Developing services for open eprint archives: globalisation, integration and the impact of links

Steve Hitchcock, Les Carr, Zhuoan Jiao, Donna Bergmark, Wendy Hall, Carl Lagoze* and Stevan Harnad*

Multimedia Research Group, Department of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK

* Digital Library Research Group, Department of Computer Science, Cornell University, Ithaca, NY 14853-7501, USA

Contact for correspondence: Steve Hitchcock sh94r@ecs.soton.ac.uk

ABSTRACT

The rapid growth of scholarly information resources available in electronic form and their organisation by digital libraries is proving fertile ground for the development of sophisticated new services, of which citation linking will be one indispensable example. Many new projects, partnerships and commercial agreements have been announced to build citation linking applications. This paper describes the Open Citation (OpCit) project, which will focus on linking papers held in freely accessible eprint archives such as the Los Alamos physics archives and the other distributed archives, and which will build on the work of the Open Archives initiative to make the data held in such archives available to compliant services. The paper emphasises the work of the project in the context of emerging digital library information environments, explores how a range of new linking tools might be combined and identifies ways in which different linking applications might converge. Some early results of linked pages from the OpCit project are reported.

KEYWORDS: electronic publishing, digital library information architectures, reference linking, distributed collections, eprint archives, Open Archives

INTRODUCTION

The process of scholarly communication, in particular the aggregation of formal academic papers in journals, is probably about to enter its fastest period of change since 1994-6. The immediate platform for this change was the emergence of the World Wide Web as a popular medium in 1994 and the subsequent conversion of most established journals to electronic facsimiles delivered over the Web, a process which began to build momentum in 1996 (Hitchcock et al. 1996). While growth in the number of e-journals continues to accelerate towards an estimated 10 000 (Maclennan 1999, Hunter 1999), it is prior developments, such as the establishment of free electronic archives, or eprint archives, that are beginning to influence wider changes.

Launched in 1991, the importance of the Los Alamos physics eprint archive, the first and preeminent archive of its kind (Ginsparg 1994), cannot be underestimated, but its practical ramifications have so far been largely confined to its home community in physics. Many have questioned, because of cultural differences between different academic disciplines (Kling and McKim 1999), whether the eprint model will be accepted beyond physics. That contention will be challenged by the most significant new eprint archives to have emerged since 1991: PubMed Central, launched at the beginning of this year and sponsored by the National Institutes of Health, covering all fields in biomedical and life sciences (Varmus 1999); and the Computing Research Repository (CoRR), sponsored by the ACM and the British Computer Society (Halpern and Lagoze 1998). An immediate effect of PubMedCentral has been the announcement by some biomedical publishers, notably the British Medical Journal (Delamothe and Smith 1999) and the Current Science Group (Anon. 1999), of freely accessible archives of electronic copies of current and past papers, both pre- and post-publication. As well as discipline-based archives, institutionally-based initiatives such as Scholars Forum are planned. (Buck et al. 1999)

The essential feature of the Los Alamos eprint archive model is author self-archiving based on a free-to-archive, free-to-access service. Compared with journals, eprint archives provide an almost wholly automated and highly efficient organisational framework and distribution mechanism based on the Internet, but without many of the additional services that journals provide, such as peer review, and other services that mostly require human intervention. As more archives attract more papers the test for journals is how they respond to the effective loss of exclusivity that most depend on, and how they cope with the new economics of journal publishing in which every facet of traditional value-adding is re-evaluated against the for-free services.

There is another dimension too: globalisation. With no geographical or financial barriers the next inevitable step is, if not universalisation where archives all adopt the same technical infrastructure, then integration of the archives based on new services. In this agenda archives are not only freely accessible to users but are open to independent, third-party computational processes - for an example, see

the UPS prototype (Van de Sompel et al. 2000) - on which these services will be built. This approach has been formalised by the Open Archives initiative, a group that includes archive managers, potential service developers and academic librarians, in its Santa Fe Convention (Van de Sompel and Lagoze 2000).

The publishing industry anticipates that links on citations within scholarly papers will be one of the primary new services driving integration between scholarly sources (Needleman et al. 1999). Linking services cannot be implemented piecemeal on such a scale. A number of journal publishers will collaborate through CrossRef to use a Digital Object Identifier (Davidson and Douglas 1998) based system to form reference links between their, separately maintained, journal contents (Atkins et al. 2000). Most traditional journals service providers - aggregators, subscription agents and secondary publishers (e.g. Brunelle and Johnson 1999) - as well as new producers, highlight the important role of citation links in their services.

This paper considers the implications of the new wave of eprint archives and the development of open archives. It focusses on an Open Archives service being developed by the Open Citation (OpCit) project, funded by the joint NSF-JISC international digital libraries programme. The project will build citation links uniting large, high-profile and distributed archives but the service is planned to be extensible to other services that provide access to refereed scholarly papers.

