Linking electronic journals

Lessons from the Open Journal project

Steve Hitchcock
Les Carr
Wendy Hall
Steve Harris
Multimedia Research Group
Department of Electronics and Computer Science
University of Southhampton

Steve Probets
David Evans
David Brailsford
Electronic Publishing Research Group
Department of Computer Science
University of Nottingham

Correspondence for all authors should go to Steve Hitchcock, *sh94r@ecs.soton.ac.uk*.

Introduction

The Open Journal project [1] has completed its three year period of funding by the UK Electronic Libraries (eLib) programme (Rusbridge 1998). During that time, the number of journals that are available electronically leapt from a few tens to a few thousand. Some of these journals are now developing the sort of features the project has been advocating, in particular the use of links within journals, between different primary journals, with secondary journals data, and to non-journal sources. Assessing the achievements of the project and considering some of the difficulties it faced, we report on the different approaches to linking that the project developed, and summarise the important user responses that indicate what works and what does not. Looking ahead, there are signs of change, not just to simple linking within journals but to schemes in which links are the basis of "distributed" journals, where information may be shared and documents built from different sources. The significance has yet to be appreciated, but this would be a major change from printed journals. If projects such as this and others have provided the initial impetus, the motivation for distributed journals comes, perhaps surprisingly, from within certain parts of the industry, as the paper shows.

Journals: adapting to the Web

Many aspects of publishing are being transformed by the arrival of the World Wide Web and its facility to distribute information electronically. For journals, the transformation has barely begun. A large number of journals are now available in electronic form of some kind, but otherwise little has changed. Recently, one leading publisher reportedly described the growth of electronic journals based on the portable document format (PDF) as the "first frontier" facing journal publishers just a couple of years ago. The second frontier is the emergence of links, he went on to say. Judged by this measure alone, journals may have appeared on the Web but have yet to adapt to it. Publishing may have changed, but readers probably experience little difference between Web-delivered journals and, in many cases, their print originals. New and better features that can be supported electronically are now being demanded by users, however, supported by librarians, as the price of acceptance of these new products (Tenopir and Ennis 1998, Weintraub 1998).

Of these new features, links are one of the most important. Since 1995, the Open Journal project has been applying original software tools and techniques to support flexible linking in e-journal applications, based on selected journals which were available electronically but which were are not all exclusively electronic. In doing so, the project foretold the impact that links will have. Links are not a superficial feature of the Web, nor are they simple add-on features for e-journals. Links have the power to alter the character of journals fundamentally, most obviously in the development of "distributed publishing" in which users can find items of interest irrespective of the publisher (Dixon 1998). Ultimately, distributed publishing may transform the way in which individual documents are compiled by sharing components or "objects", figures say, from different sources, and by using network-based software processes, or services, to enhance presentation.

Links are important for a number of reasons:

- For users, links provide faster, more direct access to more information.

- For librarians, links support more effective information retrieval, especially from large archives, and can help with identified user phenomena such as "successive search episodes" (Spink *et al*. 1998).

- For publishers, links add value to works (Hunter 1998), but in this context links need to be applied in particular ways to make it easier to maintain and manage large numbers of documents.

For a limited period, the Open Journal project made selected demonstrator Open Journals, which link journals and associated resources that are distributed across different locations on the Web, openly accessible, reporting the first user testing and evaluation of the use of links within journals on a large scale (Open Journal Project 1998). Although the project's

funded development has finished, the work continues in a different form. So this is a review of work in progress, but it presents a critical evaluation of the main results from the project, especially with respect to user responses.

The project's impact has also been marked by its reporting of e-journal developments more generally (Hitchcock *et al.* 1996, 1997a). Informed by these findings and the reported experiences of other users and publishers, elsewhere we assess the future for e-journals more broadly than for the Open Journal approach alone, asking how we can make the most of e-journals, again with the user perspective principally in mind. 2

The Open Journal Project

In the beginning

The Open Journal project was launched with three main objectives (Open Journal Project 1995):

- to provide immediate access to electronic versions of existing journals (noting that this was a time when there were few e-journals)
- to provide powerful hypermedia linking techniques
- to support faster access to information, including journals and other Web resources

Of these, the second and third were the most important. The project did not intend to become a journal producer or publisher. These objectives were largely achieved, although with some qualifications which will be explored below. The project also expressed a longer-term vision of link-based journals: "For publishers the "journal" now becomes a set of links: substitute links for the glue of the paper journal" (Hitchcock 1996). The project produced some extensive demonstrators, notably of citation linking, but it has not been possible yet to produce a convincing demonstrator of how journals might be fundamentally "re-packaged" in this way. We were lacking an important component and experience gained from the project shows us what this is, as we explain below.

