following do you think will have most impact on improving library provision of e-journals?**
Improved licensing terms for libraries 5+1
More open access (free) content 4+1
Stable archiving standards 2
More backlist digitisation 2
Other (specify) 4
*On the basis of the 'Other' responses, added 1 to score for licensing terms and 1 to more open access*

- Lower cost of commercial e-journals
- Libraries (in the UK anyway) need to incorporate 'free access' e-journals into their systems more effectively. Their procedures are geared towards resources obtained through paid subscription and many include free e-journals only on web sites (not on the OPAC). They also promote resources that they've paid for and perhaps neglect free journals.
- Adoption of some sort of architecture which will allow integrated search across indexes. There is a great possibility for this with tools like SOAP (Simple Object Access Protocol) or XML–RPC (portable way to make remote procedure calls over HTTP) I think, or with the recently released WSDL (Web Service Definition Language)

For PeP the findings of this section are somewhat uncomfortable. PeP is not predicated on solving the ‘serials crisis’, and although there seems to be support for new journal models in principle, it could all be held back by outdated thinking about the business models that can be supported. As one author remarked: “Most users have little understanding of payment systems. Sadly, changing their understanding will take much longer than the relevant technological development cycles.”

A brief note on design of the questions in this section. Two users commented: “Careful – you changed the order of options for (Q3.6) – not sure how much I'd trust the answers to this question”; and “Why mix up the ranking order between questions? Bias?” No bias was intended. The order of options varied throughout. Each question needs to be considered carefully and individually.

### *Comment on Journals and the Library*

*Target* users

- About (Q3.4): if a library provides extra help/information than the e-journals alone, then it will have a role. About (Q3.5) and (Q3.6): ‘pay-per-view’ costs more than subscriptions and site licences. About (Q3.7): if paid-for journals provide better quality/value-added services than the free source papers, then readers would still like to pay for them
- For some high-impact journals, an author-pays model may also be a viable source of revenue
- Pay-per-view will emerge quickly, but will not be the most cost-effective/beneficial method of providing access. There are other access models on the horizon such as the Electronic Journals Center of OhioLINK
- University libraries need to evolve to incorporate electronic press functions
- Cost-effective for whom? Pay-per-view requires an efficient payment system
- I really hesitated on (Q3.4). I like the subscription scheme, but I also like that the library handles all the subscription details for me. I can then supplement the library subscriptions with others as I feel it is necessary.
- Improved licensing terms would have most impact, but I don't believe that will happen

*Invited* users

- Pay-per-view is not the most viable solution, especially for libraries
- (Academic) libraries have still to prove themselves with regard to the provision of e-journals. (I'm) Disappointed at the progress made; the eLib programme was a good start in tackling some of the issues at an implementation level, but that was over half a decade ago, now :-(

*Open*

- The current controversy about commercial journal publishers' pricing stratagem (and role in the communication process) indicates a powerful move away from high price journal subscription models. May lead to a greater role for learned society/university press publishing but at lower levels of profitability.
- SOME journal prices are too high
- There will always need to be an intermediary – at the moment our users couldn't function without it – pulling everything together, making sure things continue to work, etc.

### 9.9.6 The role of PeP

This analysis is reaching the core of the evaluation. We have some idea in principle of user preferences, but how will real reactions to PeP match? **Almost every evaluator claims to have tried the browsing exercises as a means of gaining practice in using PeP (Q4.1)**, although apart from the original *target* evaluators, few others submitted the form that would indicate progress with the exercises, so nothing new can be claimed here. Fewer users were prepared to try the discovery exercises (Q4.2), and fewer still opened the WebLink applet (i.e. used the PeP link service) (Q4.3), even though the discovery exercise depends on using the applet. **The proportion of *non-target* users prepared to use the applet fell to 50% compared with 65% of users in the *target* group**.

Incidentally, the 17 *target* evaluators for stage 3 were not the same 17 evaluators that responded to the discovery exercise (stage 2a), but all 11 *target* evaluators that used the PeP link service in stage 2a participated in stage 3, so link service users were apparently not put off by the experience.

### 9.9.7 What is PeP?

**Over 85% of users are happy to use PeP as an abstracting and indexing service**, and none would not use it at all for this purpose (Q5.1). Even among link service users, support for this description is strong. Ignoring that not all records of papers in PeP have abstracts, there is little contention in this description, and it is usable by most, even with a second window for the remote index, apparently. PeP contains comment and other items of information that other services might not offer, and it is focussed on a topic. It would hardly be worth building PeP as an abstracting and indexing service, however, as it cannot compete in scale with established services.

A s**imilar numbers of users support PeP as a subject gateway**, apart from the three users who slipped into the 'would not use' category (Q5.2). Again, the established academic gateways are broader than PeP in every respect, most offer free services and are already funded. Kirriemuir (1999) and Koch (2000) describe the requirements of subject gateways. Apart from the added links, PeP offers little that is new against these benchmarks. In follow-up correspondence, Kirriemuir states: “There is increasing blurring in the edges between e-journals, gateways (portal/hubs) and other e-services. People get very upset when their pre-conceived ideas on this are challenged; it's difficult (pointless?) to compartmentalise a lot of stuff.”

Users appear not agree with Kirriemuir, **because support for PeP as a journal falls markedly**, even among link service users who, at that time, were effectively the only users able to see PeP as journal-by-design, with access to full texts (Q5.3) – subsequently the URLs for papers described in PeP records were uncovered, so the link service is no longer the only means of locating papers from PeP, although it remains the only means of viewing PeP *added* links.

Users are evenly split on whether PeP is useful or not as a journal. Just 25% of link service users think PeP is 'very useful' as a journal. This may have been anticipated from the responses to 'defining a journal' (section 9.9.3). As built, PeP is not a *peer-reviewed* journal, nor yet a *primary* journal of any substance (although it contains two original articles), which the comments accompanying section 9.9.10 indicate are the criteria applied by many users. Even those who like PeP hesitate to label it this way: “really useful but hard perhaps to call it a journal.”

### *Comment on What is PeP?*

*Target* users

- I've not seen any novel ‘primary’ research reports within PeP itself – only links to them in other archives, (which may themselves be unstable)
- PeP doesn't publish articles; it provides access to articles that are published elsewhere
- Peer review is a key element of a journal; PeP doesn't qualify on that basis
- Best of All Worlds for all topics on electronic networked publishing
- To me, PeP is not a primary journal, because it is not the primary vehicle for archiving articles as contributions to knowledge. It strikes me more as the electronic (and much more functional) equivalent of a review journal.
- In my view PeP is best typified by calling it a critical electronic abstract journal. Critical because of the comments, electronic is obvious, abstracts also. But journal? ... Because of the comments, I am inclined to call it a journal, because of the editorial work it implies, but otherwise, PeP resembles an abstracting and indexing service, which have their roots in abstract journals, which were called journals because they were published on a regular basis.
- The linking strikes me as potentially invaluable
- Does PeP select some articles each month for readers to peruse?
- PeP seems to be more of an overlay on existing sources, aiding discovery and comprehension and facilitating links between sources, rather than disseminating new content
- I'd view PeP as a set of table of contents with cross-linking independent of where the material was published or how. That's not a journal but would be a pretty useful beast in its own right.

*Invited* users

- It's what the Web was supposed to be about, according to Tim Berners-Lee's earlier visions – a quick way of moving between associated bits of information. It adds value to articles where the referencing and/or linking to other information is poor.

*Authors*

- It ain't a journal; it's either a Zine or a Resource.
- PeP is unfortunately too hard to use, confusing navigation, cluttered interface, very slow over a modem, otherwise a nice idea

*Open*

- Need for Internet Explorer negates value
- The very detailed instructions re. Links are very offputting – I did not proceed to even try to look at papers of interest. Users of e-journals want ease of use, not something that can only be used by following what appear to be complex procedures.
- Far too complicated for the average user who wants speed and would probably give up before installing the applet

### 9.9.8 PeP in the emerging scholarly e-journal infrastructure

Is the rejection of PeP as a 'journal' a rejection of the model or the label, and should PeP even be seeking to differentiate itself from established models? Considering PeP in the emerging infrastructure may reveal more. Up to this point in the evaluation PeP has not been strongly promoted as a journal, but the point is re-emphasised in this section. Users are invited to consider how PeP compares with other publishing models (see Table 4.1). Information was presented but no questions were asked. The most revealing comments indicate that users do not necessarily reject PeP as an original peer-reviewed journal, but they are confused whether these are its objectives. A fair observation is: "It's news to me that PeP is designed to be a 'journal'. I see PeP as a discovery tool and an exploratory tool."

PeP may include an excess of features, increasing complexity and cost. As well as trying to elaborate what PeP is, the next sections of the evaluation seek to identify the best balance of features.

#### *Comments on Emerging Scholarly E-journal Infrastructure*

*Target* users

- Open access sources provide various values; however, coordinates services will make them more useful and will help the readers to discover information more efficiently

- For me, the current PeP model meets a need (easy access to high-quality reports)
- PeP's primary role would need to be publishing articles for me to see it as a journal. I suspect that if PeP were widely known as a ‘journal’ and it included articles from commercial journals that significant copyright challenges could arise.
- I envision an increase in the Open Access (and Supra-Open Access) model. Concurrently, I also see modification of conventional access models to include pay-per-view and consortial 'on-site' collections (e.g. OhioLINK Electronic Journal Collection).
- I presume PeP is quite labour-intensive, so would expect that while access to the papers would be free, PeP itself would be a charged-service

*Authors*

- Working through the scenarios did not give me the impression I was working with a journal. Most of the content I saw was content I was aware that been published previously. I see PeP as a discovery tool and an exploratory tool, less as a publishing medium
- Honest to God (and I just loathe your eensy-teensy little scrolling boxes), I can't even make sense of the matrix.

### 9.9.9 Selectivity vs access vs scope

**Most users support the PeP model in three trade-off cases** – free vs peer review (Q6.1); focus vs selection (6.3); volume of content vs automation (Q6.4) – but by a majority of users in only one case (Q6.1).

This section is tricky for users, who need to appreciate the connections between the questions, typically connections that will confront publishers rather than users. Thus it is not always evident in the results that users have been able to adopt a consistent position (“Not sure I've understood Q. 6.3 and 6.4”), so some caution is needed in assessing the results for this section.

That **free access is preferred to control by peer review *for PeP***, given the findings above, suggests that users are prepared to encourage PeP to find a niche that complements peer-review journals (Q6.1). The surprising willingness of 50% of users to take a view on whether PeP will strengthen or threaten peer-review journals, the majority of this group predicting PeP will strengthen journals (Q6.2), offers support for this conclusion.

**Users clearly prefer the new model to focus on a defined topic rather than adopt a disciplinary-wide breadth (Q6.3)**. In the latter respect PeP avoids having to emulate arXiv, and also avoids the

bland option of a universal journal. Users are split on editorial selection against automation, perhaps because they are sympathetic towards the advantages of journal-like editorial control but attracted by the impact of the automated ResearchIndex system in computer science.

Some automation in PeP would save time and enable it to include **more papers faster, which is strongly supported (Q6.4)**, but at the expense of selection, comment and other features designed to inform users. Results on this last trade-off are inconclusive.

“Ideally, the widest range of papers are included, along with indicators of whether peer-reviewed.” This is an objective of PeP, but harder in practice and not yet implemented. It is difficult to know what has been peer reviewed as the range of sources is diverse and so are the policies on review. It is not appropriate to guess.

### *Comments on Selectivity vs Access vs Scope*

*Target* users

- My own major need is to be pointed to high-quality reports, peer-reviewed or not
- The real answer to (Q6.1) is include all high-quality articles whether peer-reviewed or not and whether freely available or not
- I do not believe that the PeP model (or any Open Access model) 'threatens' the peer-reviewed journal, *per se*. I believe 'challenge' is a more appropriate characterization. Publishers are responding (in part) with the development of the pay-per-view model and innovative consortial arrangements such as the Electronic Journal Center offered by OhioLINK.
- Ideally, the widest range of papers are included, along with indicators of whether freely available, peer reviewed, etc.
- (Q6.2) I chose to define ‘strength’ in terms of breadth of readership, rather than financially, because I think ultimately the strength of a journal lies in the number of people who choose it as an information source. The PeP model does however cut into the revenues journals might have otherwise received for subscriptions....
- (Q6.3) I really think the editor is what makes or breaks the journal in this case. It's really important to have someone filter out the trash. It might be possible to have an index of the PeP-like journals so that one could find the journal that covers the specific scope one is interested in.