OPEN ARCHIVES: SEPARATING CONTENT FROM SERVICES

Eprint archives are noticeably entering a new phase. Not only have significant new archives launched, but new services are being developed to complement the traditional content management functions of the archives. The key feature is that many of these new services will be independent of the underlying archive contents, and there are good reasons for this. Growth in the number of archives and greater use create opportunities for third-party developers, and such developments thus do not add to the low-cost base of most archives and can be developed either commercially or non-commercially.

More importantly, as the use of archives crosses boundaries between disciplines, the role of new services will be to enable the user to view an integrated set, or selected subset, of all archives. User customisation features to set the scope of a service are likely to become increasingly common. For archives predicated on free and open access, capitalising on this feature means not constraining users within the boundaries of a single discipline or archive. Cross-disciplinary navigation support such as indexing, searching and linking should provide a consistent interface and seamless service regardless of which archives are accessed by the user, who need have no knowledge of the structure of an archive or, apart from perhaps noting the archive source and publisher identities, from where a viewed document originated.

Integration of free eprint archives through independent services is the preferred view of the Open Archives initiative of the way in which such archives may pave the way for a unified global scholarly literature, encompassing not just the archives but journals and other contributing literatures.

This level of service will be achieved by interoperability agreements between the archives, for example the protocols through which services can communicate with the archives, and metadata forms that will expose the data structures and the contents of archives. Methods chosen must encourage widespread adoption, but also ensure conformity of safe and trusted practices that do not compromise the integrity of the archives.

A well established protocol for data communication in digital library applications is Dienst, developed at Cornell University (Dienst 1999). An implementation of part of the Dienst protocol has been adopted by the Open Archives initiative for data harvesting. The OpCit project, in which the Dienst research group is one of the principal partners, is similarly building on this core technology for its linking applications.

LINKING THE OPEN ARCHIVES: THE OPCIT PROJECT

Linking is an apparently simple concept, especially the model implemented by the Web. A unidirectional point-and-click event presents the user with the page from the location pointed at by a locator, the URL, that is authored into the linking page. This simplified form of hypertext linking, it has long been argued, is inadequate to support robust services required for large-scale information environments, such as the contents of scholarly communications organised via digital libraries.

In terms of reference linking services, an early pioneer and predecessor of the OpCit project was the Open Journal project (Hitchcock et al. 1998a). The information environment envisaged in that project centred on published journal contents but was essentially unbounded, allowing links to take users to other types of materials, such as abstracting and indexing services, dictionaries and biological databases. In practice different implementations of Open Journals had to be bounded, principally to manage the user interface more effectively.

Originally supported by four publishers, by its end 12 publishers were involved. It is safe to assume that many more scholarly publishers now recognise the importance and power of links in electronic documents, especially links which implement the long-established, non-electronic form of linking inherent to the scholarly or scientific paper, the reference.

With this wider participation has come the recognition, as Caplan and Arms (1999) show, that reference linking may not be as simple as originally envisaged. There are a number of reasons for this. Expectations are high, of links from almost every reference in every electronically-

accessible paper, both backwards in time as provided by a typical reference list in a paper, and forwards in time in a manner made familiar by the Science Citation Indexes (Hitchcock et al. 1998b). Constraining these expectations are the need fo accuracy and reliability of links, and availability of the necessary contents in electronic form (Hitchcock et al. 1998c). There may be multiple, but not identical, versions of the same document at multiple locations (Caplan and Flecker 1999). Finally, there are the financial and authorisation barriers imposed by commercial journals and services. Access requirements can differ for each user, for each access location, for each document and for each document location. Competing publishers may have become allies in cross-publisher reference linking, but although there are various possible solutions to the problem of linking across distributed collections, there is as yet no convincing demonstration or detail of how this might be achieved in an exclusively commercial environment (Atkins 1999).

In this context, linking is more than just a technical process but must be viewed as part of the social and business phenomena that are shaping the new information environments. This is recognised in the scope and partnerships that form the OpCit project. Ironically, in the first environments the project will explore, selected open archives, such is the accessibility of so much content that the project could almost reduce the problem to its technical issues, but the longer-term commitment is to collaborating with others in the wider environments discussed below.

Three principal objectives elaborated by the OpCit project (Harnad et al. 1999) concern scale-compatibility-universality:

- Scale: to hyperlink each of the over 100 000 papers in Los Alamos's unique online physics archive to every other paper in the archive that it cites.
- Compatibility: to develop and integrate a family of generic linking tools and to design author and user interfaces to enable easy adoption by other archives.
- Universality: to promote the power of this remarkable new way of navigating the scientific journal literature and induce authors in other fields to create interlinked online archives like Los Alamos across disciplines and around the world.