The Open Journal approach

In practical terms, the Open Journal approach involves four components, respectively, linking, resources, formats and location:

1. applying new linking tools, including a link service and a software-based citation "agent"
2. ... to selected primary journals and secondary resources
3. ... in HTML and PDF (and prospectively the Extensible Markup Language, or

XML)
4. ... distributed on the Web.

The link service used in the project was developed as a research tool at Southampton University and was based on technology that has been commercialised, under other names, for other applications. It quickly became apparent from project users that whatever you call the underlying linking tool, it is the functionality that counts, and that functionality must be presented to users as transparently as possible. The functionality, and the novel twist in this case, is that the links are added to documents that have been requested by users *as they are downloaded*. Up-to-date links are published within the document upon delivery rather than authored in typical HTML style because they are served by the link service separately from the document source and added on the way to the client via a proxy server (Carr *et al. 1998b*).

Resources of a wide variety of types, from primary journals to databases, were generously provided by a group of twelve publisher partners (see the Appendix) most of which supported the project throughout its full period and some of which continue to work with the project developers.

Initially, the functionality provided by the link service extended linking capabilities to documents in any format (Carr *et al*. 1995), although for journals in the project it became clear that this could be relaxed to the two main formats above. There was discussion about including capabilities for pages in TeX, which is possible, and this might be important for subject areas with more mathematical content, but this was beyond the scope of the project. The linking tools will be compliant with the linking component of XML, an important new format for Web documents that emerged during the latter months of the project (Carr *et al*. 1998a). Full compliance will be formalised following publication of the XLink and XPointer standards.

Recognising that it will never be possible to serve the information needs of users from single resources or single Web sites, a vital feature of this approach was that links could be applied to documents *wherever* they are on the Web -- the distributed journal scenario again.

Demonstrators and linking

Three demonstrator journals were produced in the subject areas of cognitive science, biology and computer science. These areas and the core journal resources contained within the respective Open Journals were effectively defined by the interests of the research teams and the publishers that supported the project. Development was further informed by subject specialists, including researchers, teachers and librarians, in each of the areas. For

the purposes of this analysis, however, it is more convenient to consider the demonstrator Open Journals from the viewpoint of the linking features they highlighted (Table 1). The three main link types in this application can best be described as citation links, keyword links, and PDF links.

Table 1. Open Journal demonstrators and linking features

Open Journal	Citation linking	Keyword linking	PDF linking	Release status
Cognitive Science	Yes			Open release; closed end May 1998
Biology		Yes	Yes	Released to selected evaluators
Computer Science	Yes	Yes	Yes	Internal project release

Citation linking

Citation links act on references contained within papers. Print journals are constrained to work backwards in time, citing earlier works; electronically, links can also point forward in time, to later papers that have cited the one being viewed, or even to later papers that have cited any of the references contained within that paper. The influences on a work and its own influence can thus be tracked (Garfield 1955). How completely a thread can be tracked in this way depends on the number of citation links that can be presented, which in turn depends on the scale of the data that can be linked. Given the small size today of the electronic full-text archive, citation linking on a meaningful scale requires a secondary database of abstracts.

The project was fortunate to be able to work with data provided by the Institute for Scientific Information (ISI) from its citation indexes. These indexes not only include abstracts from papers but the references too, and this is the basis of its distinctive services, now translated to the company's Web of Science (Hitchcock *et al*. 1998a). Working with a much smaller, but still substantial (500 MB), data set than Web of Science, the project demonstrated forward and backward linking within the secondary data but went further, extending the linking capability to remote but accessible full-text journals. For those journals **it was shown that references could be linked -- where the data set allows; in this example the data set was not big enough for comprehensive linking -- to the secondary data or, potentially, to other full-text journals independently of the established authoring and publication process**.

An important and successful component of citation linking was a software-based **citation agent** developed for use in the project. This agent recognises reference data within a downloaded paper, matches the citations against a pre-indexed database of abstracts, and links the references to entries in the database where matches are found. All of these actions are performed in "real time" as the requested paper is downloaded. Conceived as an autonomous processor, the agent was partitioned as a "library" of functions to allow it to be integrated with other programs. In the project, this approach was used to enable the agent to be used with the link service and PDF software. In continuing post-project work with publishers, described below, the citation agent is the common feature of the planned applications.

Some other research projects are developing tools for citation linking. While the Open Journal project emphasised text recognition, matching, and linking, other projects are concerned with software agents that can find cited works on the Web (Han *et al*. 1997), and improved search services and parsing of other document formats (e.g., PostScript) to build automatic and comprehensive citation indexes (Giles *et al*. 1998). Another approach to serving links separately from documents is Hyper-G, an electronic publishing package based on a Web server with an object-oriented distributed network database and a separate link database (Schmaranz 1996). Hyper-G has been used in journal projects with Springer-Verlag and Academic Press.