*Invited* users

- Helps to understand the boundaries if not too big

*Open*

- Most of the papers listed were ones I already knew about, and those from very well known sources are probably easily found without this service. It's more peripheral things that are more difficult to identify
- PeP should list all relevant papers whether or not freely accessible and whether or not peer reviewed but should flag each article to indicate whether it is peer reviewed and whether it is subject to access control

### 9.9.10 Funding PeP-like services

The questions posed here might suggest otherwise, but at the time of the evaluation there was no plan to seek any funding to support PeP. It is a model to test a thesis and these questions contribute to that test. Results for 'Journals and the Library' show that while new business strategies may be needed to support new publishing models, broadly users are disinclined towards new business models. This section aimed to find out if this view might moderate given a specific example.

For those who dislike the PeP model this section is an opportunity to query its presumptuousness. Yet only five users took the chance to express caution about the prospects for funding (Q7.1), most for practical reasons (Q7.5), others with more fundamental reservations: “unnecessary”, and “it is too early”. Even supporters of PeP advocate the 'volunteer' model.

Despite this, **85% of those asked who might pay indicated that there are reasons for institutions or funders to support PeP (Q7.1)**.

Intriguingly, institutions and users might prioritise features differently. Results indicate that **institutions would prefer to purchase a complete package** rather than component parts, i.e. probably not a link service alone, although more would be attracted by the conventional feature of original papers (Q7.2). **Users might prioritise original papers, interestingly *PeP review* papers rather than peer-reviewed papers (Q7.3).** Both institutions and users would attach lowest value to original content in the form of PeP records.

The surprise among user-recommended funders that might be willing to support open access is that 'a university' ties with 'Research funding body' with most votes (Q7.4). Clearly any funding on an open access basis would have to be justified by serving a large part of the funder's target community.

Are the options offered in these questions radical enough for a radical model? “You missed payment to be peer reviewed for inclusion in PeP. That is the best funding model. Users would pay to have their work exposed/publicised as being of quality. The models you are suggesting are just journal models revisited.” A good point, except that any funding model must be consistent with the objectives of the model (section 9.1). Paid-for peer review would not be in this case. There may be other business models that have not been considered here.

### *Comments on Funding PeP-like Services*

*Target* users

- Whoever wants to use PeP to expand the accessibility of relevant information may pay for it
- Scope-specific research networks need scope-specific PeP-like services
- Hard to believe that the technical infrastructure costs would be great. The trick is to build a volunteer human infrastructure that rewards participants through professional recognition.
- Institutions/users might also pay for PeP but I think that 'supplier' paid is the better model because it eliminates the administration of payments.
- Start out on grant money, get a track-record, then sell to institutions/organizations/individuals
- Value is added to articles by seeing them in the context of related articles. Both suppliers and consumers would seem to benefit, so finding a way to split costs would be preferable. PeP seems almost a ‘public’ good, the trick being to define the public.

*Authors*

- PeP needs to be free – subsidized indirectly through server space/bandwidth donated by research institutions and work/review time donated by participating researchers

## 9.9.11 Using PeP: readers

The prospect of paying for a service focuses attention on key features in ways that free access does not. That must be the conclusion of discovering that **readers place links and commentary above orginal content as attractive features (Q8.1)**, reversing the findings of the previous section. **The highest rating, for barrier-free access to papers, effectively vindicates the PeP model**. Original content also scores highly in this survey, which might be more significant than the ranking order and differences in score, which are small.

Active user forums would be a bonus that ought to work in a PeP-like environment, but readers probably do not expect them. Extracts from papers, as found in most PeP records, are so unusual that

this feature is easily misunderstood. Typically users might want to annotate papers themselves, and such a facility might become possible if user forums succeed.

**The main deterrent for users, predictably given earlier findings (section 9.9.6), is the PeP user interface (Q8.2)**. 'Unlikely to find anything new' is a justifiable deterrent for well-read users, until more original content is included in PeP, as is 'Incomplete coverage' for PeP in its present state, although comparison with any conventional journal would be interesting on this basis. There is no resistance at all to an electronic-only model.

Combined, reference linking and citation analysis would be the most desired new features (Q8.3). Full-text search would also be beneficial.

That **five of every six users would recommend papers to PeP (Q8.4), and four in seven are more likely to comment on other works in PeP (Q8.5)**, is not yet borne out in practice.

**PeP is a research tool, so it is positive that 'researchers' might find it most useful (Q8.6)**, including researching librarians ("Why ignore librarians? We're doing a lot of research in this area"). The score for publishing professionals might have been higher had more participated in the evaluation; and for students too had any been invited to take part.

*Comments on Using PeP: Readers*

*Target* users

- The PeP model and content is of potential value to a wide variety of users
- I really like the opportunity with PeP for the organization of the content to be a truly collaborative product of the users' participation. The ultimate goal is of course to find the information you're looking for in the most direct fashion

*Authors*

- The part I enjoyed the most was reading your comments on the included articles!

### 9.9.12 Using PeP: authors

**Fewer than 10% of respondents have so far not made any papers available in open sources** (one of those has not published any papers yet) (Q9.3). This is fertile ground for PeP. In the field covered by PeP open access papers can be found on personal, departmental and institutional servers, and in open journals and open conference proceedings (Q9.4). This is no surprise. PeP is predicated on this

observed distribution of papers, and on the lack of integration between them as well as the lack of coverage in bibliographic services.

**A large majority of authors believe they are more likely to make their papers available in open access sources in future (Q9.5)**, which if fulfilled must involve authors with 'some' papers available in open publications adding more. This finding will clearly depend on the stability and durability of appropriate open sources, and on competing non-open publications. Unfortunately, **it seems PeP would not enhance the prospect of more open access papers (Q9.7)**, unless users understood the question as asking whether they were *even more* likely to publish openly ("PeP might further encourage me to publish in open access sources – not that I need much more encouragement").

**Virtually none of the open access papers by PeP evaluators are to be found in in eprint archives**. Open Archives-compliant institutional eprint archives would be another open access source that might appeal to these authors, but there are too few for this type of source to register in this exercise. OAi-compliant archives would be a good foundation for PeP. Rather late, after the *target* evaluation, a question was added to investigate whether authors would be prepared to use an eprint archive if one was available (Q9.4a). Most respondents ignored it.

**Almost all users would be happy to see their papers listed in PeP (Q9.1);** the tiny minority who are not could be strongly opposed ("comments on a 4 year old article are of little value and may, in my view, distort the debate away from issues of importance today ... And raises the question of possible violation of copyrights"). It is curious that fewer authors than would want their papers to be listed want to tell services such as PeP about their papers (Q9.2), but the difference is minor.

**PeP offers sufficient features to benefit authors listed in it (Q9.6)**. Reader feedback is valued more highly than editorial comment, which is among the least liked features. Authors would welcome extra exposure for their work, if it appears in context, probably among selected papers by their peers.

### *Comments on Using PeP: Authors*

*Target* users

- I'm more likely to make my work available in open access resources because I think it likely that this route will be more available in future and that these will be widely used. If I had to choose between 'open access' and 'closed access' I'd have to be persuaded that the former were used by the relevant audience. So, PeP would make it more likely if I knew that it had a large, relevant audience.

### 9.9.13 Evaluators' conclusions

Given a straight choice, over **70% of respondents conclude PeP is a valid journal model (Q10.1)**, a higher proportion than might be expected on the evidence of (Q5.3). **Most users will use PeP again, but only on an occasional basis (Q10.2)**. Users were split on whether making PeP itself accessible from, or as part of, a more familiar service, another journal for example, would raise usage (Q10.3), a speculative response since there is no existing model for this. Finally, **six in every seven users agree that PeP is an original service that is not available elsewhere (Q10.4)**, with some users quick to qualify this in case it is viewed as endorsement: "Don't take this to mean that I believe I could rely exclusively on PeP, it's just another channel I have to monitor".

#### *Comments on PeP as a journal*

*Target* users

- I think that I prefer the term 'knowledge network'
- It's great, but I don't see it as a 'journal'
- It's not peer reviewed. It's a valid model of an alternative scholarly communication format.
- (Q10.1) Emphatically
- PeP is a most innovative model and should be emulated by Open Access and non-Open Access journals alike!
- If one considers a journal to be a log of a journey along a specific path I think it is extremely valid, perhaps more so than others, in the sense that it allows individuals to find their own paths through the material
- Not clear on mechanisms for receiving original content and undertaking peer review
- Start thinking about it as a combined peer-review mechanism and a new kind of secondary literature rather than a journal

*Invited* users

- More like extended bibliography with links?

*Authors*

- As a Zine/Service, fine. It isn't a journal.
- The editorial goals are unclear

*Open*

- Too many problems and not enough in context help

*Concluding comments from evaluators*

*Target* users

- PeP is a very interesting experiment!
- I strongly encourage you to continue with PeP!
- I'd be more likely to use PeP regularly if there were an alerting service so that I were informed by email of new papers
- PeP is the synthesis and integration of the best features, functionalities, and content of all worlds!
- PeP is very thought provoking, indeed
- PeP, even just linking and abstracting, would be a great addition to existing online services
- Write it so it meets W3C standards then it will ultimately work on any browser

*Invited* users

- Intriguing – used it to find some papers I didn't know about for current research am doing!
- Excellent idea – lots of links and windows confusing at first

*Authors*

- Arrghh. I gave up on this hopeless form after early portion of (7). You're using my material (which is fine with me); you're asking me to spend *way* too much effort to communicate semi-blindly with you (I LOATHE THESE DAMNABLE LITTLE BOXES)...and, frankly, I just don't see the point.

*Open*

- The PeP model appears a valuable addition, taking the best of primary, secondary and tertiary publishing with a cross-reference overlay. A useful forum for building up an interactive community in the scholarly information area. Congratulations!
- (From the email posting the open invitation for evaluators: “First reactions suggest that PeP presents too many options for users and may be too complex as a user interface”) This is absolutely true! I would like to browse the contents of PeP in two ways: a) by date of original publication (not possible at the moment); b) by date of addition to the database (WITH THIS DATE INDICATED which it isn't at the moment). All the rest of the complex interface I would happily do without.
- I suppose it offers more problems! Not impressed overall with how it looks but the idea is a valid one

## 9.10 Target users: comparing their views with other evaluators

Evaluation is hard, offering evaluators some influence but few prospects of tangible reward. As one evaluator who completed just half of one stage evaluating PeP complained: “you're asking me to spend way too much effort to communicate semi-blindly with you.” So the *target* evaluators who completed all stages have made a remarkable contribution to this study, and it is worth making special note of their views. How might their views compare with other evaluators? Given that the background of the *target* evaluators is not significantly different from other evaluators (see section 9.9.2), what this comparison might show is the effect of more sustained and focussed usage of PeP. Qualitatively, *target* evaluators appear:

- No more inclined to use all features of PeP
- More likely to be very positive about PeP as an abstracting and indexing service, but no significant difference with other groups on PeP as a ‘gateway’ or ‘journal’
- More likely to think PeP will strengthen peer-reviewed journals

- Slightly more positive towards features such as links, commentary, extracts and user forums
- Less likely to think coverage in PeP is incomplete, but more likely to think it is too narrow
- More likely to want citation and reference linking features added than full-text searching
- As authors, keener to benefit from PeP features such as reader feedback, links and comment, and more positive about PeP features generally
- More likely to make all their papers openly accessible

- Keener on the open access funding model for PeP
- No more likely to think PeP is a valid journal model
- More likely to be occasional users of PeP
- More likely to strongly agree and less likely to disagree that PeP offers an original service

## 9.11 Rebuilding PeP: recommendations

The following recommendations for improving PeP can be indirectly gleaned from the results:

*For users*

- Make it usable on multiple platforms on non-MIE browsers
- Reduce the complexity of the interface (fewer windows)
- Add more papers quickly, at the expense of comments and extracts if necessary; be selective

- Don't concentrate exclusively on peer-reviewed papers, but identify those that are

*For authors*

- Be careful and sensitive to the needs of authors, who want exposure to their peers, and feedback from readers

*The Market*

- Don't expect to be funded, but if funding is needed make sure PeP has more original papers
- Don't promote it as a journal unless there is enough original content to justify this
- Don't attempt to make individual components of the model the selling point, e.g. links, at the expense of the whole package
- Target PeP at researchers; possibly at students (presumably through teachers and libraries); possibly at publishing professionals

## 9.12 Acknowledgement of evaluators

I wish to acknowledge the extraordinary contributions of the *expert* and *target* evaluators (Table 9.9) who contributed so much to the results of this project. I gave a promise not to identify evaluators against their responses to the questionnaires. I hope they will forgive this general recognition.