Primary partners in the project are:

- Southampton University's Multimedia Research Group, with its expertise in linking applications, also home of the CogPrints eprint archive for cognitive sciences and a mirror site for the Los Alamos archives
- Cornell University, which is a major player building NCSTRL based on Dienst
- Paul Ginsparg from Los Alamos

In related work the Cornell group will apply the linking technology used in OpCit to the ACM Digital Library, and will experiment with the distributed object technology being developed in the FEDORA project (Payette and Lagoze 1998).

When the Open Journal project began in 1995 the linking tools used then were all developed at Southampton University. Now there are a range of tools created to suit different linking requirements:

- The **Distributed Link Service** (Southampton University): much modified since its use in the Open Journal project, now more modular and works with contributing software 'agent' applications, but still essentially does the job of making links visible and actionable for users in documents of various formats at any Web location. (Carr *et al.* 1998)

- **Research Index** (NEC Research, Princeton): formerly called Citeseer, an impressive collection of tools for resource discovery, data collection and indexing of documents found on the Web. (Lawrence et al. 1999) Citeseer has been referred to as building an ISI-like service on the Web, so there is another interesting parallel with the Open Journal project which produced its most successful citation linking demonstrator using data provided by ISI.

- **SFX** (University of Ghent): similar linking capabilities as DLS, but tailored for knowledge management in a host library environment. (Van de Sompel and Hochstenbach 1999a)

- **SLinkS** (Openly Informatics): a software-based intermediary service that enables the URL to be calculated of a cited document published in a journal and held within the publisher's authorised service. Unlike the DOI service there are no identifiers, but cooperating publishers and content providers specify rules that enable URLs to be calculated from standard bibliographic information. New services can be built on this open specification (Hellman 1999). More generally, objects that can be accessed by known or calculable identifiers can be mediated by the company's Linkbaton service to provide more user options.

An early objective of the OpCit project is to examine the structure of these tools at a software level and elaborate how they might work together, in technical terms, exposing the respective application programming interfaces (APIs). By publishing these details it is hoped to establish a generic foundation for emerging digital library linking applications, recognising that these research tools will constantly change and evolve and that new tools will be developed. This is a novel approach in that linking applications of all types have typically been tool based, but most such tools, although often highly functional, have tended to be used independently.

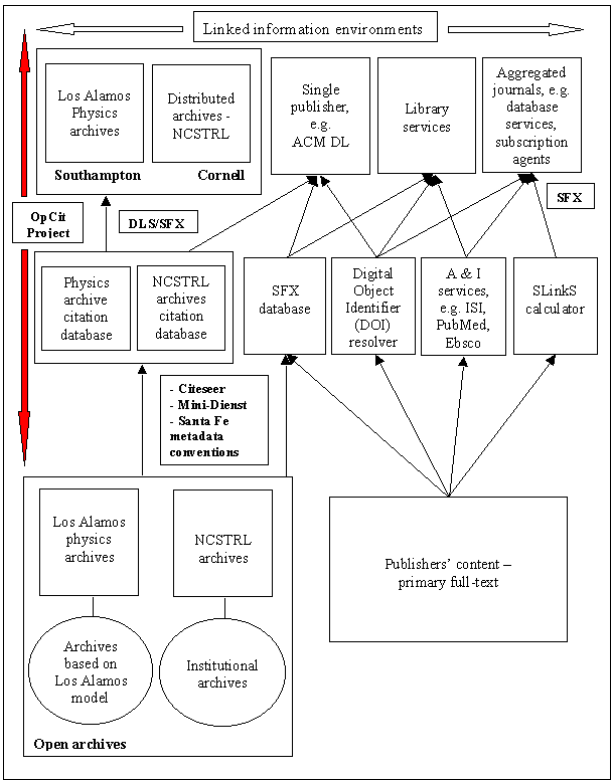

Figure 1. Information environments and link services: a range of possible scenarios / tools for OpCit and beyond

LINKING SERVICES AND INFORMATION ENVIRONMENTS

Information environments organised via digital libraries continue to be part of the Web but are distinguished by services that apply to contents that are deemed to be within them, determined not by physical location but by the nature and selection of those contents and the services that act on them. Examples of these environments might include the contents of libraries as in the UK eLib-funded Distributed National Electronic Reserve (DNER), single-publisher collections such as the ACM Digital Library, larger collections of published journal papers accessed via DOI-based services, or distributed archives such as the Networked Computer Science Technical Report Library, NCSTRL. In essence, in these environments the Web is transformed from a document delivery service into a dynamic, computational framework. The information environments being explored for citation linking by the OpCit project are outlined in Figure 1.