In one sense, serving links in this way is a rough-and-ready but practical method for implementing relationships, represented explicitly as links, between documents from different sources. This is a flexible approach but because no control is exercised over the documents, the links can be unstable and need to be rigorously mantained. At the other end of the spectrum, an ideal application-independent and stable way of identifying documents or their components might be the Digital Object Identifier (Davidson and Douglas 1998). Publishers to have prototyped applications represented in the the gallery of the International DOI Foundation include Academic Press, Elsevier, Springer-Verlag, and Wiley. The DOI is not yet an accepted standard, and to become so would require wide agreement. While links simply provide access to works, however, the DOI has a more demanding remit: "The intent of the (Association of American Publishers) AAP's Enabling Technologies Committee (which designed the DOI) was to support copyright protection, while ameliorating inconvenience to users, by supporting technology that promotes interoperability' (Rosenblatt 1997). It is not clear that these objectives can be easily reconciled without compromising user access. There are also concerns about possible limitations, such as restricting organizations that are permitted to assign DOIs to "legitimate" publishers.

Between the DOI and link serving is Hellman's (1998) proposal for the Scholarly Link Specification Framework (SLinkS) which applies DOI-like identifiers to documents from different publishers but controlled via an intermediary service.

It is already clear from a number of publishing arrangements that the electronic scholarly literature will be dominated by cross-linking on citations between different journals and services. ISI Links has been announced as its means of mediating citation linking between Web of Science, collaborating publishers, and subscribing institutions. Linking applications where links are applied between different journals and documents directly managed by a single publisher have been described for the BioMedNet service (in Hitchcock *et al*. 1998b), HighWire Press (Rubinstein 1997) and the Institute of Physics (Dixon 1998).

Some Web-based abstracts services enable third-party users to create links to entries in these sites. The best known is the National Library of Medicine's Medline service, the basis of widespread citation linking in biomedical fields. NCBI Citation Matcher allows users to find the Medline ID of any article in the database, given its bibliographic information, and to use that ID in a URL to retrieve the record. A related development, the PubMed project, additionally links back out from Medline entries to full-texts on the servers of cooperating publishers.

The Astrophysics Data System Abstract Service also helps with bibliographic code querying to link directly to abstracts from outside the abstract service.

Keyword linking

Keyword linking is more contentious than citation linking. Keywords are a means of classification that does not automatically translate to linking on the same scale as citation linking. Keywords are often created by the authors of papers, and therefore may not always be compiled systematically. Publisher and editor intervention can help by formalising the terms selected, especially in structuring the terms in a hierarchy. As a result, however, indiscriminate linking based on keywords can produce results which appear not to match authorial intent and produced, early in the project's first demonstrations, some adverse reactions from users.

Given the ubiquity of keywords within quality journals, and the frequency with which terms might appear given their position in the classification hierarchy, it became a relatively simple task programmatically to develop and display large numbers of keyword links (Hitchcock *et al*. 1998b). This creates problems for users, in this case the classic problem of information overload through too many links. Links appear to be random and are not well labelled; in other words, users are unsure where a link will take them or whether following the link will be useful.

Links to dictionaries or glossaries produced from keywords can be more intuitive and

useful, but even here the response was equivocal. Dictionaries are often assumed to be low-level texts, useful for novices and students. Specialists do not want to see links to dictionaries in research texts, as was the case in the Open Journal of Biology. Ironically, the dictionary linked in this case, the *Dictionary of Cell Biology*, is produced by and for specialists and, being available on the Web, is the most up-to-date resource of its kind in a fast-moving field in which new terms are being created constantly. For users, it seems that the benefit of even a well-labelled link is determined by their knowledge of the source being linked.

It would be easy to dismiss keyword linking on this evidence, but the opportunity for new, informed perspectives implemented through this type of link is worth pursuing. These links will almost always appear within the body of a text, not at a point that can be conveniently extracted. Citation links alone, while vital and useful and the obvious next step for e-journals, will in the long term be insufficient as a way of identifying relationships between texts or of creating new perspectives. There will be some technical refinements to the linking framework, and work continues to develop tools which enable users to reduce the number of links displayed and to apply links more selectively (Carr *et al*. 1998b). This is more of a culture gap than a technology gap, however, which on the part of the author, or link author, requires a better understanding of text structures (Renear 1997) and of the relationships between texts; and on the user side, requires more experience of this type of linking and raised expectations through better implementations.

One way forward, instead of using ready keywords, is to look again at texts for the occurrence of what might be called link words, quite a different concept from keywords and which are created with a better understanding of the linking strategy. With new demands being placed on e-journals, perhaps link words will become an editorial task as common as creating keywords today.

On a smaller scale, it would be possible to use this technique effectively within single journals, where the effect of keyword linking would be to overlay the journal index as links on the electronic text archive.