**Table 9.9 Acknowledgement of *target* and *expert* evaluators**

| ***Target* users** | ***Expert* users, IAM Dept, Southampton U.** |
|---|---|
| Ann Apps, University of Manchester, UK<br>Charles W. Bailey, Jr., University of Houston, USA<br>Tony Barry, consultant, Australia<br>Morgan C. Benton, NJIT, USA (based in Japan)<br>Donna Bergmark, Cornell University, USA<br>Rob Cameron, Simon Fraser University, Canada<br>Fred Friend, University College London, UK<br>David Goodman, Princeton University, USA<br>Leah Halliday, Loughborough University, UK<br>Wayne Heiser, University of California, Berkeley, USA<br>Terry Hulbert, Institute of Physics Publishing, UK<br>Eric Lease Morgan, NCSU, USA<br>John MacColl, University of Edinburgh, UK<br>Gerry McKiernan, Iowa State University, USA<br>Frank Norman, National Institute for Medical Research, UK<br>Hans Roes, Tilburg University, The Netherlands<br>Don Schauder, Monash University, Australia<br>John W.T. Smith, University of Kent, UK<br>Jim Till, University of Toronto, Canada<br>Andrew Treloar, Monash University, Australia | Samhaa El-Beltagy<br>Jessie Hey<br>Simon Kampa<br>Gary Wills<br>Zhuoan Jiao (Web user pre-testing) |

# 10 Conclusion

Electronic journals exist in a post-Gutenberg, post-Google information environment. The ramifications of the former have been widely discussed (e.g. Harnad 1991, Guédon 2001), the latter relatively little in terms of its impact on the journal literature. Google may be a search engine rather than a journal, but the ability to locate a specified item of information precisely and instantly among the mass of information available on the Web – by March 2001 the Internet Archive had stored 10 billion Web pages (100 terabytes of data) – has profound implications. In the electronic environment the search engine has become the *de facto* interface to information, rather than the fragmented packages that have migrated from the print world.

The task for journals is to reinvent themselves as expert filters, labellers and commentators on the literature, because when anything can be found instantly this is the service that scholarly users in particular will need most. Even in this respect Google is not content neutral, analysing authored Web links to rank results using techniques that have become familiar in scholarly citation analysis.

That few journals, while completing the massive task of creating electronic versions and services since 1996, have yet to reinvent themselves in this way is hardly surprising. The findings of this study suggest that many users are wary and unprepared for what must inevitably follow.

The volume of networked information available to users today may be large and intimidating, but in scale it is embryonic. Science is moving towards data networks on a very large scale – ‘e-science’ will increasingly be carried out through distributed global collaborations enabled by the Internet, requiring access to very large data collections, and very large-scale computing resources. It is inconceivable that the published peer-reviewed literature will be unaffected by such developments, in form as well as format (Berners-Lee and Hendler 2001, Rzepa and Murray-Rust 2001). Digital information demands to be managed as a continuum: on this scale there is no other option.

The goal of this project was to reinvent the ‘journal’ as an interface to selected, evaluated and commented freely-accessible papers in an open, Google-indexed Web, and to discover user reactions by setting a series of browsing and, using editorially *added* links, discovery tasks. The proposed model was implemented as *Perspectives in Electronic Publishing* (PeP), based on a database describing the selected papers and supplemented by a link service providing access to the full texts.

The thesis opens by anticipating users' unease at meeting such a service, where works are not fixed in time or owned by the service, and where the service might materially alter the users' view of a given document. Such an approach at first appears to contravene many vested interests prevalent in producing and consuming the scholarly literature – preservation, rights and licencing, commercial exploitation. Given the added complexity of the link service interface and browser restrictions, coupled with an uncompromising decision for the duration of the evaluation not to offer links to the full-text papers other than by means of the link service, a degree of hostility could be expected.

The project was underpinned by a conviction that more effective ways of interconnecting the scholarly literature are necessary, and that to do so will challenge existing preconceptions. PeP offered the first chance to confront and assess those challenges.

That conviction is reinforced by an assessment of emerging technologies and changing publishing practices. The convergence of publishing and technology points to open systems and open access as critical characteristics of network digital services (chapter 2). In this respect journals may have changed very little, even after digitisation, but there is a growing recognition that journals are no longer isolated entities but part of a wider information environment. Examples of 'overlay' journals, which provide a peer review service for papers deposited and managed by independent services, have been shown to demonstrate high scholarly impact (chapter 3).

Centred on digital libraries, a formative infrastructure for disseminating scholarly publications has emerged almost surreptitiously (chapter 4), which is beginning to realise the growing influence of information interfaces and services above content. In this respect scholarly information services are well placed to adopt generic Web services technologies and to exploit progress of the semantic Web (Miller *et al*.). These advances have yet to be reinforced by a concerted effort to decouple the journal literature from the constraints of packaging and ownership and to build an open, defragmented corpus of materials on which these services can operate (chapter 5), although the physics and biomedical communities have adopted different models to promote free access to scholarship, based on open eprint archives and delayed open access journals, respectively.

Set against this background user reactions to the model represented by PeP should be measured by the criteria set out (chapter 9), in particular the three goals of usability, support for browsing and discovery, and whether it is a valid journal model.

### *Goal 1: Ensure ease of use*

It was found that *target* users were able to navigate the information space bounded by PeP using the Web interface and the remote index. Another result showed that users who scored highest on the information-seeking exercises were also among the quickest to assess effective use of the tools.

### *Goal 2: Inform browsing, enhance discovery and improve retrieval*

Results of the browsing exercise show that PeP can be used for browsing effectively as designed. The discovery exercises using the PeP link service are trickier to assess. Only 11 from 17 *target* users managed to open a link session due to browser restrictions.

The primary discovery exercise was a guided tour, which required use of the link service. All but three users completed the tour without at any point losing the trail. Users who followed the guided tour were unanimous in finding the *added* links to be useful.

To benchmark the response of users able to follow the guided tour against those who were not, all users were invited to select a topic and seek references supporting an expressed viewpoint on that topic. The link service was not required, nor was PeP mandated as the information service to be used for this exercise. A majority of users who used PeP, at least in part, for the topic selection exercise found that it fulfilled its brief: discovery of new papers and faster access, and a majority of users noted that the exercise revealed unexpected papers. It could be argued that the benefit of the link service will become more apparent as the collection of selected papers grows. At least as many new papers as are already listed in PeP remain to be assessed for inclusion.

Tests designed to evaluate browsing and discovery were both followed by user satisfaction tests, which confirmed the hypothesis that familiarity with PeP leads to higher usability scores, but that usability is adversely affected by the addition of the link service, at least on first experience.

### *Goal 3: Establish PeP as a working journal*

This is the most contentious area of the study. While there is no formal definition of a scholarly journal, for many the pre-requisites are original papers and peer review. Peer review was not regarded exclusively by *target* users as the defining function of a journal – quite a few believe that dissemination is the defining characteristic – but peer review dominated to a greater extent when the wider group of evaluators was asked to indicate the *most* important function of a journal, and the largest number of users believe that peer review will continue to be 'critical' for journals.

PeP is not peer reviewed, and currently offers few original papers. Nor does it store or preserve papers. It is produced by a sole editor, another factor that counts against its acceptance as a journal. None of these is designed to be as limiting in the general model as in this particular implementation.

PeP was not conceived as a conventional primary journal, and in its present form its claim to be a journal rests on two features in common with other journals: bounded access to full-text papers, and an editorial 'voice', in this case expressed through selection, commentary and *added* links.

PeP builds on papers in open access sources. Selecting from over 150 recognised sources, as well as authors' personal pages, suggests that this is a promising approach, with a large and growing volume of material to choose from. Over 90% of test users claim to have made papers available in open sources, and a large majority of these authors believe they are more likely to make their papers available in open access sources in future, although virtually none of their papers are to be found in organised eprint archives. In terms of coverage, users clearly prefer the new model to focus on a defined topic rather than adopt a disciplinary-wide breadth.

Do users accept PeP as a valid journal model? In principle, yes: 70% of respondents concluded that PeP is a valid journal model. In use, however, support for PeP as a journal fell markedly, even among link service users who, during the evaluation, were effectively the only users able to see PeP as journal-by-design, with access to full-text papers. Over 85% of users are happy to use PeP as an abstracting and indexing service, and similar numbers support PeP as a subject gateway.

Six in every seven users agree that PeP is an original service that is not available elsewhere. The main deterrent for users is the PeP user interface and the insistence on use of the link service to retrieve full-text papers. This has since been relaxed.

A less compartmentalised terminology adopted since the evaluation describes PeP as a *journal-centred portal*, recognising that while PeP features 'portal' elements strongly, the number of original papers – which could be peer reviewed – needs to be boosted to reinforce its role as a journal. It was always expected that PeP would originate review and comment papers.

### *Prospects: can PeP acquire a value consistent with its value-adding role?*

Clearly there are implications in these results for the transition of the PeP model from project to full service, and for the prospects of similar models. Of those asked who might pay for PeP, 85% indicated that there are reasons for institutions or funders to support PeP.

Institutions might prioritise features differently. Results indicate that institutions might prefer to purchase a complete package rather than component parts, i.e. probably not a link service alone, and more might be attracted by the inclusion of original papers.

For obvious reasons most funding agencies want the results of research to be disseminated, but also to make an impact. Historically the formal mechanism for this has been peer-reviewed publication. Funders have therefore tended to support this mechanism by purchasing the publications that appear to confer impact. Equating the need of the funder with the need of the author in this way, rather than with the end user of the information, is likely to hinder new models of publication aimed at users.

As a secondary service – a subject gateway or knowledge network – that exploits open access sources, PeP would itself benefit from funding that would enable it too to be open access, but may instead have to sell value-adding services in the market.

### *Limitations of the evaluation*

Evaluation is alternately heartening and daunting. Few journals are evaluated as systematically as PeP has been, but few depart from the established journal model as radically. The most noticeable limitation of any evaluation is that it interferes with real use. Often the task-based form instructing the user becomes the object of the evaluation rather than the object itself. The evaluation of PeP was not immune to this. More specifically to PeP, it would have helped had PeP been available to users for some months prior to inviting evaluation.

A constant dilemma for this evaluation was whether to present PeP to evaluators as a functional service or as an experiment; whether to present it as a 'journal' and invite users to dispute this, or to leave open the possibilities of what it might be so that users could decide. It has to be admitted that over the extended period of the project this was never fully resolved.

### *Future work*

The evaluation reveals a number of ways in which a more complete implementation would improve the impact and service provided by PeP:

- Simplify the user interface, e.g. browser independence for the link service
- Cover more papers, more originals
- Greater link control for users

In the time available and with limited resources it was not possible to build all the anticipated features of the model as effectively as intended. Planned supplementary features include:

- An improved framework for user forums (section 7.2.2.3) and support for user annotation
- More *added* links, notably on authors (section 7.2.3.4) and category links (section 7.2.3.5)
- More editors and contributors

As well as more original papers, more types of content, including less formal documents such as news items, product descriptions and author pages, would be added to future versions of PeP.

Beyond the bounded model, there are exciting opportunities to adapt PeP for use within the emerging library framework of 'integrated evaluated content' (Hamaker 1999). There is scope to apply more sophisticated reference linking and citation analysis tools developed in other projects (Brody, *et al*. 2001), and correspondence prompted by this work suggests the possibility of open linking services such as PeP participating in extended OpenURL services (section 8.6).

Keyword links in PeP are manually selected but could be enhanced by semantic linking based on a coordinated ontology. Examples that demonstrate semantic linking applied to journal-like collections include ScholOnto, an environment for scholarly discourse (Motta, *et al*. 2000) and OntoPortal, which 'projects a semantic meta-layer over existing Web resources' (Kampa, *et al*. 2001). Collaboration with these projects would be beneficial. There is great scope for enhancing the PeP link service to embrace semantic linking as well as adding to some of the native PeP linking features.