The scope of the project and that shown in Figure 1 is intentionally wide, although of immediate concern is the area bounded by the left-hand vertical arrows and the information environments denoted by 'Southampton' and 'Cornell'. The first results reported below specifically relate to Southampton's work with the Los Alamos physics archives.

These scenarios are highly flexible and may be viewed in other ways by different applications. For example, the connecting arrows are all possible scenarios, although such applications may not have been implemented yet. An SFX application linking various resources - notably a number of abstracting and indexing (A&I) services, some publishers' full-text content as well as the Los Alamos archives - via an SFX database for library environments was described by Van de Sompel and Hochstenbach (1999b). The multi-publisher CrossRef linking initiative is likely to route through a DOI resolver to a library or aggregated journal environment. ISI has announced a number of agreements with publishers to link between Web of Science and full-texts. Each of these applications can be identified in some form in Figure 1.

In addition, the roles imputed to the linking tools in Figure 1 may not reflect their wider capabilities. Citeseer and SFX variously contain information retrieval, database and linking functions. This is not shown for Citeseer. Demonstrator services based on these tools have created user interfaces. In this respect the tools could legitimately be indicated as part of the information environments at the top of the diagram, but are not in this view. Apart from Citeseer, methods for data extraction from the archives - Dienst and the Santa Fe metadata conventions - that will be used to create citation databases are ongoing developments of the Open Archives initiative.

Thus it can be seen how interchangeable these components are, and it is anticipated that this flexibility will drive significant innovation in citation linking. Figure 1 should be considered a perspective on environments for citation linking held by the OpCit project but not necessarily by others.

REFERENCES

[1] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).

[2] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2752 (1977).

[3] P. Kraus and F. Wilczek, Nucl. Phys. **B433**, 403 (1995).

[4] P. Kraus and E. Keski-Vakkuri, Nucl. Phys. **B491**, 249 (1997).

[5] S. Massar and R. Parentani, gr-qc/9903027.

[6] P. Painlevé, C. R. Acad. Sci. (Paris), **173**, 677 (1921).

[7] P. Kraus and F. Wilczek, Mod. Phys. Lett. **A 9**, 3713 (1994).

[8] W. G. Unruh, Phys. Rev. D **14**, 870 (1976).

[9] P. Kraus and F. Wilczek, Nucl. Phys. **B437**, 231 (1995).

Reference section of article hep-th/9907001 with added links. The link boxes are coloured to indicate varying success in identifying real links (see Figure 5)

OPCIT: EARLY IMPLEMENTATIONS AND RESULTS

The process of adding citation links dynamically to documents retrieved from an archive involves parsing the document during download to identify and read citations. The data are compared with a precompiled link or citation database, and a link to the cited work added where an exact match is found. For more details, one method for doing this was described by Hitchcock *et al.* (1997).

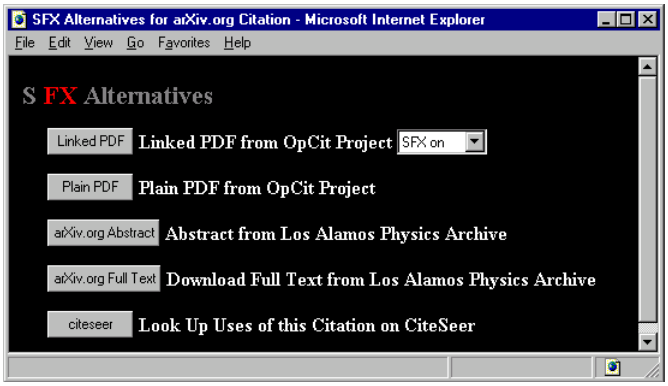

Figure 3. Activating the link for reference [5] in Figure 2 optionally returns an 'SFX' router page offering the user a choice of sources. If the SFX option is switched off the user is taken directly to the chosen linked pdf version.

A similar method has been adopted for OpCit, but in this case the application demands that a larger, richer citation database is compiled. Broadly, the stages involved in the compilation of this database are:

linking, effectively a highly automated version of Garfield's famous citation analyses:

- which are the most frequently cited papers?
- what papers refer to the current paper (forward linking)?
- who co-reference the same papers, and to what extent? (identifying similar research interests)
- in what context is a paper is cited? (i.e. citation context)
- which journals are the most popular?