PDF linking

PDF linking is different again, concerned with the technical issue of managing data in a particular format rather than the content of that data. Of course, the content could be a citation link or a keyword link. Most electronic journals are in PDF (Hitchcock *et al.* 1997a) so it was natural for the project to treat this as an important linking issue. Although Adobe designed PDF to include links, links are not a primary feature of the format as they are within HTML, and it is notable that few PDF journals have links. Among the few that have citation links, some such as the American Physical Society and the Institute of Physics in the UK have reformatted reference data into HTML for the purpose. There may

be good reason for this. PDF may be chosen for ease of conversion from print sources, but once in that form the data is complex and difficult to manipulate in the way required by the project's approach to linking on a large-scale, an approach which is more sophisticated than linking supported intrinsically by the format.

It required a major effort, but the project produced a working service for linking from PDF ([Probets *et al*. 1998](#)). It remains to be seen whether the dominance of PDF prevails for e-journals, but since it is more cost-effective to work with given formats rather than convert then PDF linking could be an important tool. Although converting references to HTML is one alternative, it does not address the need for links, keyword links perhaps when the implementation is refined, within the body of papers.

There is, however, the danger that the PDF tools will have to be updated every time there is a change to the linking framework, as happened during the project, or each time Adobe changes the specification for PDF or the way in which it supports the format. In principle, because Open Journal linking is a server-side process ([Carr *et al*. 1998b](#)), this approach is both independent of the platform used and of the version of a given software application on the user's machine. In practice, communication between Adobe client and server software has been version dependent -- there are incompatibilities between Adobe Acrobat Reader versions 2 and 3, for example.

Adobe PDF may be a *de facto* standard as far as e-journals are concerned, but it is still a proprietary format, and this is not ideal in an environment such as the Web which promotes the use of open, public standards that are intended to allow improved interoperability between applications. This might be rectified by plans for a vector graphics standard for the Web, and moves toward "structured" PDF, that are intended to improve PDF to HTML and Web interchange.

Summary of link types

Table 2 summarises the application of different link types in the project, showing what they were used for, what worked, and which areas require further investigation.

<p align="center">Table 2. Link types demonstrated in the Open Journal project</p>

Link type	Use	For	Against

Citation linking	• linking on references in papers, backward *and* forward, in time	• builds on established journal authoring procedures • already a prominent feature of advanced online journals	• needs to be comprehensive • large data sources required
Keyword linking	• indexing links for individual journals or collections of journals • links to dictionaries or glossaries	• works in small-scale, well-defined cases with careful editorial control	• can produce too many links • difficult to control • difficult to describe or label link destinations • difficult to manage user expectations • authors wary of links superimposed on text
PDF linking	• enables Open Journal links, say citation or keyword links, to be added to documents in PDF format in the same way as they are for HTML	• most online journals are in PDF, so this *could* be important for publishers of PDF journals	• difficult to maintain • dependent on support for proprietary software - may be version dependent • complex and expensive in terms of computing resources

The user view

User responses were gathered by the project in a number of ways: meetings with individual

users, selected evaluation groups, and open access demonstrators, mediated by electronic mail and Web forms; and less directly through meetings with other interested groups, conference presentations, seminars, and workshops. Evaluators included specialists in the subject areas covered, as well as teachers, librarians and publishers. Below is a summary of the important points to have emerged. The comments added for illustration, and those shown additionally for illustration in Table 3, are abstracted from user returns, via a Web questionnaire, from the open access demonstration of citation linking in the Open Journal of Cognitive Science during April and May 1998. This Open Journal comprised two primary journals and the database of abstracts provided by ISI (Hitchcock *et al*. 1998a, 1997b). The full test results for this demonstrator are presented in the project's final report to its funding body, JISC and the Electronic Libraries programme (Open Journal Project 1998).

Important results from user testing

- **Users are very demanding.**
  - "It would have to be much faster."
  - "Speed of links -- initial link to an abstract, have all the abstracts loaded with the initial page"
- **Citation links are popular, and users will always want more links.**
  - When services are simple, intuitive and useful, it is hard to overstate the impact, even for new users. The Web as a whole is one example. Citation linking will be another case in point. Before our link service existed there was great anticipation among e-journal advocates. By the time the project was, briefly, able to release its demonstrator openly, real users were more sanguine. Why so few links? (It was due to limited data sets.) The demand for links in this context will grow *very rapidly*.
  - "Bigger and better please."
  - "Make it BIGGER!!!"
  - "I wish that a visual cue could be used to discriminate between links to the reference section of the article alone, and those citations for which an abstract could be found. I was a bit frustrated by wanting to look up an abstract, interrupting my reading to find it, only to learn that it was unavailable."
- **Link direct.**
  - "One could imagine even making the link go directly to the abstract/full-text, rather than first to the list of references at the end of the paper."
  - "Add a tiny little document icon next to those links for which there are abstracts or full-text available, and make the little icon jump directly to the abstract/full-text."
- **Technology must be transparent: users want better services without having to install new software or change computer settings.**

It was common practice for Web users, especially back in 1995 when the project began, to download and install software from the Web to improve the Web experience. With the exception of Acrobat it seems the practice of software download does not apply to typical e-journal users, as the project soon discovered from publishers and librarians. So the link service software was rewritten to work at the server end, mediated by a user-set browser proxy. Even this was insufficient. Libraries do not want settings on shared machines to be altered, and proxy settings can interfere with firewalls in corporate environments. In the latest version the direction to the proxy server is attached to the URL, leaving the user to browse the Web conventionally and do nothing to receive the link service, barring starting from the right place!