Speculatively, an underlying open eprint archive could be launched for papers relevant to PeP, but not exclusively for PeP. This might reduce authors' concerns over the management, storage and longer-term preservation of open access papers. As electronic versions begin to displace usage of printed journals, users are most concerned about the long-term maintenance of published works. As an electronic-only service, PeP is the type of product that most users are increasingly comfortable with, but does not contribute anything to the issue that concerns them most. The Open Archives initiative has yet to produce convincing evidence for institutional eprint archives against disciplinary archives such as arXiv. PeP could live with either outcome, but could usefully contribute to efforts to assure at least one open model prevails.

It would be intriguing to discover if the model and existing framework can be used by others to produce a series of PeP-like examples, e.g. *Perspectives in Digital Libraries*, which would overlap and complement PeP. A more open database framework would be the first requirement.

### *Does PeP inform or improve research?*

If by no other measure, PeP would have achieved its objective if it could satisfy this condition. To answer this question more convincingly, however, would require further evaluation of a version of PeP enhanced by the findings of the study reported here.

Despite efforts to improve the supporting technology and user interface, there is a more fundamental feature that will ultimately determine the success or otherwise of the model on which PeP is based, or indeed any electronic publishing model: *selection is key*. The influence of Google on the digital information environment again becomes evident. When any scholarly work can be exposed publicly and located instantly, what should be the basis for selection? This transcends peer review. A peer-reviewed journal can deny recognition but can no longer deny publication. Works are not uniformly good or bad, which peer review suggests. Authors and readers deserve a more open dialogue that enables works to be viewed in context. In an open information environment, is fully automated citation analysis, such as demonstrated by ResearchIndex, to be preferred to editorial selection?

PeP seeks to inform research by anticipating improved access to papers in many new contexts. Users anticipate the empowerment of electronic access, but not the obligations that will accompany it, the need to interpret and assess works rapidly, identifying new insights and making connections, and responding within a framework that builds on that connectivity.

Scholarly research is difficult. Pursuing scholarly research requires skills in managing information, and maintaining a comprehensive and up-to-date knowledge of all information pertaining to the specified research. In the emerging scholarly electronic information environment, in which it can be predicted that access to research papers will eventually become easier, these skills will not become redundant. Neither will services such as PeP make these skills redundant. Rather, easier access will place higher demands on these skills, as researchers will need to mine vast data sources more extensively, more forensically, seeking previously unidentified connections. PeP may not be a ‘journal’, or the ideal implementation or the perfect model for a service designed to assist the researcher in coping with these new demands. It is certain that services like it will not just be desirable, but necessary.

# References

**Ackerman, L. V. and Simonaitis, A.**, 1997, RSNA Electronic Journal: Beyond Paper Images: Radiology on the Web. *Journal of Electronic Publishing*, Vol. 3, No. 1, September http://www.press.umich.edu/jep/03-01/RSNA.html

**Adair, J. R.**, 1997, TC: A Journal of Biblical Textual Criticism: A Modern Experiment in Studying the Ancients. *Journal of Electronic Publishing*, Vol. 3, No. 1, September http://www.press.umich.edu/jep/03-01/TC.html

**Allard, S.**, 2001, Erasing the Barrier between Minds: Freeing Information, Integrating Knowledge. *American Communication Journal*, Vol. 4, No. 2, Winter http://www.acjournal.org/holdings/vol4/iss2/articles/allard.htm

**Amiran, E. and Unsworth, J.**, 1991, Postmodern Culture: Publishing in the Electronic Medium. *The Public-Access Computer Systems Review*, Vol. 2, No. 1, 67–76 http://info.lib.uh.edu/pr/v2/n1/amiran.2n1

**Anderson, K., *et al*.**, 2001, Publishing Online-Only Peer-Reviewed Biomedical Literature: Three Years of Citation, Author Perception, and Usage Experience. *Journal of Electronic Publishing*, Vol. 6, No. 3, March http://www.press.umich.edu/jep/06-03/anderson.html

**Arms, W. Y.**, 2000, Economic Models for Open Access Publishing. *Information Impacts***,** March http://www.cisp.org/imp/march_2000/03_00arms.htm

**Arms, W. Y., Blanchi, C. and Overly, E. A.**, 1997, An Architecture for Information in Digital Libraries. *D-Lib Magazine*, February http://www.dlib.org/dlib/february97/cnri/02arms1.html

**Austin, S.**, 1996, An Editor/User Perspective. *APS News: APS Online* (insert), November, 7

**Ayars, J., *et al*.** (eds), 2001, Synchronized Multimedia Integration Language (SMIL 2.0). *W3C Recommendation*, 7th August http://www.w3.org/TR/smil20/

**Bachrach, S., *et al*.**, 1998, Who Should Own Scientific Papers? *Science*, Vol. 281, No. 5382, 4 September, 1459–1460 http://www.sciencemag.org/cgi/content/full/281/5382/1459

**Bacon, R. A.**, 1994, Software Teaching of Modular Physics. *Computers in Physics Education*, September http://www.ph.surrey.ac.uk/stomp/papers/cpe2rev.htm

**Bailey, C. W. Jr**, 1994, Scholarly Electronic Publishing on the Internet, the NREN, and the NII: Charting Possible Futures. *Serials Review*, Vol. 20, No. 3, 7–16 http://info.lib.uh.edu/cwb/schpub.htm

**Balasubramanian, V., Bashian, A., and Porcher, D.,** 1997, A Large-Scale Hypermedia Application Using Document Management and Web Technologies. *Hypertext 97, the Proceedings of the eighth ACM conference on Hypertext* (New York, NY: Association for Computing Machinery), pp.134–145

**Barlow, J. P.**, 1994, The Economy of Ideas. *Wired*, 2.03, March http://wwww.wired.com/wired/2.03/features/economy.ideas.html

**Barron, D.**, 1989, Why use SGML? *Electronic Publishing*, Vol. 2, No. 1, April, 3–24

**Bauwens, M.**, 1996, The Three Laws of the Cyber-Economy. *Computer-Mediated Communication Magazine*, Vol. 3, No. 6, June http://www.december.com/cmc/mag/1996/jun/bauwens.html

**Beit-Arie, O., *et al*.**, 2001, Linking to the Appropriate Copy: Report of a DOI-Based Prototype. *D-Lib Magazine*, Vol. 7, No. 9, September http://www.dlib.org/dlib/september01/caplan/09caplan.html

**Bekhradnia, B.**, 1995, Pilot National Site Licence Initiative for Academic Journals. *Serials*, Vol. 8, No.3, November, 247–250

**Berger, A. and Smith, R**., 1999, New Technologies in Medicine and Medical Journals. *British Medical Journal*, Vol. 319, No. 7220, 13th November http://bmj.com/cgi/content/full/319/7220/0

**Berners-Lee, T. and Hendler, J.,** 2001, Scientific Publishing on the Semantic Web. *Nature*, 12th April, 2001 http://www.nature.com/nature/debates/e-access/Articles/bernerslee.htm

**Berners-Lee, T., *et al*.**, 1994, The World Wide Web. *Comm. ACM*, Vol. 37, No. 8, August, 76–82

**Bertocco, S**., 2001, Torii, an Open Portal over Open Archives. *HEP Libraries Webzine*, issue 4, June http://library.cern.ch/HEPLW/4/papers/4/

**Bieber, M., *et al*.**, 1997, Fourth Generation Hypermedia: Some Missing Links for the World Wide Web. *International Journal of Human-Computer Studies*, Vol. 47, 31–65 http://ijhcs.open.ac.uk/bieber/bieber.html

**Bjork, B.-C. and Turk, Z.**, 2000, How Scientists Retrieve Publications: An Empirical Study of How the Internet is Overtaking Paper Media. *Journal of Electronic Publishing*, Vol. 6, No. 2, December http://www.press.umich.edu/jep/06-02/bjork.html

**Borghuis, M**., ***et al***., 1996, *The TULIP final report* http://www.elsevier.nl/inca/homepage/about/resproj/tulip.shtml#FinalReport

**Bosak, J.,** 1997, XML, Java and the Future of the Web http://sunsite.unc.edu/pub/sun-info/standards/xml/why/xmlapps.htm

**Bosak, J**., 2001, Text markup and the cost of access. *Nature*, 28th August http://www.nature.com/nature/debates/e-access/Articles/bosak.html

**Boyce, P. B., Owens, E., and Biemesderfer, C.**, 1996, Electronic Publishing: Experience is Telling Us Something http://www.aas.org/~pboyce/epubs/sr-art.html

**Boyce, P. B**., 1998, Urania, a Linked, Distributed Resource for Astronomy. *Library and Information Services in Astronomy III*, ASP Conference Series, Vol. 153 http://www.eso.org/gen-fac/libraries/lisa3/boycep.html

**Bradley, D**., 2000, Preprint and be Damned. *The Alchemist*, 7th July http://www.chemweb.com/alchem/2000/catalyst/ct_000707_preprint.html

**Brailsford, D. F.**, 1994, Experience with the use of Acrobat in the CAJUN Publishing Project. *Proceedings of ECHT '94*, pp. 228–232

**Bray, T., *et al***. (eds), 2000, Extensible Markup Language (XML) 1.0 (Second Edition). *W3C Recommendation*, 6th October http://www.w3.org/TR/REC-xml

**Brody, T., *et al*.**, 2001, Enhancing OAI Metadata for Eprint Services: Two Proposals. *Experimental OAI-based Digital Library Systems Workshop*, Darmstadt, September http://opcit.eprints.org/ecdl-oai/oai-ecdl10.html

**Brown, C**., 2001, The Coming of Age of E-Prints in the Literature of Physics. *Issues in Science and Technology Librarianship*, Summer http://www.library.ucsb.edu/istl/01-summer/refereed.html

**Browning, J.**, 1996, Teleconomics: Information Marketeers Focus on Content rather than Access. *Scientific American*, February, 21–23

**Buck, A. M., *et al*.**, 1999, Scholar's Forum: A New Model For Scholarly Communication http://library.caltech.edu/publications/ScholarsForum/

**Buckingham Shum, S. and Sumner, T**., 2001, JIME: An Interactive Journal for Interactive Media. *First Monday*, Vol. 6, No. 2, February http://firstmonday.org/issues/issue6_2/buckingham_shum/

**Burg, J., *et al*.**, 2000, The State of the Art in Interactive Multimedia Journals for Academia. *Proceedings of ED-MEDIA 2000*, Vol. 1 (Association for the Advancement of Computing in Education), pp. 1–5 http://www.mthcsc.wfu.edu/~burg/papers/edmedia99.ps

**Bush, V.**, 1945, As We May Think. *Atlantic Monthly*, July, 101–108 http://www.isg.sfu.ca/~duchier/misc/vbush/

**Butler, D**., 1999, The Writing is on the Web for Science Journals in Print. *Nature*, Vol. 397, 21st January, 195–200

**Butler, D**., 2001a, The Future of the Electronic Scientific Literature. *Nature*, April http://www.nature.com/nature/debates/e-access/Articles/opinion2.html

**Butler, D**., 2001b, Budget Proposal Casts Doubt Over Physics Portal's Future. *Nature*, June http://www.nature.com/nature/debates/e-access/Articles/butler.html

**Cameron, R. D.**, 1997, A Universal Citation Database as a Catalyst for Reform in Scholarly Communication. *First Monday*, Vol. 2, No. 4 http://www.firstmonday.dk/issues/issue2_4/cameron/index.html

**Campbell, R**., 2001, Information Access: What is to be Done? *Nature*, 27th April http://www.nature.com/nature/debates/e-access/Articles/campbell.html

**Caplan, P. and Flecker, D**., 1999, Choosing the Appropriate Copy. *NISO News***,** September http://www.niso.org/DLFarch.html

**Carr, L. A.**, 1994, Structure and Hypertext. *PhD thesis, Southampton University* http://journals.ecs.soton.ac.uk/lacethesis/

**Carr, L., *et al*.,** 1995, The Distributed Link Service: a Tool for Publishers, Authors and Readers. *Proceedings of the Fourth World Wide Web conference*, Boston. http://www.w3.org/pub/Conferences/WWW4/Papers/178/

