DPC/PADI What's new in digital preservation

Issue no. 4 (November 2002-January 2003)

compiled by Gerard Clifton (National Library of Australia)
and Michael Day (UKOLN, University of Bath)

20th February 2003

This is a summary of selected recent activity in the field of digital preservation compiled from the Preserving Access to Digital Information (PADI) Gateway and the digital-preservation and padiforum-l mailing lists. Additional or related items of interest may also be included.

1. Organisations

2. Projects

3. Specific Areas of Activity

4. Events

5. Other recent publications

1. Organisations

1.1 Joint Information Systems Committee (JISC)

The UK Joint Information Systems Committee (JISC) adopted a 3-year continuing access and digital preservation strategy in October 2002. The strategy proposes the role that JISC should undertake on behalf of the funding councils and institutions as part of a national digital preservation programme. The full strategy document can be downloaded in RTF.

Neil Beagrie, *A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-2005*. Joint Information Systems Committee (JISC), London, 1 October 2002.
Executive Summary available (in HTML) at:
http://www.jisc.ac.uk/index.cfm?name=pres_continuing
Full Strategy available (in RTF) at:
http://www.jisc.ac.uk/uploaded_documents/dpstrategy2002b.rtf

Details of the JISC Digital Preservation and Records Management Programme 2002-3 have also been added to the JISC Web site. This includes details of recent or current project and activities undertaken or supported by the JISC Digital Preservation Focus as part of its implementation of the JISC Continuing Access and Digital Preservation Strategy. Details and deliverables for the programme will continue to be added and announced individually - the current list of projects and feasibility studies includes: Archiving E-Publications, Web-archiving, Archiving of E-prints,

Curation of e-science Primary Research Data, File Format and Rendering, and Revision of the Records Life Cycle Report.

JISC Digital Preservation and Records Management Programme: http://www.jisc.ac.uk/index.cfm?name=about_preservation

JISC also launched its new web site on 5 February 2003, with improved navigation, searching and customisation facilities. The new site contains information on collections, projects, funding opportunities, advice and guidance, including that absorbed from the former DNER, and will continue to evolve over time to meet user needs.

JISC Web site: http://www.jisc.ac.uk/

1.2 Medical Research Council

In a bid to encourage and promote the sharing and preservation of research data, the UK Medical Research Council (MRC) is launching an initiative that aims to establish a set of collaborative, science-led activities to produce generic data sharing and preservation solutions for the research community. The MRC is keen to bring together the best appropriate expertise from among data-set creators, users, archivists and document management and informatics experts. MRC's draft policy statement and guidance on the initiative are to be found on its Web site.

MRC Data sharing and preservation: http://www.mrc.ac.uk/txt/index/strategy/strategy-science_strategy/strategy-strategic_implementation/strategy-data_sharing.htm

1.3 Digital Preservation Coalition

A Survey of Information Technology Vendors prepared for the Digital Preservation Coalition (DPC) is now available on the DPC Web site. The study was conducted between July and October 2002 to help promote the DPC's aim of working with industry, and followed a forum held in June 2002 with a cross-section of industry on future research and development for digital asset preservation. The survey's primary objective was to establish the extent of awareness and perception within industry of its need to become more aware of and engaged with issues of digital longevity; and their views on how the DPC can interact with industry, improve awareness, and foster partnerships. The public version of the report makes the management summary and findings available online and provides an interesting overview of opinion from across-section of industry vendors of both hardware and software on digital retention and preservation.

Extracts from: Philip Lord, *A Survey of Information Technology Vendors* (28 October 2002): http://www.dpconline.org/graphics/reports/

1.4 National Digital Information Infrastructure and Preservation Program

A press release from the Library of Congress announced approval from the US Congress for its "Plan for the National Digital Information Infrastructure and Preservation Program (NDIIPP)."