○ "I don't know what "the proxy" is."

- **Links help navigation but users also need orientation.**

  It was stated above that the project missed something important. The project was first and foremost about building and applying links, and the links appear in others' texts. So the project did not concentrate on building custom journal, or Open Journal, interfaces which, it was felt, might have detracted from the identities of the contributing journals and publishers. The interface is the user's starting point, and Web interface design has advanced enormously in the last year or two, especially for large resources, where so-called "gateways" have leapt in importance. An Open Journal needs not just links but a distinctive entrance, or gateway, too.

  ○ "It was not easy to discover how to start into the archive."
  ○ "Took a while to see what was on offer."
  ○ "Without additional navigation aids it is easy to lose track of where you are in the cluster."

- **Reliability is a critical issue.**

  ○ "Unable to access. "

- **Links must be clearly labelled and unambiguous: users are suspicious of unexpected links in texts.**

Table 3. Illustrative user responses for and against citation linking in the Open Journal of Cognitive Science

"It's a great service!"	"It is a WONDERFUL idea. However..."

- "The potential is clear and exciting. … Thanks for letting us see how the future might look!"
- "An excellent way to trace ideas and how the scientific community has reacted to them."
- "the forward search aspect is terrific."
- "I'd like to have such a link structure for EVERYTHING I read."
- "Looks like a wonderful way to find info that sometimes is elusive using keyword searches of databases."

- "It is a WONDERFUL idea. However, (for) two separate target articles … it simply didn't work, and was slow doing it."
- "Powerful tool which better exploits the capability of the medium … (but) not fully implemented."
- "I could have obtained the same information (and in a more controlled way since I would have been deciding where and when to search) if I simply had one window looking at an item and had another window open into BIDS/ISI and used the citation search feature."
- "The search engine needs to be improved before a serious trial can be initiated. The single keyword restriction made the output set way too big."

- "It would be a good idea to have an opportunity for marking citations."

Building better Open Journals

If the project has ended, the ideas have not. Some of the work continues to be developed within different frameworks. The linking tools may be commercialised, and specific link publishing projects will proceed with some of the publishers that participated in the project. A question that must be asked before continuing is: could we have produced better Open Journals?

Yes, they could have been better. Mostly this answer is informed by results, as a development project should be, but other elements might have been foreseen. The Open Journals would have been better if we had:

- access to more resources
- built better user interfaces

- stable, open systems technology for the Web, links, document formats, etc.
- more link editing tools

The first two points are concerned with structural journal issues and publishing; the other two points with the technology framework, especially the Web and links. Both environments altered significantly during the project's three years, as already described, and affected the project work. While the project was clear about its approach to linking, the definition of an Open Journal was altogether more arbitrary, an idea based on linking, or integrating, information elements from different journal sources selectively and flexibly. The formation of Open Journals was to be left to evolution and experience, and therefore more vulnerable to changing circumstances.

This could be discussed at length, but of wider importance, given the results reported here, is whether the concepts demonstrated in the project might be more broadly applicable and enduring. There are two reasons why they might be:

- **There will be greater integration of journal resources.** This will be between journals and other database sources on the Web and is already happening, as discussed above. Ultimately it may require a sophisticated gateway and comprehensive access to resources. The stakes are *very* high, and there are a number of possible competing claims for this prize, but there is no evidence yet of a convincing strategy or a dominant player in this respect.

- **Link services will become more important.** HTML is evolving towards the more sophisticated XML, which includes a framework for linking. One component of this framework, XLink, governs how you insert links into your XML document, in principle supporting the use of link services within the proposed XLink standard. XML became a recommended standard early in 1998 (W3C 1998), although the linking components are not yet included in that standard.

Link services could emerge strongly when the technology framework for Web publishing becomes more stable. Essentially what is required are:

- open systems services -- some link services are, but PDF is not and the Web is "open" only in some respects;
- based on open standards -- the Web is but link services are not.

For publishers, it is never too late to switch to open system services based on open standards.

Links can be created programmatically in large numbers, but results suggest that more precision and control over the presentation of links is needed. To assist, link editing tools

have to be further developed to enable authors, editors and other content and information service developers, as well as programmers, to manage links for different applications.