**Carr, L., *et al*.**, 1996, Open Information Systems. *Computer Networks and ISDN Systems*, Vol. 28, 1027–1036 http://www5conf.inria.fr/fich_html/papers/P12/Overview.html

**Carr, L., Hall, W. and Miles-Board, T**., 2000, Writing and Reading Hypermedia on the Web. *Southampton University Technical Report No. ECSTR-IAM00-1* http://www.bib.ecs.soton.ac.uk/data/3368/html/WRWH.html

**Case, M. M**., 1998, ARL Promotes Competition through SPARC: The Scholarly Publishing & Academic Resources Coalition. *ARL Newsletter*, No. 196 http://arl.cni.org/newsltr/196/sparc.html

**Clement, G.**, 1994, Evolution of a Species: Science Journals Published on the Internet. *Database*, October/November, 44–54

**Conklin, J.**, 1987, Hypertext: An Introduction and Survey. *IEEE Computer*, Vol. 20, No. 9, September, 17–41

**Crawford, W**., 2001, Getting Past the Arc of Enthusiasm. *Cites & Insights*, Vol. 1, No. 5, May, 4–8 http://home.att.net/~wcc.techx/civ1i5.pdf

**Davidson, L. and Douglas, K**., 1998, Digital Object Identifiers: Promise and Problems for Scholarly Publishing. *Journal of Electronic Publishing*, Vol. 4, No. 2, December http://www.press.umich.edu/jep/04-02/davidson.html

**Davis, H. C. *et al***., 1992, MICROCOSM: An Open Hypermedia Environment for Information Integration. *CSTR 92-15, Computer Science Technical Report, University of Southampton* http://www.bib.ecs.soton.ac.uk/records/1317

**Davis, H. C., Knight, S. and Hall, W**., 1994, Light Hypermedia Link Services: A Study of Third Party Application Integration. In *Proceedings of the 1994 European Conference on Hypermedia Technology*, Edinburgh http://www.bib.ecs.soton.ac.uk/data/1331/html/html/

**Delamothe, T., *et al***., 1999, Netprints: The Next Phase in the Evolution of Biomedical Publishing. *British Medical Journal*, Vol. 319, 11th December, 1515–1516 http://www.bmj.com/cgi/content/full/319/7224/1515

**Denning, P. J., and Rous, B.**, 1995, The ACM Electronic Publishing Plan. *Communications of the ACM*, Vol. 38, No. 4, April, 97–103 http://www.acm.org/pubs/epub_plan.html

**DeRose, S., Maler, E. and Orchard, D**., 2001, XML Linking Language (XLink) Version 1.0. *W3C Recommendation*, 27th June http://www.w3.org/TR/2001/REC-xlink-20010627/

**Dyson, E.**, 1994, Intellectual Property on the Net. *Release 1.0 report*, December http://www.eff.org/Intellectual_property/ip_on_the_net.html

**Edmonds, B**., 2000, A Proposal for the Establishment of Review Boards. *Journal of Electronic Publishing* Vol. 5, No. 4, June http://www.press.umich.edu/jep/05-04/edmonds.html

**Eisen, M. and Brown, P**., 2001, Should the Scientific Literature be Privately Owned and Controlled? *Nature*, 4th May http://www.nature.com/nature/debates/e-access/Articles/Eisen.htm

**Elliott, V**., 2001, National Site Licensing, *National Scholarly Communications Forum Round Table 12: Structures Of Scholarly Communication*, Canberra, August http://www.anu.edu.au/caul/nscf/elliott-nsl.doc

**Engelbart, D. C**., 1975, NLS Teleconferencing Features: The Journal, and Shared-Screen Telephoning. *IEEE Catalog* No. 75CH0988-6C, pp. 173–176 http://www.bootstrap.org/augment-33076.htm

**Ensor, P. and Wilson, T**., 1997, Public-Access Computer Systems Review: Testing the Promise. *Journal of Electronic Publishing*, Vol. 3, No. 1, September http://www.press.umich.edu/jep/03-01/pacs.html

**Entlich, R**, ***et al***., 1995, The Chemistry Online Retrieval Experiment: a Summary of the CORE Project (1991–1995). *D-Lib Magazine*, December http://www.dlib.org/dlib/december95/briefings/12core.html

**Festa, P**., 1998, Web Design Not What You Pay For. *CNet News.com*, 16th April http://news.com.com/2100-1023-210235.html?legacy=cnet&sas.mail

**Fillmore, L.**, 1993, Internet Publishing: How We Must Think. *Meckler's Internet World Conference*, New York, December http://www.press.umich.edu/jep/works/fillmore.think.html

**Flohr, U.**, 1995, Hyper-G Organizes the Web. *Byte*, Vol. 20, No. 11, November, 59, 60, 62, 64 http://www.byte.com/art/9511/sec5/art4.htm

**Fosmire, M. and Young, E**., 2000, Free Scholarly Electronic Journals: An Annotated Webliography. *Issues in Science and Technology Librarianship*, Fall http://www.library.ucsb.edu/istl/00-fall/internet.html

**Fosmire, M. and Yu, S**., 2000, Free Scholarly Electronic Journals: How Good Are They? *Issues in Science and Technology Librarianship*, Summer http://www.library.ucsb.edu/istl/00-summer/refereed.html

**Frank, M**., 2001, No Free Lunch! *Nature*, 20th August http://www.nature.com/nature/debates/e-access/Articles/frank.html

**Frazier, K**., 2001, The Librarians' Dilemma: Contemplating the Costs of the "Big Deal". *D-Lib Magazine*, Vol. 7, No. 3, March http://www.dlib.org/dlib/march01/frazier/03frazier.html

**Gannon, F**., 2001, Boycott! *Nature*, 5th April http://www.nature.com/nature/debates/e-access/Articles/gannon.html

**Gardner, T**., 2001, An Introduction to Web Services. *Ariadne*, issue 29, October http://www.ariadne.ac.uk/issue29/gardner/

**Gardner, W.**, 1990, The Electronic Archive: Scientific Publishing for the 1990s. *Psychological Science*, Vol. 1, No. 6, 333–341

**Garfield, E.**, 1955, Citation Indexes for Science: A New Dimension in Documentation Through Association of Ideas. *Science*, Vol. 122, 15th July, 108–111 http://www.garfield.library.upenn.edu/papers/science_v122(3159)p108y1955.html

**Garfield, E**., 1996, The Significant Scientific Literature Appears in a Small Core of Journals. *The Scientist*, 2nd September http://www.the-scientist.com/yr1996/sept/research_960902.html

**Ginsparg, P.**, 1994, First Steps Towards Electronic Research Communication. *Computers in Physics*, Vol. 8, No. 4, July/August, 390–396 http://xxx.lanl.gov/blurb/blurb.ps.Z

**Ginsparg, P.**, 1996, Los Alamos XXX APS ONLINE. *APS News: APS Online* (insert), November, 8 http://www.arxiv.org/blurb/sep96news.html

**Ginsparg, P**., 2001, Creating a Global Knowledge Network. *Proceedings of the Second ICSU-UNESCO International Conference on Electronic Publishing in Science*, Paris, February http://associnst.ox.ac.uk/%7Eicsuinfo/ginspargfin.htm

**Glass, G. V**., 2000, Anonymous FTP: Can We Risk It? Dare We Publish Pre-prints? *Annual meeting of the American Educational Research Conference*, New Orleans http://glass.ed.asu.edu/gene/papers/aera2000/

**Goldhaber, M. H.**, 1997, The Attention Economy and the Net. *First Monday*, Vol. 2, No. 4 http://www.firstmonday.dk/issues/issue2_4/goldhaber/

**Graham, G**., 2001, Cutting Through the Electronic Hype. *100-day Dialogue, Topic 3, What's Next For Publishing: Terminal Decline or Golden Age?* 27th July http://www.honco.net/100day/03/2001-0727-graham.html

**Green, D**., 1967, Death of an Experiment. *International Science and Technology*, Vol. 65, May, 82–88

**Guédon, J.-C**., 2001, In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing. *Creating the Digital Future, Association of Research Libraries, Proceedings of the 138th Annual Meeting*, Toronto, Ontario, May http://www.arl.org/arl/proceedings/138/guedon.html

**Guthrie, K. M**., 2000a, Challenges and Opportunities Presented by Archiving in the Electronic Era. *JSTOR Participants' Meeting*, January http://www.jstor.org/about/archiving.html

**Guthrie, K**., 2000b, Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR. *The Economics and Usage of Digital Library Collections (PEAK) conference*, Ann Arbor, Michigan, March http://www.si.umich.edu/PEAK-2000/guthrie.pdf

**Halasz, F. G.**, 1988, Reflections on NoteCards: Seven Issues for the Next Generation of Hypermedia Systems. *Communications of the ACM*, Vol. 31, No. 7, July, 836–852

**Halasz, F. G.**, 1991, 'Seven issues' Revisited. *Hypertext '91conference, closing plenary* http://www.parc.xerox.com/spl/projects/halasz-keynote/default.html

**Hall, W.**, 1994, Ending the Tyranny of the Button. *IEEE Multimedia*, Vol. 1, No. 1, Spring, 60–68

**Hall, W., and Davis, H.**, 1994, Hypermedia Link Services and their Application to Multimedia Information Management. *Information and Software Technology*, Vol. 36, No. 4, 197–202

**Hamaker, C.,** 1999, Integrated Evaluated Content. *Charleston Advisor*, Vol. 1, No. 2, October http://www.charlestonco.com/features.cfm?id=17&type=ed

**Harnad, S**., 1990, Scholarly Skywriting and the Prepublication Continuum Of Scientific Enquiry. *Psychological Science*, Vol. 1, No. 6, 342–343 http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad90.skywriting.html

**Harnad, S**., 1991, Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge. *Public-Access Computer Systems (PACS) Review*, Vol. 2, No.1, 39–53 http://info.lib.uh.edu/pr/v2/n1/harnad.2n1

**Harnad, S.**, 1995a, Premature polls. *VPIEJ-L Usenet newsgroup archive*, 14th August http://borg.lib.vt.edu/ejournals/vpiej-l/vpiej-l.log9508.html

**Harnad, S**., 1995b, Interactive Cognition: Exploring the Potential of Electronic Quote/Commenting. In *Cognitive Technology: In Search of a Humane Interface*, edited by B. Gorayska and J.L. Mey (Elsevier), pp. 397–414

**Harnad, S.**, 1996, Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals. *Electronic Publishing Confronts Academia: the Agenda for the Year 2000*, edited by R. Peek and G. Newby (Cambridge, MA: MIT Press) http://cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad96.peer.review.html

**Harnad, S**., 1998a, For Whom the Gate Tolls? *September98-Forum*, August 25th http://www.cogsci.soton.ac.uk/~harnad/Hypermail/Amsci/0000.html

**Harnad, S**., 1998b, The Invisible Hand Of Peer Review. *Nature*, 5th November http://www.nature.com/nature/webmatters/invisible/invisible.html

**Harnad, S**., 2000, Ingelfinger Over-Ruled: The Role of the Web in the Future of Refereed Medical Journal Publishing. *The Lancet Perspectives* 256, December (Supplement), s16 http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.lancet.htm

**Harnad, S**., 2001a, Why I Think Research Access, Impact and Assessment are Linked. *Times Higher Education Supplement*, No. 1487, 16 http://www.cogsci.soton.ac.uk/~harnad/Tp/thes1.html

**Harnad, S**., 2001b, Skyreading and Skywriting for Researchers: A Post-Gutenberg Anomaly and How to Resolve it. *Text-e Virtual Symposium*, November http://text-e.org/conf/index.cfm?fa=printable&ConfText_ID=7

**Harnad, S. and Carr, L**., 2000, Integrating, Navigating and Analyzing Eprint Archives Through Open Citation Linking (the OpCit Project). *Current Science Online*, Vol. 79, No. 5, 10th September http://tejas.serc.iisc.ernet.in/~currsci/sep102000/629.pdf

**Harnad, S., Carr, L. and Brody, T**., 2001, How and Why to Free All Refereed Research from Access- and Impact-Barriers Online, Now. *HEP Libraries Webzine*, issue 4, June http://library.cern.ch/HEPLW/4/papers/1/

**Harnad, S., Varian, H. and Parks, B**., 2000, Academic Publishing in the Online Era: What Will Be For-Fee and What Will Be For-Free? *Culture Machine*, issue 2 http://culturemachine.tees.ac.uk/Cmach/Backissues/j002/Articles/art_harn.htm

