

The Open Citation Project

Second Year Report to JISC

Project Manager: Steve Hitchcock
Duration of Award: 10/99-09/02
Period of Report: 09/00-08/01

Lead Institution: Southampton University

Version history of this report

This version 2.0 submitted to JISC 5th October 2001

Version 1.0 submitted to JISC 31st August 2001

First year report <http://opcit.eprints.org/y1report/y1report-final.pdf>

Aims, Objectives and Methodology

Reference Linking for Open Archives

<http://opcit.eprints.org/>

The Open Citation project between Southampton University, Cornell University and arXiv.org is ahead of its development plan. In its second year the project made significant progress on its primary deliverables and has initiated some important new developments.

The project has achieved its principal objective of producing reference-linked version of the physics arXiv, and has added citation analysis and citation ranked search (cite-baseSearch <http://cite-base.ecs.soton.ac.uk/cgi-bin/search>). Progress has been made towards enabling the original source arXiv with data from the OpCit citation database to a degree unanticipated in the original plan. The project is working on an enhanced metadata format (the Academic Metadata Format, AMF) to transfer the data using the Open Archives protocol. If successful this would put the results of the project before a large user base on a long-lasting basis.

Some of the software tools used to build this demonstrator have been made openly available to project partners.

Other original objectives of generalising the user interface and author deposit facilities for eprint archives are progressing in conjunction with our colleagues at Cornell (see that team's annual report to the NSF, June 30, 2001 at <http://www.cs.cornell.edu/cdlrg/Reference%20Linking/AnnualReportYear2.ps>), and with the development of EPrints.org software.

During Y2 the project consolidated its work on bibliometric analysis of citation and usage of the physics arXiv with a survey of users of eprint archives (<http://www.eprints.org/results/>).

The project aims to capitalise on its extensive development and data collection during Y2 by reaching more users to inform further evaluation, and by reporting the results more widely to promote awareness and usage of Open Archive services across the whole scholarly community.

Highlights, Outcomes and Important Findings from project

The most important new feature of the linked arXiv demonstrator (v. 2.0, <http://arabica.ecs.soton.ac.uk/>) is the ability to discover what later papers in arXiv have cited a selected paper from arXiv. Citations that lead a user forward in time map the development of an active area of research to its present by means of the most significant, most highly cited, papers. ISI has demonstrated the value of such services for many years. **OpCit is the first to successfully apply this approach to a large-scale eprint archive.**

Development of EPrints archive-creating software v.1.1 was accelerated to comply with the requirements of Open Archives initiative Metadata Harvesting Protocol v. 1.0. EPrints v. 2.0, with a completely new database and support for multiple archives on one server and international language support, is at alpha-testing stage (<http://www.eprints.org/>).

In terms of activities, the highlights of the year were:

- Development of a richer, more stable arXiv citation database
- Prototype citation-ranked search engine for arXiv
- Enhanced citation-linked arXiv demonstrator, includes 'cited by'
- Progress towards integration of OpCit data in arXiv
- Demonstrator reference linking API in a presentation/rendering application
- Open sourcing of key software modules for reference linking
- New version releases of EPrints software and support tools
- Survey of users of eprint archives
- A one-day seminar for our technical collaborators
- Proposals for extended metadata transport between OAi service providers

Changes to original award

There have been no major changes to the overall plan. Attention to technical developments in Y2 has resulted a slight restructuring of previously reported annual work plans. In Y3 the project will emphasise analysis, evaluation and dissemination.

Project staff

All full-time researchers who joined at the outset remain with project. There have been additional contributions from temporary contributors, including undergraduates. All those involved in the people are listed at <http://opcit.eprints.org/opcitpeople.shtml>

In lieu of the 0.5 research assistant position in the original proposal, temporary appointments have given the project flexibility to tackle specific tasks and support costs of the EPrints developer, which will continue to be met by the project until JISC Opsis funding begins.

The principal changes at Southampton during Y2 were:

- At Southampton Chris Gutteridge replaced Rob Tansley as EPrints developer.
- Students: two final year undergraduate students built projects working with OpCit: Tim Brody (cite-baseSearch); Catherine Hunt (survey of users of eprint archives)

Main changes involving our partners:

- The ArXiv team, including Paul Ginsparg, have moved from Los Alamos to Cornell University.
- Herbert Van de Sompel joined the Cornell team and contributed to project development, now becomes the first e-director of the British Library, where we hope to continue some joint development.

Involvement with Programme

- From the JISC-NSF programme, useful contacts were established with the Harmony and Cross Domain Resource Discovery projects. All three projects were subsequently identified to join DNER Z projects cluster. OpCit will contribute to this cluster.
- Close collaboration with EPrints.org will be maintained. Development and open sourcing of the core EPrints code will continue within the JISC Opsis project. A bid to extend EPrints from the archiving of pre- and post-refereeing research to the archiving and sharing of research data has been submitted to the EU-DataGrid Project through the UK e-Science programme.