Publishers: the way forward

The publishing background

Since the project began it has prominently reported the phenomenal growth of journals available on the Web. During this period, the number of scholarly refereed e-journals has jumped from around 100 (Hitchcock *et al.* 1996), measurably towards 1300, prospectively 3000 (Hitchcock *et al.* 1997a), with perhaps more than 5000 worldwide according to some speculation. 3 Estimates indicate that more than 80 per cent of e-journals are presented in PDF (Hitchcock *et al.* 1997a), which is the *de facto* standard for e-journals reproduced from print. It is against this background that the project's link application must be considered. If the second frontier is the emergence of links, it has not been breached yet. A new path may be being mapped out, however, in post-project applications with publishers.

Post-project plans of publishers

A number of applications will continue to be investigated with individual publishers. The dilemma for these applications, whether to continue with a static format such as PDF or to migrate to a more flexible format for Web publishing, is highlighted by two contrasting observations of the project developers:

- PDF proved to be quite a complicated format for the project to work with. The format is optimised for display; links are specified with respect to page coordinates, not to words or phrases, and in-depth knowledge of the PDF file format is required to access and create the necessary structures to add links around words and phrases.

- Webmasters in some established publishing environments are reticent about using databases of external links as a fundamental navigation resource. Instead they prefer to produce pre-linked documents from bespoke database scripts.

Clearly the latter has a bearing on immediate applications because it reflects existing publisher cultures and practices. Most of the continuing publisher projects involve working with PDF. These projects are supported by the original project developers but with a view to the publishers taking control and applying the tools themselves, and all the projects have one thing in common: they do not yet plan to use the full capability of the link service but instead will initially use a cut-down version of the tools:

- the text recognition properties of the bibliography agent to extract citation data

from PDF journals.
- the agent technology to match extracted citations within data sources.
- the link service to insert links in static pre-processed PDF pages.
- work is ongoing with a JISC-funded digitisation project to build link trails in a large PDF archive.

All this suggests that e-journals are not yet ready to become distributed, dynamically changing resources in a native Web format rather than PDF. This could change quickly, however. Publishers, now more familiar with working in multiple media, are attracted to the reusability features of the SGML format. Journal publishers in particular are formatting selected elements of published papers, notably header data, in SGML, and some have plans to generate HTML for the Web from an SGML original (Hitchcock *et al*. 1997a).

First examples of reformatting to support linking include conversion of reference sections to HTML, as demonstrated by the Institute of Physics' e-journals service. This service, which uses PDF to present papers, links citations from extracted HTML to a database of abstracts held by the publisher.

Alternatively, publishers may be tempted towards the Web-based successor to SGML, XML, if it delivers more cost-effective production, particularly if e-journals can generate their own independent income streams to support this development. Potentially, XML, with its linking components, offers significantly more native capability for linking applications than does PDF, or even HTML, but it is not yet widely used.

There is belated recognition that e-journals must offer more than the printed equivalent, and citation linking will be the first example. There are a number of possible effects. As more data is shared, how will it be managed, by whom and where? As shared data sources become larger, will static linking be adequate in a fast-changing, expanding data environment?

Widening the boundaries for e-journals

Although the application of project technology was not without its difficulties, and the demonstrators need to be followed up by real applications, a bigger barrier to adoption of an Open Journal approach remains cultural. The cultural shift required to embrace it is perhaps best highlighted by a description of the advantages of using a link service:

- Although the link service approach may seem over-complicated, an advantage is that links can be applied directly to citations in any documents, from any publisher, not just those over which the user has editorial control.

Allowing others to exercise some control over already-published materials would appear to

run counter to publishing ethos. This may be one reason why project publishers appear reluctant to unleash the full link service on their Web journals in the follow-up projects. Internationally there are political and legal moves to strengthen the exercise of rights and control over data published on computer networks. Yet nobody can know the impact that universal adoption of the Web, a uniquely user driven service, will have as a communications medium.

It is possible that anarchic users will stretch the limits of what is acceptable, but the motivations for change can be seen even among established publishers, who invariably have limited access to Web users. What if a publisher could extend its reach by placing links to its works *directly* into other services, library services for example? How could these links be maintained, updated and managed? Is it possible that respected publishers might want to do this, interact with other services?

One is. The Institute of Physics' Stacks service -- "the ultimate linking service" -- generates tables of contents (TOCs) with embedded hyperlinks, and is aimed at librarians, other publishers, aggregators, abstracting and indexing services and producers of information gateways. In contrast to the project's link service which can manage link inclusion in Web pages independently of the data creator, Stacks delivers TOCs and link data via email or file transfer for implementation by the local service provider.