**Harter, S. P., and Kim. H. J.**, 1996, Electronic Journals and Scholarly Communication: A Citation and Reference Study. *Proceedings of the ASIS Midyear Meeting*, San Diego, CA, May, pp. 299–315 http://ezinfo.ucs.indiana.edu/~harter/harter-asis96midyear.html

**Hellman, E. S**., 1999, The S-Link-S Framework for Reference Linking: Architecture and Implementation. *Proceedings of the ICCC/IFIP conference Electronic Publishing '99. Redefining the Information Chain - New Ways and Voices*, edited by John W.T.Smith, Anders Ardö and Peter Linde (Washington D.C.: ICCC Press)

**Hellman, E**., 2001a, Less is More! Re: Reader-Designated HyperLinking In/Between/Among E-Journals. *Web4Lib Electronic Discussion Archive*, 13th February http://sunsite.berkeley.edu/Web4Lib/archive/0102/0218.html

**Hellman, E**., 2001b, Building a Database for E-journals. *Web4Lib Electronic Discussion Archive*, 17th October http://sunsite.berkeley.edu/Web4Lib/archive/0110/0175.html

**Hill, G**, *et al*., 1997, Exploiting Serendipity Among Users to Provide Support for Hypertext Navigation. *Hypertext 97, the Proceedings of the eighth ACM conference on Hypertext* (New York: Association for Computing Machinery), pp. 212–213

**Hitchcock, S.,** 1996, Web Publishing: Speed Changes Everything. *IEEE Computer*, Vol. 29, No. 8, August, 91–93

**Hitchcock, S**., 2001a, Concise Harnad Reader on Scholarly Electronic Publishing. *PeP Review*, February http://aims.ecs.soton.ac.uk/pep.nsf/special/harnad?opendocument

**Hitchcock, S**., 2001b, Napster: Implications for Journal Publishers. *PeP Commentary*, February http://aims.ecs.soton.ac.uk/pep.nsf/special/napster?opendocument

**Hitchcock, S. and Hall, W**., 2001, How Dynamic E-journals can Interconnect Open Access Archives: Introducing Perspectives in Electronic Publishing – conference slides. *ICCC/IFIP 5th Conference on Electronic Publishing ELPUB'01*, Canterbury, UK, July http://www.ecs.soton.ac.uk/~sh94r/elpub-slides/title.html

**Hitchcock, S., Carr, L. and Hall, W.**, 1996, A Survey of STM Online Journals: The Calm Before the Storm. *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, edited by D. Mogge, 6th edition (Washington, D.C.: Association of Research Libraries), pp. 7-32 http://journals.ecs.soton.ac.uk/survey/survey.html

**Hitchcock, S., Carr, L. and Hall, W**., 1997a, Web Journals Publishing: A UK Perspective. *Serials*, Vol. 10, No. 3, November, 285–299 http://journals.ecs.soton.ac.uk/uksg.htm

**Hitchcock, S., *et al*.,** 1997b, Citation Linking: Improving Access to Online Journals. *Proceedings of the Second ACM international conference on Digital Libraries*, edited by Robert B. Allen and Edie Rasmussen (New York: Association of Computing Machinery), pp. 115–122 http://journals.ecs.soton.ac.uk/acmdl97.htm

**Hitchcock, S., *et al*.**, 1998a, Towards Universal Linking for Electronic Journals. *Serials Review*, Vol. 24, No. 1, Spring, 21–33 http://journals.ecs.soton.ac.uk/IFIP-SerRev98.html

**Hitchcock, S., *et al*.**, 1998b, Linking Electronic Journals: Lessons from the Open Journal Project. *D-Lib Magazine*, December http://www.dlib.org/dlib/december98/12hitchcock.html

**Hitchcock, S., *et al*.**, 2000, Developing Services for Open Eprint Archives: Globalisation, Integration and the Impact of Links. *5th ACM Conference on Digital Libraries*, San Antonio, Texas, June http://opcit.eprints.org/dl00/dl00.html

**Hitchcock, S. M**., 1997, Progress in Online Publishing: the Case for Remodelling Journals for a New Medium. *University of Southampton, Mini-thesis*, December

**Holoviak, J. and Seitter, K. L.**, 1997, Earth Interactions: Transcending the Limitations of the Printed Page. *Journal of Electronic Publishing*, Vol. 3, No. 1, September http://www.press.umich.edu/jep/03-01/EI.html

**Hopkins, C.**, 2001, Healthy Warning: 'This journal supports full text, tariff-free archives". *Nature*, 18th September http://www.nature.com/nature/debates/e-access/Articles/hopkins.html

**Hughes, G., *et al*.**, 1999, Academic Information Management: an Open Linking Approach. *Webnet 99 World Conference on the WWW and Internet 10*, pp. 526–531

**Hugo, J. and Newell, L.**, 1991, New Horizons in Adult Education: The First Five Years (1987–1991). *The Public-Access Computer Systems Review*, Vol. 2, No. 1, 77–90 http://info.lib.uh.edu/pr/v2/n1/hugo.2n1

**Hunter, K.**, 1998, Adding Value by Adding Links. *Journal of Electronic Publishing*, Vol. 3, No. 3, March http://www.press.umich.edu/jep/03-03/hunter.html

**Hunter, P.**, 2001, The Management of Content: Universities and the Electronic Publishing Revolution. *Ariadne*, issue 28, June http://www.ariadne.ac.uk/issue28/cms/

**Inger, S.**, 2001, The Importance of Aggregators. *Proceedings of the Second ICSU-UNESCO International Conference on Electronic Publishing in Science*, Paris, February http://associnst.ox.ac.uk/%7Eicsuinfo/ingerfin.htm

**Instone, K.**, 1997, Usability Heuristics for the Web. *Web Review* http://www.webreview.com/1997/10_10/strategists/10_10_97_2.shtml

**Jacsó, P.**, 2001, Electronic Shoes for the Cobbler's Children: Treatment of Digital Journals in Library and Information Science Databases. *Online*, July http://www.onlinemag.net/OL2001/jacso7_01.html

**Jennings, E. M.**, 1991, EJournal: An Account of the First Two Years. *The Public-Access Computer Systems Review*, Vol. 2, No. 1, 91–110 http://info.lib.uh.edu/pr/v2/n1/jennings.2n1

**JHEP Executive Office**, 2000, The Journal of High Energy Physics: A Successful Experience of Electronic Publishing. *High Energy Physics Libraries Webzine*, issue 1, March http://library.cern.ch/HEPLW/1/papers/3/

**Johnson, R. K.**, 2000, A Question of Access: SPARC, BioOne, and Society-Driven Electronic Publishing. *D-Lib Magazine***,** Vol. 6, No. 5, May http://www.dlib.org/dlib/may00/johnson/05johnson.html

**Kahan, J. *et al*.**, 2001, Annotea: An Open RDF Infrastructure for Shared Web Annotations. *WWW10 conference*, Hong Kong, May http://www.www10.org/cdrom/papers/488/index.html

**Kampa, S., *et al*.**, 2001, Linking with Meaning: Ontological Hypertext for Scholars. *University of Southampton* http://www.bib.ecs.soton.ac.uk/data/5163/PDF/lwm.pdf

**Kappe, F., Maurer, H., and Sherbakov, N.**, 1992, Hyper-G – a universal hypermedia system. *HCM Technical Report, Graz University of Technology, Austria*

**Kaser, R. T.**, 2001, When Allegory Replaces Rational Thought, Science Had Better Watch Out. *Nature*, 18th September http://www.nature.com/nature/debates/e-access/Articles/kaser.html

**Katz, J.**, 1994, Online or Not, Newspapers Suck. *Wired*, Vol. 2.09, September http://www.wired.com/wired/archive/2.09/news.suck.html

**Kiernan, V**., 1999, Why Do Some Electronic-Only Journals Struggle, While Others Flourish? *Journal of Electronic Publishing*, Vol. 4, No. 4, June http://www.press.umich.edu/jep/04-04/kiernan.html

**Kirriemuir, J**., 1999, A Brief Survey of Quality Resource Discovery Systems, version 2 (final). *Resource Discovery Network Centre*, September http://www.rdn.ac.uk/publications/studies/survey/

**Kirstein, P. and Montasser-Kohsari, G**., 1996, The C-ODA Project: Online Access to Electronic Journals. *Communications of the ACM*, Vol. 39, No. 6, June, 87–99

**Kling, R. and Covi, L.**, 1995, Electronic Journals and Legitimate Media in the Systems of Scholarly Communication. *The Information Society*, Vol. 11, No. 4, 261–271 http://www.ics.uci.edu/~kling/klingej2.html

**Kling, R. and McKim, G**., 2000, Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science*, Vol. 51, No. 14 http://xxx.lanl.gov/abs/cs.CY/9909008

**Koch, T**., 2000, Quality-Controlled Subject Gateways: Definitions, Typologies, Empirical Overview. *Online Information Review*, Vol. 24, No. 1, February http://www.lub.lu.se/~traugott/OIR-SBIG.txt

**Koman, R**., 1998, The Scent of Information: Helping Users Find Their Way by Making Your Site "Smelly". *Web Review*, 15 May http://www.webreview.com/1998/05_15/strategists/05_15_98_1.shtml

**Krichel, T. and Warner, S**., 2001a, Disintermediation of Academic Publishing Through the Internet: An Intermediate Report from the Front Line. *ICCC/IFIP 5th Conference on Electronic Publishing*, Canterbury, UK, July http://openlib.org/home/krichel/sants.html

**Krichel, T. and Warner, S. M**., 2001b, A Metadata Framework to Support Scholarly Communication. *DC-2001, Proceedings of the International Conference on Dublin Core and Metadata Applications*, Tokyo, October http://www.nii.ac.jp/dc2001/proceedings/product/paper-22.pdf

**Krol, E.**, 1994, *The Whole Internet: User's Guide and Catalog*, second edition (Sebastopol, CA: O'Reilly & Associates)

**Lagoze, C. and Van de Sompel, H**., 2001, The Open Archives Initiative: Building a low-barrier interoperability framework. *Joint Conference on Digital Libraries*, Roanoke, VA http://www.cs.cornell.edu/lagoze/papers/oai-final.pdf

**Lancaster, F.W**., 1985, The Paperless Society Revisited. *American Libraries*, Vol. 16, No. 8, September, 553

**Landow, G.**, 1992, *Hypertext: The Convergence of Contemporary Critical Theory and Technology* (Baltimore, MD: Johns Hopkins University Press)

**Langer, J**., 2000, Physicists in the New Era of Electronic Publishing. *Physics Today*, Vol. 53, No. 8, Part 1, August http://www.aip.org/pt/vol-53/iss-8/p35.html

**Lassila, O. and Swick, R. R**. (eds), 1999, Resource Description Framework (RDF) Model and Syntax Specification. *W3C Recommendation*, 22nd February http://www.w3.org/TR/REC-rdf-syntax/

**Lawrence, S**., 2001, Free Online Availability Substantially Increases a Paper's Impact. *Nature*, 31st May http://www.nature.com/nature/debates/e-access/Articles/lawrence.html

**Lawrence, S., Giles, C. L. and Bollacker, K**., 1999, Digital Libraries and Autonomous Citation Indexing. *IEEE Computer*, Vol. 32, No. 6, 67–71 http://www.neci.nj.nec.com/~lawrence/papers/aci-computer98/

**Leiner, B. M.**, *et al*., 1997, The Past and Future History of the Internet. *Communications of the ACM*, Vol. 40, No. 2, February, 102–108

**LeJeune, L**., 1999, Who Owns What? *Journal of Electronic Publishing*, Vol. 4, No. 3, March http://www.press.umich.edu/jep/04-03/glos0403.html

**Lessig, L**., 2001, A Commitment to Candid Speech. *The Filter*, No. 4.4, 30th August http://cyber.law.harvard.edu/filter/lessig-sklyarov.html

**Lie, H. W., and Bos, B.**, 1997, *Cascading Style Sheets: Designing for the Web* (Harlow, UK: Addison-Wesley). Web updates http://www.awl.com/css/

**Liu, X., *et al***., 2001, Arc – An OAI Service Provider for Digital Library Federation. *D-Lib Magazine*, Vol, 7, No. 4, April http://www.dlib.org/dlib/april01/liu/04liu.html

