

# DPC/PADI What's new in digital preservation

A joint service of the **Digital Preservation Coalition** and [PADI](#)

## Issue no. 1 (December 2001-April 2002)

compiled by Michael Day ([UKOLN](#), University of Bath)

3rd May 2002

---

This is a summary of selected recent activity in the field of digital preservation compiled from the [digital-preservation](#) and [padiforum-l](#) mailing lists and the Preserving Access to Digital Information ([PADI](#)) Gateway.

## 1. Organisations

### 1.1 The Digital Preservation Coalition

The Digital Preservation Coalition (DPC) was officially launched at the House of Commons on the 27 February 2002. This event was very successful and gained a large amount of press-coverage for digital preservation issues.

On the 25 March 2002 in London, the coalition organised a DPF Forum on Web-archiving. Presentations included a general introduction to Web-archiving issues and the UK Web domain; also descriptions of Web-archiving activity in the BBC and the Bibliothèque nationale de France. A workshop report and links to all presenters' PowerPoint slides are available on the DPC Web-site: [Webforum](#) page.

A more detailed review of recent DPC activity can be found in:

Neil Beagrie, "An update on the Digital Preservation Coalition," *D-Lib Magazine*, 8 (4), April 2002. <http://www.dlib.org/dlib/april02/beagrie/04beagrie.html>

### 1.2 The US National Digital Information Infrastructure and Preservation Program (NDIIPP)

This initiative began in late 2000, when Congress called for the Library of Congress (LC) to take the lead in a national collaborative planning effort for the long-term preservation of digital content. The April 2002 issue of *D-Lib Magazine* contained a progress report by Amy Friedlander (Council on Library and Information Resources). Friedlander outlines the results of some stakeholder meetings held last November, including the support for a national initiative from stakeholder groups that are not part of the traditional scholarly community, e.g. the entertainment industry. A research programme - which will be a key part of NDIIPP - also aims to be collaborative in nature and LC is already working with the National Science Foundation (NSF) and other federal agencies in drawing up a research agenda. An invitational workshop to discuss the research agenda was held in April 2002 in Washington and information is

being posted on a Web-site mounted at the University of Michigan (<http://www.si.umich.edu/digarch/>). An additional theme in NDIIPP is the importance of building operational systems. It is acknowledged that mistakes may be made, but that it is important to learn lessons from these. LC have also worked on devising a conceptual framework in order to see how the many and varied entities and functions related to the long-term preservation of digital content might interact. This is also described briefly in this paper. Amy Friedlander, "The National Digital Information Infrastructure Preservation Program: expectations, realities, choices and progress to date," *D-Lib Magazine*, 8 (4), April 2002.

<http://www.dlib.org/dlib/april02/friedlander/04friedlander.html>

### **1.3 OCLC/RLG Working Groups**

In April 2002, the OCLC/RLG Preservation Metadata Working Group published a proposed metadata element set for what the OAIS model refers to as 'Preservation Description Information' (PDI). Previous documents from the group had provided a state-of-the-art survey of preservation metadata activities and a recommendation for OAIS 'Content Information.' Publication of the PDI recommendation means that the group has almost completed its commissioned task. A final document bringing together both metadata recommendations is currently being compiled. All working group documents are available in PDF from:

<http://www.oclc.org/research/pmwg/documents.shtm>

The other joint OCLC/RLG digital preservation initiative, the Digital Archive Attributes Working Group, published a draft document entitled *Attributes of a Trusted Digital Repository* in August 2001. This has been very well received and is available in PDF from:

<http://www.rlg.org/longterm/attribswg.html>

## **2. Projects:**

### **2.1 The Cedars project**

Work on the Cedars (CURL Exemplars in Digital Archives) project finished in March 2002. The project had been going for almost four years and a final workshop was held in Manchester on the 25-26 February in order to disseminate information about the project, put that work into a wider context and to look forward to what should happen after the project had ended. A short summary of this event has been published in the April edition of RLG DigiNews, while a longer version is available on the Cedars Project Web-site:

Michael Day and Maggie Jones, *Cedars Final Workshop, Manchester Conference Centre, UMIST, Manchester, 25-26 February 2002*, Leeds: Cedars Project, 22 April 2002. <http://www.leeds.ac.uk/cedars/pubconf/umist/finalWorkshopRep.html>