Is this a more practical approach than the project has applied, more likely to appeal to publisher needs, or is it simply more limited and less flexible? Whichever, **an important principle has been demonstrated by two developments independently: data, not just computers, are becoming perpetually more distributed on the Web. No data provider can survive alone.** Data will be shared and interactive, and not just at the user level. Again, this is recognised in the XML initiatives ([Khare and Rifkin 1997](#)). The sooner this is more widely recognised, the more likely that established cultures can begin to change and efforts can be directed towards building an online information environment in which new opportunities to serve users can flourish, rather than trying to constrain this environment by imposing other publishing models.

The legacy of the Open Journal project may eventually be commercial applications built by publishers and supported by commercial tools[4] first tested in the project. Perhaps a broader legacy will be to have contributed to developments leading towards distributed data, by motivating the user benefits at a time when the prevailing culture, especially among information providers, was difficult to reconcile with the emerging needs.

Notes

[1] The Open Journal project was funded in the UK by JISC's Electronic Libraries Programme (eLib) award ELP2/35. More information about the project is available at

http://journals.ecs.soton.ac.uk.

2 This paper has been developed from a presentation given at a one-day seminar *Making the most of e-journals* in April 1998, at Loughborough University, UK, organised jointly by the UK Online User Group (UKOLUG) and the UK Serials Group. To get the complete story from that presentation, this paper can be read in conjunction with [a paper published elsewhere](#) in which we draw a broader picture of the needs of e-journals, not just from the Open Journal perspective, and how we might make more of them. We discover that the capabilities which it seems are most widely desired remain limited by the prevailing framework for commercial journals publishing. Some non-prescriptive solutions to this problem are suggested. [The slides from the Loughborough talk](#) are also available.

3 Personal correspondence with Ann Okerson in January 1998. Based on NewJour data published and unpublished at that time, Ann said: "If there is any way to put numbers on this ejournal movement, I would say that 5,000 is very conservative -- but that 5,000 would be 'real' journals and that number will be sky high a year from now."

4 A version of the link service software is available from [Multicosm Ltd](#), although it does not currently support journal applications as developed in the project. Negotiation continues with the company with a view to commercialising the link service for publishers, possibly with the additional components built by the project. This process will be informed by demand from publishers, particularly those experimenting with their own applications.

References

Carr, L., Davis, H. C., De Roure, D. and Hall, W. (1998a) Application-independent link processing. *Seventh International World Wide Web Conference*, Brisbane, Australia, April   http://www.elsevier.nl:80/cas/tree/store/comnet/free/www7/1883/com1883.htm

Carr, L., De Roure, D., Hall, W. and Hill, G. (1998b) Implementing an Open Link Service for the World-Wide Web. *World Wide Web*, Vol. 1, No. 2 , 61-71 http://www.staff.ecs.soton.ac.uk/~lac/imp.pdf

Carr, L., De Roure, D., Hall, W. and Hill, G. (1995) The Distributed Link Service: a Tool for Publishers, Authors and Readers. *World Wide Web Journal* (special issue, Proceedings of the Fourth International World Wide Web Conference) No. 1, Winter 1995/96 http://www.w3.org/pub/Conferences/WWW4/Papers/178/

Davidson, L. A. and Douglas, K. (1998) Digital Object Identifiers and Their Role in the Implementation of Electronic Publishing. *Socioeconomic Dimensions of Electronic Publishing Workshop,* held in cooperation with the *1998 IEEE International Conference*

on Advances in Digital Libraries, April 1998
http://www.lita.org/igs/doiieeefp1.pdf
or see the updated version version in html
Digital Object Identifiers: Promise and Problems for Scholarly Publishing. *Journal of Electronic Publishing*, Vol. 4, issue 2, December   http://www.press.umich.edu/jep/04-02/davidson.html

Dixon, A. (1998) The Wannabee Culture: Why No-One Does What They Used to Do. *Issues in Science and Technology Librarianship*, Winter
http://www.library.ucsb.edu/istl/98-winter/article4.html

Garfield, E. (1955) Citation indexes for science: a new dimension in documentation through association of ideas. *Science*, Vol. 122, 15 July, 108-111

Giles, C. L., Bollacker, K. D. and Lawrence, S. (1998) CiteSeer: An Automatic Citation Indexing System. *Proceedings of the third ACM International Conference on Digital Libraries*, Pittsburgh, USA, June (ACM: New York)

Han, Y., Loke, S. W. and Sterling, L. (1997) Agents for Citation Finding on the World Wide Web. In *PAAM 97: Proceedings of the Second International Conference on the Practical Applications of Intelligent Agents and Multi-Agent Technology* (Practical Application Company: Blackpool, UK), pp. 303-317

Hellman, E. (1998) Scholarly Link Specification Framework (SLinkS), public draft #1.5, November 24 http://www.openly.com/SLinkS/

Hitchcock, S. (1996) Open Journals. *Ariadne*, issue 5, September
http://www.ariadne.ac.uk/issue5/open/