Publications and Publicity

The following lists papers and articles published by the project in the period reported. A full list of project publications, including conference presentations, is at

<http://opcit.eprints.org/opcitpapers.shtml>

[1] Tim Brody, Zhuoan Jiao, Steve Hitchcock, Les Carr and Stevan Harnad
Enhancing OAI Metadata for Eprint Services: two proposals. *Experimental OAI-based Digital Library Systems Workshop*, Darmstadt, September 2001, held in conjunction with the *5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL)*
<http://opcit.eprints.org/ecdl-oai/oai-ecdl01.html>

[2] Donna Bergmark and Carl Lagoze
An Architecture for Automatic Reference Linking. Cornell University Technical Report, TR2001-1842, 2001, presented at the *5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL)*, Darmstadt, September 2001
<http://www.cs.cornell.edu/cdlrg/Reference%20Linking/tr1842.ps>

[3] Stevan Harnad, Les Carr and Tim Brody
How and Why To Free All Refereed Research From Access- and Impact-Barriers Online, Now. *High Energy Physics Libraries Webzine*, Issue 4, June 2001
<http://library.cern.ch/HEPLW/4/papers/1/>

[4] Donna Bergmark
Automatic Extraction of Reference Linking Information from Online Documents. Cornell University Technical Report, TR 2000-1821, November 2000
<http://www.cs.cornell.edu/cdlrg/Reference%20Linking/extraction.pdf>

[5] Donna Bergmark, William Arms and Carl Lagoze
An Architecture for Reference Linking. Cornell University Technical Report, TR 2000-1820, October 2000
<http://www.cs.cornell.edu/bergmark/ReferenceLinkingArchitecture.ps>

[6] Stevan Harnad and Leslie Carr
Integrating, navigating and analyzing eprint archives through Open Citation Linking (the OpCit Project). *Current Science Online*, Vol. 79 No. 5, 10th September, 2000 (special issue in honour of Eugene Garfield)
<http://tejas.serc.iisc.ernet.in/~currsci/sep102000/629.pdf>

Newspaper and viewpoint articles

[1n] Richard Poynder

INSIDE TRACK: A mission to free scientific ideas: INTERVIEW STEVAN HARNAD: A university professor is championing do-it-yourself scholarly publishing using the internet. *Financial Times*, July 24, 2001

<http://globalarchive.ft.com/globalarchive/articles.html?id=010724001385&query=harnad>

[2n] Stevan Harnad

Why I think research access, impact and assessment are linked. *Times Higher Education Supplement*, Vol. 1487, 18 May 2001, p. 16

Extended version at <http://www.cogsci.soton.ac.uk/~harnad/Tp/thes1.html>

[3n] Stevan Harnad

The Self-Archiving Initiative. *Nature*, Vol. 410, 1024-1025, 2001; *Nature Web Debates* version, 26 April 2001

<http://www.nature.com/nature/debates/e-access/index.html>

Engagement with Potential Outcomes Users

Contact with potential users. Complementing the mining of arXiv user data by the project ('mining the social life of an eprint archive') reported last year, OpCit supported another undergraduate project to survey users of arXiv and of the CogPrints eprint archive. The earlier work was based on evidence of what users actually do; this survey explored the users' own views and perceptions of what they *think* they do.

Conference, workshop presentations. The work of the project has been presented at the following conferences during the reporting period:

- *OpCit tech seminar: Interoperable data for scholarly communication*, Southampton, July 2001
- *UKOLN/DNER Open Archives Meeting: Developing an agenda for institutional e-print archives*, London, July 2001
- *NSF Digital Libraries Initiative 2/IMLS Principal Investigators Meeting*, Roanoke, VA, June 2001
- *JISC/DNER Synthesis Meeting for International Digital Libraries Research Projects*, Bath, UK, May 2001
- *Workshop on The Open Archives Initiative (OAI) and Peer Review Journals in Europe*, Geneva, March 2001
- *OAI Open Meeting* held to mark the European public release of the specifications of the OAI interoperability architecture, Berlin, February 2001
- *Coordination meeting for arXiv mirrors*, Lyon, October 2000

and elsewhere <http://opcit.eprints.org/opcitpapers.shtml>

Detailed Progress and Future Plans

In Y3 the project will emphasise analysis, evaluation and dissemination. With its linking demonstrators, data mining results and user surveys, the project has amassed a rich resource that demands to be investigated more fully.

In addition the project will undertake to widen the scope of its interoperability activities, promoting OAI, EPrints and OpenURL. These activities are outlined with a plan of work for Y3 in Table 1.

Table 1. Plan of work for OpCit Y3

	Oct - Dec'01	Jan- Mar'02	Apr- Jun'02	Jul- Sep'02
Evaluation, analysis, dissemination of data mining, user survey, OpCit demonstrators and other OpCit results	■	■	■	■
Integrate OpCit with arXiv: develop and promote AMF	■			
Add OpenURL services: links to OpCit linked demonstrator; work with OpenURL resolver services; build an advanced OpenURL generator to turn references in PDF/TeX/LaTeX/HTML papers to OpenURL requests when viewed	■	■		
Advanced citation analysis– new measures of impact	■	■	■	
Implement and test EPrints components for reference checking	■	■		
Evaluate OpCit and Opsiis project software			■	
Migrate non-OAi archives (e.g. NCSTRL)	■			