In a preliminary implementation of the linking model, where one of the objectives was to integrate services provided by some of the linking tools described above, a successfully linked citation directs the user to an intermediate page offering the user a choice: either download the text from the archive or look up some contextual information on the citation. In this example the links in the original document are added by the DLS, and the intermediate page is produced from an SFX-like database which will maintain some knowledge of the user privileges and can offer all versions of the cited paper that are accessible to the user. In this case only the archive

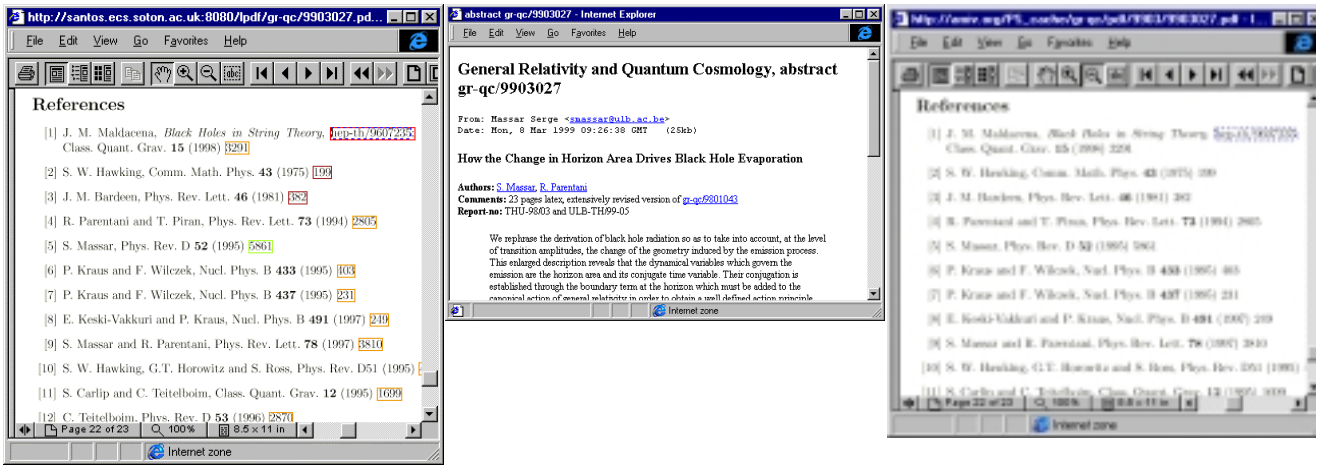

Fig 4a **Fig 4b** **Fig 4c**

Figure 4. **Selecting options from Figure 3: a, PDF version enhanced with reference links; b, abstract of cited paper; c, the original full-text from the archive (the dashed link is created by the archive on explicitly cited identifiers)**

1. transforming original documents to a format (e.g. plain text) for extracting citations;
2. parsing documents to identify and read citations;
3. design database schema for storing citation information in an information-rich, easy-to-use and flexible manner, and accommodating future extensions.

Citeseer excels in this respect, but the algorithms have reportedly not been so successful when applied to the Los Alamos physics archive because the references contain too little information. The project aims to build tools to supplement the results produced by Citeseer for other archives.

The importance of stage 3 is that richer citation databases can provide users with useful information, apart from versions - abstract, and link-enhanced and authors' original full texts are available. In principle, if the user or a library subscribes to the journal in which the cited paper was published that version could be linked from SFX too; also, other versions of the paper, in abstracting services for example. The different stages of retrieval are shown in Figures 2-4. Contextual information would be retrieved from a database compiled by Citeseer (those results are not shown here, but an example of this service can be tried online at `http://citeseer.nj.nec.com/cs`).

CITATION LINK ANALYSIS

Working initially just within the Los Alamos physics eprint archives, citations were analysed from a subset of papers submitted during 1999 from one section of the archive, hep-

th (theoretical high-energy physics), a total of over 2200 papers. These papers contained over 65 000 citations.
Citations in physics are notoriously terse, citing author names, then either an archive identifier or a standard abbreviation for the journal title followed by some undifferentiated numbers, usually volume number, start page number and the year of publication (see Figure 2). Sometimes both the archive identifier and journal data are included.

For our subset of documents the relative success in automatically recognising and resolving to the corresponding document in the citation database compiled from the archive can be gauged from Figure 5. The colours in this chart correspond to the link colours in Figure 2. Some links were simply derived from explicitly cited identities for the archive documents, others were derived purely from the bibliographic data in the citation. Where archive identities are not included directly, they can alternatively be derived from data in the archive journal-ref metadata or from other more intensively-maintained, overlapping bibliographic databases in physics such as SPIRES. (SPIRES is maintained by the Stanford Linear Accelerator Center (SLAC), another associate partner in the OpCit project.)