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., Carr, L. and Hall, W. (1996) A Survey of STM Online Journals 1990-95: the Calm Before the Storm. In *Directory of Electronic Journals, Newsletters and Academic Discussion Lists*, sixth edition, edited by D. Mogge, (Washington, D.C.: Association of Research Libraries), pp. 7-32, http://journals.ecs.soton.ac.uk/survey/survey.html

Hitchcock, S., Carr, L., Harris, S., Hey, J. M. N. and Hall, W. (1997b) Citation Linking: Improving Access to Online Journals. In *Proceedings of the Second ACM International Conference on Digital Libraries*, Philadelphia, USA, July (ACM: New York), pp. 115-122 http://journals.ecs.soton.ac.uk/acmdl97.htm

Hitchcock, S., Kimberley, R., Carr, L., Harris, S. and, Hall, W. (1998a) Webs of Research: Putting the User in Control. In *Proceedings of IRISS'98: Internet Research and Information for Social Scientists*, Bristol, UK, March
http://sosig.ac.uk/iriss/papers/paper42.htm

Hitchcock, S., Quek, F., Carr, L., Hall, W., Witbrock, A. and Tarr, I. (1998b) Towards Universal Linking for Electronic Journals. *Serials Review*, Vol. 24, No. 1, Spring, 21-33
http://journals.ecs.soton.ac.uk/IFIP-SerRev98.html

Hunter, K. (1998) Adding Value by Adding Links. *Journal of Electronic Publishing*, Vol. 3, Issue 3, March http://www.press.umich.edu/jep/03-03/hunter.html

Open Journal Project (1995) An Open Journal Framework: Integrating Electronic Journals with Networked Information Resources. JISC/eLib sheet flyer
http://journals.ecs.soton.ac.uk/flyer.html

Open Journal Project (1998) Open Journal Project: Final Report to eLib, August
http://journals.ecs.soton.ac.uk/yr3/3rd-year-open.htm

Khare, R. and Rifkin, A. (1997) Capturing the State of Distributed Systems with XML. *World Wide Web Journal*, Vol. 2 , No. 4, Fall, 207-218
http://www.cs.caltech.edu/~adam/papers/xml/xml-for-archiving.html

Probets, S., Brailsford, D. F., Carr, L. and Hall, W. (1998) Dynamic Link Inclusion in Online PDF Journals. In *Proceedings of EP'98, the seventh International Conference on Electronic Publishing, Document Manipulation and Typography*, St Malo, France, April
http://www.ep.cs.nott.ac.uk/~sgp/ep98.pdf

Renear, A. (1997) The Digital Library Research Agenda: What's Missing -- and How Humanities Textbase Projects Can Help. *D-Lib Magazine*, July/August
http://www.dlib.org/dlib/july97/07renear.html

Rosenblatt, B. (1997) The Digital Object Identifier: Solving The Dilemma Of Copyright Protection Online. *The Journal of Electronic Publishing*, Vol. 3, Issue 2, December
http://www.press.umich.edu/jep/03-02/doi.html

Rubinstein, E. (1997) Notice the Library Sprouting on Your Desktop? *HMS Beagle*, issue 15, September
http://www.biomednet.com/hmsbeagle/15/webres/insitu.htm (registration required)

Rusbridge, C. (1998) Towards the Hybrid Library. *D-Lib Magazine*, July/August
http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html

Schmaranz, K. (1996) Professional Electronic Publishing in Hyper-G: The Next
Generation Publishing Solution on the Web. *WebNet 96*, San Francisco, CA
http://aace.virginia.edu/aace/conf/webnet/html/130.htm

Spink, A., Wilson, T., Ellis, D. and Ford, N. (1998) Modeling Users' Successive Searches
in Digital Environments. *D-Lib Magazine*, April 1998
http://www.dlib.org/dlib/april98/04spink.html

Tenopir, C. and Ennis, L. (1998) The Digital Reference World of Academic Libraries.
Online, Vol. 22, No. 4, July
http://www.onlineinc.com/onlinemag/OL1998/tenopir7.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

W3C, the World Wide Web Consortium (1998) Extensible Markup Language (XML) 1.0.
REC-xml-19980210, W3C Recommendation 10-February-1998
http://www.w3.org/TR/1998/REC-xml-19980210

Appendix: Publisher partners in the Open Journal project

- Academic Press
- BIDS (Bath Information & Data Services)
- BioMedNet Ltd
- British Computer Society
- Cambridge University Press
- Chapman & Hall (CompSciNet)
- Company of Biologists
- Stevan Harnad, Cognitive Science Centre, Southampton University
- Institute for Scientific Information
- MCB University Press
- Oxford University Press
- John Wiley & Sons Ltd

D-Lib Magazine Access Terms and Conditions

hdl:cnri.dlib/december98-hitchcock