**Luce, R. E**., 2001, E-prints Intersect the Digital Library: Inside the Los Alamos arXiv. *Issues in Science and Technology Librarianship*, Winter, 2001 http://www.library.ucsb.edu/istl/01-winter/article3.html

**Lucier, R. E., and Brantley, P.**, 1995, The Red Sage Project: An Experimental Digital Journal Library for the Health Sciences. *D-Lib Magazine*, August http://www.dlib.org/dlib/august95/lucier/08lucier.html

**Lugar, L. and Thomes, K**., 2001, Access to Scholarly Communication Information on ARL Member Library Websites. *SPARC E-News*, October-November http://www.arl.org/sparc/core/index.asp?page=g21#6

**Lynch, C**., 1999, On the Threshold of Discontinuity: The New Genres of Scholarly Communication and the Role of the Research Library. *ACRL National Conference*, Detroit, April http://www.ala.org/acrl/clynch.html

**Lynch, C**., 2001a, The Battle to Define the Future of the Book in the Digital World. *First Monday*, Vol. 6, No. 6, June http://firstmonday.org/issues/issue6_6/lynch/

**Lynch, C. A**., 2001b, Metadata Harvesting and the Open Archives Initiative. *ARL Bimonthly Report*, No. 217, August http://www.arl.org/newsltr/217/mhp.html

**Maclennan, B**., 1999, Presentation and Access Issues for Electronic Journals in a Medium-Sized Academic Institution. *Journal of Electronic Publishing* Vol. 5, No. 1, September http://www.press.umich.edu/jep/05-01/maclennan.html

**Malcolm, K. C., Poltrock, S. E., and Schuler, D.,** 1991, Industrial Strength Hypermedia: Requirements for a Large Engineering Enterprise. *Hypertext '91 Proceedings* (New York: Association for Computing Machinery), pp. 13–24

**Maly, K., Zubair, M. and Liu, X**., 2001, Kepler – An OAI Data/Service Provider for the Individual. *D-Lib Magazine*, Vol, 7, No. 4, April http://www.dlib.org/dlib/april01/maly/04maly.html

**Margolis, M. and Resnick, D**., 1999, Third Voice: Vox Populi Vox Dei? *First Monday* Vol. 4, No. 10, October http://firstmonday.org/issues/issue4_10/margolis/

**Marshall, C. C., and Shipmann, F. M. III**, 1997, Spatial Hypertext and the Practice of Information Triage. *Hypertext 97, the Proceedings of the eighth ACM conference on Hypertext* (New York: Association for Computing Machinery), pp.124–133

**Marshall, C.**, 1997, Looking Forward: Five Practices for Safer Hypertext (summary of keynote paper). *Hypertext 97, the Proceedings of the eighth ACM conference on Hypertext* (New York: Association for Computing Machinery), p. 241

**Martin, D**., 2000, DSSSL for XML: Why Not? *XML.com*, 2nd May http://www.xml.com/pub/a/2000/05/03/dsssl/index.html

**Maurer, H.** (Ed.),1996, *HyperWave – The Next Generation Web Solution* (Harlow, UK: Addison-Wesley)

**McCabe, M. J**., 1999, The Impact of Publisher Mergers on Journal Prices: An Update. *ARL Newsletter*, No. 207 http://www.arl.org/newsltr/207/jrnlprices.html

**McKiernan, G**., 2001, ePROFILE: Perspectives in Electronic Publishing: An Open Access-Dynamic-Virtual Electronic Journal. *Library Hi Tech News*, Vol. 18, No. 9, 2001

**McKiernan, G**., 2002, E is for Everything: The Extra-Ordinary, Evolutionary [e-] Journal. *The Serials Librarian*, to be published

**McLuhan, M.**, 1964, *Understanding Media, The Extensions of Man*, reissued 1994 (Cambridge, MA: MIT Press)

**Mermin, N. D.,** 1991, Publishing in Computopia**.** *Physics Today*, May, 9, 11

**Mermin, N. D.**, 1992, What's Wrong in Computopia? *Physics Today*, April, 9, 11

**Meyrowitz, N.**, 1989, The Missing Link: Why We're All Doing Hypertext Wrong. *The Society of Text: Hypertext, Hypermedia and the Social Construction of Information* (Cambridge, MA: MIT Press), pp. 107–114

**Miller, E., *et al*.,** undated, Semantic Web: Introduction. *W3C overview* http://www.w3.org/2001/sw/

**Minton, S. and Wellman, M. P**., 1999, JAIR at Five. *AI Magazine*, May http://www.isi.edu/sims/minton/papers/jairfive.pdf

**Moret, B. M. E**., 1997, ACM's Journal Of Experimental Algorithmics: Bridging the Gap Between Theory and Practice. *Journal of Electronic Publishing*, Vol. 3, No. 1, September http://www.press.umich.edu/jep/03-01/JEA.html

**Mossberg, W. S**., 2001 New Windows XP Feature Can Re-Edit Others' Sites. *Wall Street Journal*, 7th June http://ptech.wsj.com/archive/ptech-20010607.html

**Motta, E., Buckingham Shum, S. and Domingue, J.,** 2000, Ontology-Driven Document Enrichment: Principles, Tools and Applications. *International Journal of Human-Computer Studies*, Vol. 52, No. 6, June, 1071–1109 http://kmi.open.ac.uk/projects/scholonto/docs/OntoDocEnrich-IJHCS-2000.pdf

**Nabe, J**., 2001, E-Journal Bundling and its Impact on Academic Libraries: Some Early Results. *Issues in Science and Technology Librarianship*, Spring http://www.library.ucsb.edu/istl/01-spring/article3.html

**Naylor, B**., 1995, In *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*, edited by Okerson, A. and O'Donnell, J. (Washington, DC: Association of Research Libraries), chapter XIV http://www.arl.org/scomm/subversive/sub14.html

**Naylor, B., and Geller, M.**, 1995, A Prehistory of Electronic Journals: the EIES and Blend projects. In *Advances in Serials Management*, ed. M. Tuttle, Vol. 5 (Greenwich, CT: JAI Press), pp. 27–47

**Negroponte, N.**, 1995, *Being Digital* (London: Coronet)

**Nelson, T. H.**, 1987, *Literary Machines*, edition 87.1

**Nelson, T. H**., 1999, Xanalogical Media: Needed Now More Than Ever: Parallel Documents, Deep Links to Content, Deep Versioning, and Deep Re-Use. *ACM Computing Surveys*, Vol. 31, No. 4, December http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/60.html

**Nielsen, J**., 1994a, How to Conduct a Heuristic Evaluation http://www.useit.com/papers/heuristic/heuristic_evaluation.html

**Nielsen, J**., 1994b, Ten Usability Heuristics http://www.useit.com/papers/heuristic/heuristic_list.html

**Nielsen, J**., 1997, Be Succinct! (Writing for the Web). *Alertbox*, March 15 http://www.useit.com/alertbox/9703b.html

**Nielsen, J**., 1998, Bill Gates' Shopping List to Build the Internet Desktop. *Alertbox*, 13th December http://www.useit.com/alertbox/981213.html

**Nielsen, J**., 2001, The End of Homemade Websites. *Alertbox*, 14th October http://www.useit.com/alertbox/20011014.html

**Nurnberg, P. J., Leggett, J. J., and Schneider, E. R.,** 1997, As We Should Have Thought. *Hypertext 97, the Proceedings of the eighth ACM conference on Hypertext* (New York: Association for Computing Machinery), pp.96–101

**O'Leary, D. E.**, 1997, The Internet, Intranets, and the AI renaissance. *IEEE Computer*, Vol. 30, No. 1, January, 71–78

**O'Reilly, T.**, 1996, Publishing Models for Internet Commerce. *Communications of the ACM*, Vol. 39, No. 6, June, 79–86 http://www.ora.com/oracom/inet/pubmod.html

**Oder, N**., 2002, The New Wariness. *Library Journal*, 15th January http://libraryjournal.reviewsnews.com/index.asp?layout=article&articleid=CA188739

**Odlyzko, A.**, 1995, Tragic Loss or Good Riddance? The Impending Demise of Traditional Scholarly Journals. *International Journal of Human-Computer Studies* (formerly I*nt. J. Man-Machine Studies*) Vol. 41 http://www.research.att.com/~amo/doc/tragic.loss.txt

**Odlyzko, A**., 1999, Competition and Cooperation: Libraries and Publishers in the Transition to Electronic Scholarly Journals. *Journal of Electronic Publishing***,** Vol. 4, No. 4, June http://www.press.umich.edu/jep/04-04/odlyzko0404.html

**Okerson, A.**, 1991, The Electronic Journal: What, Whence, and When? *Public-Access Computer Systems Review*. Vol. 2, No. 1, 5–24 http://info.lib.uh.edu/pr/v2/n1/okerson.2n1

**Okerson, A.**, 1992, Synopsis. In *University Libraries and Scholarly Communication: A Study Prepared for The Andrew W. Mellon Foundation* (Washington, D.C.: Association of Research Libraries) http://etext.lib.virginia.edu/reports/mellon/synopsis.html

**Okerson, A. and O'Donnell, J.**, (eds), 1995, *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing* (Washington, DC: Association of Research Libraries) http://www.arl.org/scomm/subversive/toc.html

**Oppenheim, C**., 1996, Moral Rights and the Electronic Library. *Ariadne*, issue 4, July http://www.ariadne.ac.uk/issue4/copyright/

**Overbye, D**., 2001, An Online Archive with Mountain Roots. *New York Times*, 28th August http://www.nytimes.com/2001/08/28/science/physical/28PAUL.html

**Pearl, A.**, 1989, Sun's Link Service: A Protocol for Open Linking. *Hypertext 89, the Proceedings of the second ACM conference on Hypertext* (New York: Association for Computing Machinery), pp. 137–146

**Peek, R**., 2001, The Tangled Scholarly Publishing Mess. *Information Today*, Vol. 18, No. 4, April http://www.infotoday.com/it/apr01/peek.htm

**Pentz, E**., 2001a, CrossRef: A Collaborative Linking Network. *Issues in Science and Technology Librarianship*, Winter http://www.library.ucsb.edu/istl/01-winter/article1.html

**Pentz, E**., 2001b, Evolution and Revolution: Pragmatism versus Dogmatism. *Nature*, 28th August http://www.nature.com/nature/debates/e-access/Articles/pentz.html

**Phelps, C. E**., 1998, Achieving Maximal Value from Digital Technologies In Scholarly Communication. *ARL Proceedings*, October http://www.arl.org/arl/proceedings/133/phelps.html

**Phillips, M**., 2001, Review of 'Carlson, Scott. "JSTOR's Journal-Archiving Service Makes Fans of Librarians and Scholars"'. In *Current Cites*, Vol. 12, No. 7, July http://sunsite.berkeley.edu/CurrentCites/2001/cc01.12.7.html

**Piternick, A. B**., 1989, Attempts to Find Alternatives to the Scientific Journal: A Brief Review. *Journal of Academic Librarianship*, November, 260–266

**Powell, A**., 2001, OpenResolver: a Simple OpenURL Resolver. *Ariadne*, issue 28, June http://www.ariadne.ac.uk/issue28/resolver/

**Powell, A. and Lyon**, L., 2001, The DNER Technical Architecture: Scoping the Information Environment. *Distributed Systems Group, UKOLN*, 18th May http://www.ukoln.ac.uk/distributed-systems/dner/arch/dner-arch.html

**Press, L.**, 1995, McLuhan Meets the Net. *Communications of the ACM*, Vol. 38, No. 7, July, 15–20 http://som1.csudh.edu/fac/lpress/mac.htm

**Public Library of Science**, 2000, Open Letter. *publiclibraryofscience.org*, undated http://www.publiclibraryofscience.org/plosLetter.htm

**Quinn, F**., 1995, Roadkill on the Electronic Highway: The Threat to the Mathematical Literature. *Notices of the American Mathematical Society*, Vol. 42, No. 1, January http://www.ams.org/notices/199501/forum.pdf

**Rambler, M**., 1999, Do It Yourself? A New Solution to the Journals Crisis. *Journal of Electronic Publishing*, Vol. 4, No. 3, March http://www.press.umich.edu/jep/04-03/rambler.html

**Randle, Q**., 2001, A Historical Overview of the Effects of New Mass Media Introductions on Magazine Publishing During the 20th Century. *First Monday*, Vol. 6, No. 9, September http://firstmonday.org/issues/issue6_9/randle/