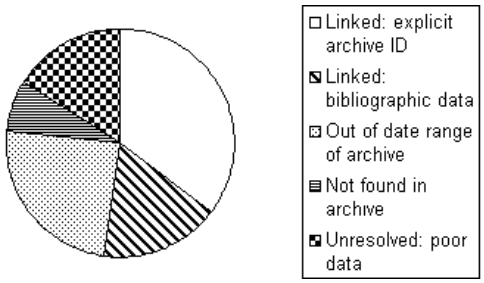

Figure 5. Resolving and linking citations in a subset of hep-th papers: what proportion could be linked, what could not and why

Where resolution of reference data against the database, and therefore linking, was unsuccessful there could a number of reasons, and the relative occurrence of these problems is also shown in Figure 5. In some cases a citation was correctly recognised, but the cited paper is too old to feature in the archive, or a reference was recognised but not found in the database (the cited paper is not in the archive). The remaining citations could not be resolved, possibly due to poor formatting, incorrect data, etc. It can be seen that just over half of the references were successfully linked within the archive. The number of successful links could be increased significantly if the archives were supplemented with other, older sources, online archival journals say. About 16 per cent of citations from this subset may never be resolvable.

These percentages might be generalisable across the physics archive but not necessarily to other archives or applications, although it is interesting to compare broad measures of success in citation linking such as this example (52 per cent of citations linked within the archive) with that reported by Electronic Press for Medline linking in which "on average only about 60% of references in a typical medical paper are contained in the Medline data". Of these Medline citations only 85% were reliably resolved by the linking software (Hitchcock *et al.* 1998c). Factors that control these figures include the size and accessibility of the archive and other document sources, and the accuracy, quality and completeness of the reference data.

OPCIT DESIGN AND EVALUATION

There is another way tackle potentially unresolvable citations for new and future submissions to the archive: at source, when the papers are deposited by authors. As well as developing linking services, additional tasks for the project are citation analysis, interface design, and user testing and evaluation.

Designing user interfaces for the linked archives concerns not just navigation but author deposit too. It is argued that a barrier to wider participation in eprint archives beyond physics is the need for more user-friendly procedures for authors. Given that original submissions to the archives, i.e. the preprints rather than the later refereed reprints, are unedited and unmoderated, the responsibility for the quality and accuracy of such submissions lies wholly with its authors. So the challenge is to make the deposit process both easier and more intuitive but with more immediate feedback to encourage high standards of correctness and completeness.

Reference linking obviously depends strongly on the accuracy of data provided by authors and so improving correctness in this area is of particular interest to the project. One idea is is to provide dynamic checking for references in newly submitted papers using the same process as used to produce citation links and inviting the author to respond. In this case correct, linkable citations would appear as standard links, but other citations might be highlighted by different link colours, indicating possible problems with a reference, suggesting a reason for the problem and highlighting which ones the author could usefully amend. Such a scheme might look like that shown in Figure 2 with link types as defined in Figure 5 (although the model implementation was not designed for this purpose).

While each of the components developed by the project will be subjected to user evaluation by standard means, in citation analysis there is an inherent means of evaluating common practice and usage of the archive by its most important constituency: its authors.

Since the project is experimenting with stored data from the archive rather than real live data there are other effects that can be used to monitor user behaviour. The stored data is a snapshot of the archive contents at that time and using simple difference computation can be compared with later datasets to reveal the extent of changes, minor or major. Of

particular interest is the proportion of pre-publication papers in the archive that are replaced by the final published version. Only the authors of a paper can change or update it in the archive, so there is no standard procedure for replacing pre- with post-publication copies, but analysis is revealing some common patterns.

CONCLUSION

The Internet is regarded by some as introducing a paradigm shift in communication (Valauskas 1997). With major new free-to-use eprint archives joining the well established Los Alamos archives, and with progress towards integration and globalisation of these archives motivated by the Open Archives initiative, the new paradigm beckons for scholarly communication. For scholarly communication these developments bring closer the critical and long heralded (Dyson 1994) transition promised by the Internet in which primary value accrues to services rather than just to content. For any such transition the main issue is access. The archives provide access to content and projects such as OpCit and others are demonstrating how that can be exploited and value added through new services. (Harnad 1998)

For commercial scholarly publishing, which by its nature imposes financial barriers to access, the picture is less clear. It is ironic that as some major publishers agree to explore possible mechanisms for managing links between their journals, they are failing to respond to the issue of improving access. For example, users can access primary journal papers via abstracting, indexing and aggregation services now transformed into electronic forms and sporting new brand names but mostly supported by familiar corporate interests. A possible scenario was sketched by Morrow (1999):

> "Sites who have signed up for the ScienceDirect trial through BIDS now have access through four different routes ...
> * Selecting the Elsevier option for the BIDS route to ScienceDirect does NOT preclude access to ScienceDirect via Web of Science through MIMAS. Even after July 2000, Web of Science users (with the appropriate linking licence) at sites who have selected the Elsevier licensing option will continue to be able to access ScienceDirect material (in exactly the same way as those who sign a NESLI/Swets agreement)."