Michael Day, "The Final Cedars Workshop: a report from Manchester, UK," *RLG DigiNews*, 6 (2) April 2002. <http://www.rlg.org/preserv/diginews/diginews6-2.html#conference>

In the first quarter of 2002, the Cedars project has also published a series of guides to various digital preservation issues. Available in print form (and in PDF) are guides to intellectual property rights, preservation metadata and digital collection management. Each of these is about 20 pages long, and are intended to provide non-technical introductions for anyone interested in aspects of digital preservation, including librarians, archivists, records managers and the creators of digital content. The guides describe some specific outcomes of the Cedars project (e.g. the draft metadata specification) but also attempt to provide a more general view and give indications of further reading. In the same series, a guide to digital preservation strategies is now available in HTML and an introduction to the Cedars digital archive prototype is under preparation. These guides are available in digital form (PDF or HTML) from the Cedars project Web-site:

<http://www.leeds.ac.uk/cedars/pubconf/pubconf.html>

## **2.2 ERPANET**

ERPANET (Electronic Resource Preservation and Access NETwork) has been funded by the European Commission to help bring together all types of organisation interested in digital preservation issues. It will primarily provide awareness about digital preservation by providing information and advice services, thematic workshops, training seminars, guidelines, etc. The project started in November 2001. Project partners are the Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow, the Schweizerisches Bundesarchiv (Swiss Federal Archives), the Rijksarchiefdienst (National Archives of the Netherlands) and the Institute for Archival and Library Science at the University of Urbino. More information on ERPANET can be found on the project's Web pages at:

<http://www.erpanet.org/>

## **2.3 Preservation of electronic scholarly journals**

The Andrew W. Mellon Foundation has funded seven major US libraries to investigate the development of digital repositories for e-journals. Work on these projects is continuing, but the Harvard University E-Journal Archiving project has recently (December 2001) published a report produced by Inera, Inc. on the feasibility of developing a common archival article Document Type Definition (DTD). The report recommended the creation of an XML DTD (or Schema) which would permit "successful conversion of significant intellectual content from publisher SGML and XML files into a common format for archival purposes." Also in December, the Harvard project published a draft proposal for the technical specifications of a Submission Information Package (SIP) that defined data formats, file naming conventions, metadata, etc. Both of these documents are available in PDF from the Digital Library Federation (DLF) Web-site:

Inera, Inc., *E-Journal Archive DTD feasibility study: commissioned by the Harvard University Library, Office for Information Systems, E-Journal Archiving Project*, 5 December 2001. <http://www.diglib.org/preserve/hadtdfs.pdf>

Harvard University Library, *Harvard E-Journal Archive: Submission Information Package (SIP) specification*, v. 1.0 draft, 19 December 2001.

<http://www.diglib.org/preserve/harvardsip10.pdf>

General information on the Mellon-funded programme can be found on the DLF Web-site:

<http://www.diglib.org/preserve/presjour.htm>

### **3. Other events**

A meeting of the US National Information Standards Organization (NISO) Book Industry Study Group (BISG) took place during the American Library Association's Midwinter 2002 Conference on the 20 January. This was entitled 'Archiving Electronic Publications' and included progress reports from two of the Mellon funded e-journal projects: Harvard University's E-Journal Archiving project and Elsevier Science's collaboration with Yale University Library. A final presentation reported on collaboration between OCLC and the US Government Printing Office (GPO) on a Web Document Digital Archive pilot project. A short summary of the meeting can be found at:

<http://www.niso.org/presentations/niso-bisg-rpt.html>

### **4. Other recent publications:**

Michael K. Bergman, "The deep Web: surfacing hidden value," *Journal of Electronic Publishing*, 7 (1), August 2001. <http://www.press.umich.edu/jep/07-01/bergman.html>

This 'white paper' is concerned with the so-called 'deep Web,' whereby information is buried deep within dynamically generated sites and which can not, therefore, be easily reached by standard search engines. The paper is essentially marketing a product (search technology from a company called BrightPlanet) and is not about preservation, but it may be able to inform harvesting-based Web-preservation initiatives on the nature of dynamic or database-driven Web-sites.