**Raney, R. K**., 1998, Into a Glass Darkly. *Journal of Electronic Publishing*, Vol. 4, No. 2, December http://www.press.umich.edu/jep/04-02/raney.html

**Richardson, M**., 2001, Impacts of Free Access. *Nature*, April http://www.nature.com/nature/debates/e-access/Articles/richardson.html

**Rizk, A., and Sauter, L.**, 1992, Multicard: An Open Hypermedia System. *Proceedings of the European conference on Hypertext, ECHT '92* (New York: ACM), pp. 4–10

**Roberts, R. J. *et al***., 2001, Building A "GenBank" of the Published Literature. *Science*, Vol. 291, No. 5512, 23rd March, 2318–2319 http://www.sciencemag.org/cgi/content/full/291/5512/2318a

**Roes, H.**, 1995, Electronic journals: A Survey of the Literature and the Net. *Journal of Information Networking*, Vol. 2, No. 3, 169 –186 http://cwis.kub.nl/~dbi/users/roes/articles/ej_join.htm

**Rous, B**., 1999, ACM: a Case Study. *Journal of Electronic Publishing*, Vol. 4, No. 4, June http://www.press.umich.edu/jep/04-04/rous.html

**Russo, E**., 2001, New Adventures in Science Publishing: Efforts of Grassroots Group Fall Short; New Online Journal in the Works. *The Scientist*, Vol. 15, No. 21, 29th October http://www.the-scientist.com/yr2001/oct/index_011029.html

**Rzepa, H. S. and Murray-Rust, P**., 2001, A New Publishing Paradigm: STM Articles as part of the Semantic Web. *Learned Publishing*, Vol. 14, No. 3, July, 177–182 http://www.catchword.com/cgi-bin/cgi?ini=alpsp&body=linker&reqidx=/catchword/alpsp/09531513/v14n3/s3/p177

**Sanville, T. J**., 2001, Use of Electronic Journals in OhioLINK'S Electronic Journal Center. *67th IFLA Council and General Conference*, Boston, August http://www.ifla.org/IV/ifla67/papers/118-141e.pdf

**Savage, L.**, 1991, The Journal of the International Academy of Hospitality Research. *The Public-Access Computer Systems Review*, Vol. 2, No. 1, 54–66 http://info.lib.uh.edu/pr/v2/n1/savage.2n1

**Schauder, D.**, 1994, Electronic Publishing of Professional Articles: Attitudes of Academics and Implications for the Scholarly Communication Industry. *Journal of the American Society for Information Science*, Vol. 48, No. 2, 73–100

**Senders, J. W.**, 1977, An Online Scientific Journal. *Information Scientist*, Vol. 1, No. 1, 3–9

**Senge, P. M.**, 1990, *The Fifth Discipline: the Art and Practice of the Learning Organization* (USA: Doubleday)

**Shackel, B.**, 1991, Blend-9: Overview and Appraisal. *British Library Research Paper No. 82*

**Sheriden, T., *et al.***, 1981, Experimentation With a Multidisciplinary Teleconference and Electronic Journal on Mental Workload. Unpublished report to the NSF, June

**Shirky, C**., 2001, Web Services: It's so Crazy, It Just Might Not Work. *XML.com*, 3rd October http://www.xml.com/pub/a/2001/10/03/webservices.html

**Shulenburger, D. E**., 1998, Moving With Dispatch to Resolve the Scholarly Communication Crisis: From Here to NEAR. *ARL Proceedings*, October http://www.arl.org/arl/proceedings/133/shulenburger.html

**Shulenburger, D. E**., 2001, Principles for a New System of Publishing for Science. *Proceedings of the Second ICSU-UNESCO International Conference on Electronic Publishing in Science*, Paris, February http://associnst.ox.ac.uk/%7Eicsuinfo/shulenburgerfin.htm

**Singh, R., *et al***., 1998, Interactive Multi-Media Journals in Science and Technology: Prospects and Issues. *Ejournal*, Vol. 8, No. 2, August http://www.ucalgary.ca/ejournal/archive/v8/immj3.htm

**Smith, A**., 2000, The Journal as an Overlay on Preprint Databases. *Learned Publishing*, Vol. 13, No. 1, January http://giorgio.catchword.com/vl=95473899/cl=14/nw=1/rpsv/cgi-bin/cgi?ini=alpsp&body=linker&reqidx=/catchword/alpsp/09531513/v13n1/s6/p43

**Smith, J. W. T**., 1999, The Deconstructed Journal – a New Model for Academic Publishing. *Learned Publishing*, Vol. 12, No. 2, April http://tamino.catchword.com/vl=5968037/cl=52/nw=1/rpsv/cgi-bin/cgi?ini=alpsp&body=linker&reqidx=/catchword/alpsp/09531513/v12n2/s3/p79

**Smith, P. N., *et al***., 1993, Journal Publishing with Acrobat: The CAJUN Project. *Electronic Publishing-ODD*, Vol. 6, No. 4, December, 1–13 http://cajun.cs.nott.ac.uk/compsci/epo/papers/samples/ep6x4pns.pdf

**Spink, A.,** 1998, Modeling Users' Successive Searches in Digital Environments. *D-Lib Magazine*, April http://www.dlib.org/dlib/april98/04spink.html

**Story, G. A., *et al***., 1992, The RightPages Image-Based Electronic Library for Alerting and Browsing. IEEE *Computer*, September, 17–25

**Suber, P**., 2001, Editorial Position of the FOS Newsletter, version 1.6 http://www.earlham.edu/~peters/fos/index.htm#scope

**Taubes, G.**, 1993, E-mail Withdrawal Prompts Spasm. *Science*, Vol. 262, No. 5131, 8th October, 173

**Taubes, G.**, 1996a, Electronic Preprints Point the Way to 'author empowerment'. *Science*, Vol. 271, No. 5250, 9th February, 767

**Taubes, G.**, 1996b, Speed of Publication – Stuck in First Gear. *Science*, Vol. 271, No. 5250, 9th February, 765

**Tenopir, C. and King, D. W**., 2000, Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers. *Psycoloquy*, Vol. 11, issue 084 http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?11.084

**Tenopir, C. and King, D. W**., 2001, Lessons for the Future of Journals. *Nature*, 18th October http://www.nature.com/nature/debates/e-access/Articles/tenopir.html

**Till, J. E**., 2001, Predecessors of Preprint Servers. *Learned Publishing*, Vol. 14, No. 1, 7–13 http://cherubino.catchword.com/vl=93366182/cl=6/nw=1/rpsv/catchword/alpsp/09531513/v14n1/s2/p7

**Turoff, M., and Hiltz, S. R.**, 1982, The Electronic Journal: A Progress Report. *Journal of the American Society for Information Science*, Vol. 33, July, 195–202 http://eies.njit.edu/~turoff/Papers/ElectronicJournal.html

**Tuttle, M**., 1991, The Newsletter on Serials Pricing Issues. *The Public-Access Computer Systems Review*, Vol. 2, No. 1, 111–127 http://info.lib.uh.edu/pr/v2/n1/tuttle.2n1

**Ubell, R**., 1997, Missing Bodies: Scientists Access Data – They Don't Read Bodies of Text. *HMS Beagle: The BioMedNet Magazine*, No. 6, 18th April; archived 2nd May http://news.bmn.com/hmsbeagle/06/people/pressbox.htm

**Valauskas, E. J**., 1997, Waiting for Thomas Kuhn: First Monday and the Evolution of Electronic Journals. *Journal of Electronic Publishing*, Vol. 3, No. 1, September http://www.press.umich.edu/jep/03-01/FirstMonday.html

**Van de Sompel, H. and Beit-Arie, O**., 2001a, Open Linking in the Scholarly Information Environment Using the OpenURL Framework. *D-Lib Magazine*, Vol. 7, No. 3, March http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html

**Van de Sompel, H. and Beit-Arie, O**., 2001b, Generalizing the OpenURL Framework beyond References to Scholarly Works: The Bison-Futé Model. *D-Lib Magazine*, Vol. 7, No. 7/8, July/August http://www.dlib.org/dlib/july01/vandesompel/07vandesompel.html

**Van de Sompel, H. and Hochstenbach, P**., 1999, Reference Linking in a Hybrid Library Environment, Part 2: SFX, a Generic Linking Solution. *D-Lib Magazine*, Vol. 5, No. 4, April http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html

**Van de Sompel, H., *et al***., 2000, The UPS Prototype: An Experimental End-User Service across E-Print Archives. *D-Lib Magazine*, Vol. 6, No. 2, February http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html

**Varmus, H., *et al***., 1999, Journals Online: PubMed Central and Beyond. *HMS Beagle: The BioMedNet Magazine*, No. 61, 3rd September http://news.bmn.com/hmsbeagle/61/viewpts/page1

**Varmus, H**., 1999, E-BIOMED: A Proposal for Electronic Publications in the Biomedical Sciences. *National Institutes of Health*, May 5 http://www.nih.gov/about/director/ebiomed/53ebio.htm

**Walker, T. J**., 1998, Free Internet Access to Traditional Journals. *American Scientist*, Vol. 86, No. 5, September-October http://www.amsci.org/amsci/articles/98articles/walker.html

**Warren, A**., 1995, Microcosm: Creating an Application. *Connections, The Journal of Southampton University Computing Services*, Vol. 6, No. 2, February http://www.soton.ac.uk/~connect/v6i2/mcosm3.html

**Weibel, S**., 1995, Metadata: The Foundations of Resource Description. *D-Lib Magazine*, July http://www.dlib.org/dlib/July95/07weibel.html

**Weintraub, J**., 1998, The Development and Use of a Genre Statement for Electronic Journals in the Sciences. *Issues in Science and Technology Librarianship*, Winter http://www.library.ucsb.edu/istl/98-winter/article5.html

**Wells, A**., 1999, Exploring the Development of the Independent, Electronic, Scholarly Journal. *MSc in Information Management, Sheffield University* http://panizzi.shef.ac.uk/elecdiss/edl0001/index.html

**Wheary, J., *et al***., 1998, Living Reviews In Relativity: Thinking and Developing Electronically. *Journal of Electronic Publishing*, Vol. 4, No. 2, December http://www.press.umich.edu/jep/04-02/wheary2.html

**Wiederhold, G**., 1995, Digital Libraries, Value and Productivity. *Communications of the ACM*, Vol. 38, No. 4, 85–96 http://www-db.stanford.edu/pub/gio/1995/acmfinal.ps

**Wiil, U. K., and Leggett, J. J**., 1997, Workspaces: The HyperDisco Approach to Internet Distribution. *Hypertext '97 Proceedings* (New York: ACM), pp. 13–23

**Wilkinson, S. L**., 1998, Electronic Publishing Takes Journals into a New Realm: Publications Slip Off Restrictions of Print World and Carve Out a Unique Identity. *Chemical & Engineering News*, 18th May

**Wilson, T**., 1997, "In the Beginning was the Word..." Social and Economic Factors in Scholarly Electronic Communication. *Journal of Electronic Publishing*, Vol. 3, No. 2, December http://www.press.umich.edu/jep/archive/wilson.html

**Woodward, H., *et al***., 1998, Café Jus: An Electronic Journals User Survey. *Journal of Digital Information*, Vol. 1, No. 3, September http://jodi.ecs.soton.ac.uk/Articles/v01/i03/Woodward/

**Wusteman, J.,** 1997, The Evolving Use of SGML in Electronic Journals. *ICCC/IFIP conference on Electronic Publishing '97: New Models and Opportunities*, Canterbury, UK, April

**Yankelovich, N., Meyrowitz, N., and van Dam, A.**, 1985, Reading and Writing the Electronic Book. *IEEE Computer*, Vol. 18, No. 10, October, 15–30

**Young, R**., 1999, Giving It Away: How Red Hat Software Stumbled Across a New Economic Model and Helped Improve an Industry. *Journal of Electronic Publishing*, Vol. 4, No. 3, March http://www.press.umich.edu/jep/04-03/young.html

**Youngen, G**., 1997, Citation Patterns of the Physics Preprint Literature with Special Emphasis on the Preprints Available Electronically. *University of Illinois at Urbana-Champaign Physics and Astronomy Library* http://www.physics.uiuc.edu/library/preprint.html