Individually these are perfectly good services that are legitimately exploring new solutions at a critical moment of change. The question it raises, however, is does this help users? This is not a paradigm shift, but a muddle that is the result of self-preservation driven by service providers rather than users. It arises because it fails to recognise the shift in value from content to services. Some publishers that maintain close links with the research community, typically professional societies, have been required to pay close attention to the development of the archives. They understand the distinction and have acted (e.g. the BMJ and the Current Science Group), or are prepared to act (Doyle 1999), to free their electronic archives and to free authors to self-archive.

This paper has highlighted the real paradigm shift taking place in scholarly communication, towards more open and accessible information. This benefits linking services, but many other digital library services can benefit too. A new phase of convergence in digital library services is beginning and is being driven by those embracing pragmatic, widely shared interests throughout the scholarly community.

REFERENCES

1. Anon. (1999) "Science Publishing - Beginning of a Revolution". Current Science Group, press release, 26th April http://www.genomebiology.com/pressrelease26apr99.asp
2. Atkins, Helen (1999) "The ISI Web of Science - Links and Electronic Journals". *D-Lib Magazine*, Vol. 5, No. 9, Septembe r http://www.dlib.org/dlib/september99/atkins/09atkins.html
3. Atkins, Helen, *et al*. (2000) "Reference Linking with DOIs: A Case Study". *D-Lib Magazine*, Vol. 6 No. 2 February http://www.dlib.org/dlib/february00/02risher.html
4. Brunelle, Bette and Johnson, Dana (1999) "Connecting the Docs: New Models and New Tools to Link Bibliographic Databases and Full Text Journals". *CNI Fall 1999 Task Force Meeting: Project Briefings* http://www.cni.org/tfms/1999b.fall/PBrief99Ftf.html
5. Buck, Anne M., Flagan, Richard C. and Coles, Betsy (1999) "Scholar's Forum: A New Model For Scholarly Communication". March http://library.caltech.edu/publications/ScholarsForum
6. Caplan, Priscilla and Arms, William Y. (1999) "Reference Linking for Journal Articles". *D-Lib Magazine*, Vol. 5, No. 7/8, July/August http://www.dlib.org/dlib/july99/caplan/07caplan.html
7. Caplan, Priscilla and Flecker, Dale (1999) "Choosing the Appropriate Copy". Draft version circulated to Ref-Links listserv <ref-links@doi.org>, 14th October http://www.doi.org/mail-archive/ref-link/msg00060.html
8. Carr, L. A., Hall, W. and Hitchcock, S. (1998) "Link Services or Link Agents?" *Ninth ACM Conference on Hypertext*, Pittsburgh, June http://www.staff.ecs.soton.ac.uk/~lac/LinksOrAgents.pdf
9. Davidson, Lloyd and Douglas, Kimberly (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
10. Delamothe, Tony and Smith, Richard (1999) "Moving Beyond Journals: the Future Arrives with a Crash: New Ways to Disseminate Research from NIH and the BMJ". *British Medical Journal*, Vol. 318, 19th June, 1637-1639 http://www.bmj.com/cgi/content/full/318/7199/1637
11. Dienst (1999) "Dienst Overview and Introduction". December http://www.cs.cornell.edu/cdlrg/dienst/DienstOverview.htm
12. Doyle, Mark (1999) "Re: PubMedCentral". Message

posted on September 1998 American Scientist Forum listserv <SEPTEMBER98-FORUM@LISTSERVER.SIGMAXI.ORG>, 22nd September http://listserver.sigmaxi.org/scripts/wa.exe?A2=ind99&L=september98-forum&D=1&O=D&F=l&P=26434