Hilary Berthon, Susan Thomas and Colin Webb, "Safekeeping: a cooperative approach to building a digital preservation resource," *D-Lib Magazine*, 8 (1), January 2002. <http://www.dlib.org/dlib/january02/berthon/01berthon.html>

This paper describes the National Library of Australia's Safekeeping project, which has funding from the Council on Library and Information Resources (CLIR). The project is trying to facilitate a distributed network of 'safekept' resources relating to digital preservation (selected from the PADI database) by encouraging resource owners to take responsibility for providing long-term access - or to nominate third parties who could do so on their behalf. The co-operative model of the Safekeeping project is interesting because it might encourage the creators and owners of resources to face up to the responsibilities that they hold with regard to maintaining long-term access.

Stewart Granger, "Digital preservation and deep infrastructure," *D-Lib Magazine*, 8 (2), February 2002. <http://www.dlib.org/dlib/february02/granger/02granger.html>

This is an 'opinion' piece by Stewart Granger of the University of Leeds.

Anne R. Kenney, Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze and Sandra Payette, "Preservation risk management for Web resources: virtual remote control in Cornell's Project Prism," *D-Lib Magazine*, 8 (1), January 2002.

<http://www.dlib.org/dlib/january02/kenney/01kenney.html>

This paper suggests that Web preservation strategies could use risk management methodologies. It is based on the work of Cornell University's Project Prism, funded as part of the second phase of the US Digital Libraries Initiative.

Julia Martin and David Coleman, "Change the metaphor: the archive as an ecosystem," *Journal of Electronic Publishing*, 7 (3), April 2002.

<http://www.press.umich.edu/jep/07-03/martin.html>

The authors of this paper are researchers at the University of New South Wales and the University of Sydney. The paper argues that there is unlikely to be any single solution to the digital preservation problem but that rapid technological change will mean that preservation solutions will need to be in a state of constant change.

Michael L. Nelson and B. Danette Allen, "Object persistence and availability in digital libraries," *D-Lib Magazine*, 8 (1), January 2002.

<http://www.dlib.org/dlib/january02/nelson/01nelson.html>

This paper - produced by researchers working at the NASA Langley Research Center - looked at the persistence and continued availability of 1,000 digital library objects. These were mostly found in Web-based e-print services like arXiv, CogPrints and PubMed Central. The authors found that in just over one year, 3% of the tested objects no longer appeared to be available. With an assumption that objects placed in e-print services should persist longer than the average Web page, the authors cautiously conclude that this finding may have relevance for those concerned with long-term preservation. However, Nelson and Allen consider that more detailed studies of digital library object persistence need to be made.

Elizabeth Yakel, "Digital preservation," *Annual Review of Information Science and Technology*, 35, 2001, 337-378.

A general overview of digital preservation issues by an assistant professor in the School of Information at the University of Michigan.

## 5. Other links:

From the Digitale Duurzaamheid Digital Preservation Testbed

(<http://www.digitaleduurzaamheid.nl/>):

*Migration context and current status*. Digital Preservation Testbed White Paper, 5 December 2001. <http://www.digitaleduurzaamheid.nl/bibliotheek/Migration.pdf>

*Approaches towards the long term preservation of archival digital records*. Digital Preservation Testbed Infosheet, v. 1.7, 19 September 2001.

<http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=186&categorie=2>

Also (from the *digital-preservation@jiscmail.ac.uk* and *padiforum-l@nla.gov.au* e-mail lists):

Andreas Aschenbrenner, *Long-Term Preservation of digital material - building an archive to preserve digital cultural heritage from the Internet*, Masters Thesis, Technical University Vienna, December 2001. Available in various formats from: <http://www.ifs.tuwien.ac.at/~aola/publications.html>

Arthur Smith, *Long Term Archiving of Digital Documents in Physics*, report of an IUPAP (International Union of Pure and Applied Physics) Conference held in Lyon, 5-6 November 2001. [http://publish.aps.org/IUPAP/ltaddp\\_report.html](http://publish.aps.org/IUPAP/ltaddp_report.html)

Dollar Consulting, *Archival preservation of Smithsonian web resources: strategies, principles, and best practices*. Washington, D.C.: Smithsonian Institution Archives, 20 July 2001. <http://www.si.edu/archives/archives/dollar%20report.html>

VERS (Victorian Electronic Records Strategy) Web-site:  
<http://www.prov.vic.gov.au/vers/>

---

*Problem links last disabled: 01 October 2008*

Warning! Website links tend to have very short lifetimes - see [web page](#) for further info.