13. Dyson, Esther (1994) "Intellectual Property on the Net". *Release 1.0*, December http://www.edventure.com/release1/1294.html
14. Ginsparg, Paul (1994) "First Steps Towards Electronic Research Communication". *Computers in Physics*, Vol. 8, No. 4, July/August http://xxx.lanl.gov/blurb/blurb.ps.Z
15. Halpern, Joseph Y. and Lagoze, Carl (1998) "The Computing Research Repository: Promoting the Rapid Dissemination and Archiving of Computer Science Research". *CoRR eprint*, cs.DL/9812020, December http://xxx.lan.gov/abs/cs/9812020
16. Harnad, S. (1998) "On-Line Journals and Financial Fire-Walls". *Nature*, Vol. 395, 10th September, 127-128 http://www.cogsci.soton.ac.uk/~harnad/nature.html
17. Harnad, Stevan *et al*. (1999) "Integrating and Navigating Eprint Archives Through Citation-Linking". Proposal to NSF-JISC International Digital Libraries Research Programme http://www.cogsci.soton.ac.uk/~harnad/citation.html
18. Hellman, Eric (1999) "The S-Link-S™ Framework for Reference Linking: Architecture and Implementation". *Proceedings of the ICCC/IFIP Conference on Electronic Publishing '99. Redefining the Information Chain - New Ways and Voices*, edited by John W.T. Smith, Anders Ardö and Peter Linde, Ronneby, Sweden, May (Washington D.C.: ICCC Press) http://www5.hk-r.se/elpub99/ap.nsf/08c6c2f88424ad99c12566ff002a0c10/ee358528a7c0afccc12566ff00380898/$FILE/68-73.pdf
19. Hitchcock, Steve, Carr, Leslie and Hall, Wendy (1996) "A Survey of STM Online Journals 1990-95: the Calm Before the Storm". *Directory of Electronic Journals Newsletters and Academic Discussion Lists*, edited by D. Mogge, sixth edition (Washington, D.C.: Association of Research Libraries), pp. 7-32 http://journals.ecs.soton.ac.uk/survey/survey.html
20. Hitchcock, Steve, Carr, Leslie, Harris, Steve, Hall, Wendy, Probets, Steve, Evans, David and Brailsford, David (1998a) "Linking Electronic Journals: Lessons from the Open Journal Project". *D-Lib Magazine*, December http://www.dlib.org/dlib/december98/12hitchcock.html
21. Hitchcock, S., Carr, L., Harris, S., Hey, J. M. N. and Hall, W. (1997) "Citation Linking: Improving Access to Online Journals". *Proceedings of the 2nd ACM International Conference on Digital Libraries*, edited by Robert B. Allen and Edie Rasmussen, 1997 (New York, USA: Association for Computing Machinery), pp. 115-122 http://journals.ecs.soton.ac.uk/acmdl97.htm
22. Hitchcock, Steve, Kimberley, Robert, Harris, Steve, Carr, Leslie and Hall, Wendy (1998b) "Webs of Research: Putting the User in Control". *Internet Research and Information for Social Scientists (IRISS) conference*, Bristol, March http://sosig.ac.uk/iriss/papers/paper42.htm
23. Hitchcock, Steve, Quek, Freddie, Carr, Leslie, Hall, Wendy, Witbrock, Andrew, Tarr, Ian (1998c) "Towards Universal Linking for Electronic Journals". *Serials Review*, Vol. 24, No. 1, Spring, 21-33 http://journals.ecs.soton.ac.uk/IFIP-SerRev98.html
24. Hunter, Karen (1999) "Journals Online: PubMed Central and Beyond, page 4". *HMS Beagle: The BioMedNet Magazine*, No. 61, 3rd September http://www.biomednet.com/hmsbeagle/61/viewpts/page4 (free registration required)
25. Kling, Rob and McKim, Geoffrey (1999) "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* http://xxx.lanl.gov/ftp/cs/papers/9909/9909008.pdf
26. Lawrence, Steve, Giles, C. Lee and Bollacker, Kurt (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/
27. Maclennan, Birdie (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
28. Morrow, Terry (1999) "BIDS services now linked to Elsevier's ScienceDirect". Message posted on lis-scitech listserv <lis-scitech@mailbase.ac.uk>, 27th September http://www.mailbase.ac.uk/lists/lis-scitech/1999-09/0025.html
29. Needleman, Mark (1999) "Meeting Report of the NISO Linking Workshop". Washington, D.C., February http://www.niso.org/linkrpt.html
30. Payette, Sandra and Carl Lagoze (1998) "Flexible and Extensible Digital Object and Repository Architecture". *Second European Conference on Research and Advanced Technology for Digital Libraries*, Crete, Greece, September, in *Lecture Notes in Computer Science*, Vol. 1513 (Springer-Verlag) http://www.cs.cornell.edu/payette/papers/ECDL98/FEDORA.html
31. Valauskas, Edward 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
32. Van de Sompel, Herbert, *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
33. Van de Sompel, Herbert and Hochstenbach, Patrick (1999a) "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

34. Van de Sompel, Herbert and Hochstenbach, Patrick (1999b) "Reference Linking in a Hybrid Library Environment: Part 3: Generalizing the SFX solution in the 'SFX@Ghent & SFX@LANL' experiment". *D-Lib Magazine*, Vol. 5 No. 10, October http://www.dlib.org/dlib/october99/van_de_sompel/10van_de_sompel.html
35. Van de Sompel, Herbert and Lagoze, Carl (2000) "The Santa Fe Convention of the Open Archives Initiative". *D-Lib Magazine*, Vol. 6, No. 2, February http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html
36. Varmus, Harold (1999) "Journals Online: PubMed Central and Beyond, page 1". *HMS Beagle: The BioMedNet Magazine*, No. 61, 3rd September http://www.biomednet.com/hmsbeagle/61/viewpts/page1 (free registration required)