Library of Congress, *Library Announces Approval of Plan to Preserve America's Digital Heritage*, Press Release (14 February 2003): http://www.loc.gov/today/pr/2003/03-022.html

The NDIIP Plan is available at: http://www.digitalpreservation.gov/ndiipp/repor/repor_plan.html

2. Projects

2.1 InterPARES Project

The first phase of the InterPARES project (International Research on Permanent Authentic Records in Electronic Systems), now known as InterPARES 1, was completed in 2001 and the consolidated project findings became available during 2002. A second phase of the project, InterPARES 2, began on 1 January 2002 and will continue until 31 December 2006.

InterPARES 1 focussed on the authenticity of non-current records from databases or document management systems that resulted from administrative and legal activities and which were destined for permanent preservation. InterPARES 2 differs in that it will address issues of reliability and accuracy in addition to issues of authenticity throughout the life cycle of records, and will focus on records produced in dynamic, interactive environments resulting from artistic, scientific and government activities.

The InterPARES 2 web site contains details of project's scope and participants:

InterPARES 2: Experiential, Interactive and Dynamic Records . InterPARES 2 Project, 2002. Available at: http://www.interpares.org/ip2_index.cfm

2.2 Task Force in Digital Repository Certification

The Task Force in Digital Repository Certification is a joint task force created by RLG and NARA, as part of ongoing work with the OAIS reference model. The Task Force is charged to consult regarding standards, criteria, and mechanisms for certification, to identify certifiable elements, to create a standard certification process or framework, to develop a certification plan and to identify the conditions for revocation of certification.

A description of the project is available at: http://www.rlg.org/longterm/certification.html

2.3 PDF-Archive Project (PDF/A)

The PDF-Archive project is a joint activity of the Association for Suppliers of Printing, Publishing and Converting Technologies (NPES) and the Association for Information and Image Management, International (AIIM, International), this project aims to develop an international standard defining the use of Adobe's Portable Document Format (PDF) for archiving and preservation of electronic documents. The project will address support of multipage documents featuring combinations of text and graphics and the requirements for reading devices to render archived documents. The project's web site contains presentations, meeting minutes and a draft standard for comment.

Project web site: http://www.aiim.org/standards.asp?ID=25013

The Draft ISO Standard for Comment is available in PDF format from the same site: International Organization for Standardization (ISO). *Document - PDF-Archive - Operational Use*, Draft 1, 3 December 2002.
http://www.aiim.org/documents/standards/ISO_PDFArchive_(E).pdf

2.4 Digital Archive for Chinese Studies (DACHS)

Based at the University of Heidelberg's Institute for Chinese Studies, DACHS "aims at identifying, archiving and making accessible Internet resources relevant for Chinese Studies, with special emphasis on social and political discourse as reflected by articulations on the Chinese Internet." The archive has been operating since August, 2001, and collected resources include websites, discussion boards, journals, newsletters and single documents. On overview of the archive's collection policy, workflow and technical infrastructure is available.

Visit the DACHS website at the Institute of Chinese Studies, University of Heidelberg: http://www.sino.uni-heidelberg.de/dachs/

A brief overview article about the archive and its aims appears in the January 2003 issue of *D-Lib Magazine*:

Hanno Lecher, "Preserving the Chinese Internet: The DACHS Project " *D-Lib Magazine* 9(1), January 2003.
http://www.dlib.org/dlib/january03/01inbrief.html#LECHER

2.5 DigiCULT reports

The DigiCULT Project has recently issued two reports focused on topics investigated by experts at round tables organised by the DigiCULT secretariat. The first of these looks at the preservation of the integrity and authenticity of digital objects, and contains a position paper by Seamus Ross (HATII, University of Glasgow) and a summary of the DigiCULT Forum held in Barcelona in May 2002. The second concerns digital asset management systems (DAMS) and results from a DigiCULT Forum held in Essen in September 2002.

Integrity and Authenticity of Digital Objects, Thematic Issue 1, DigiCULT Consortium, August 2002: http://www.digicult.info/downloads/thematic_issue_1_final.pdf

Digital Asset Management Systems for the Cultural and Scientific Heritage Sector, Thematic Issue 2, DigiCULT Consortium, December 2002:
http://www.digicult.info/downloads/thematic_issue_2_021204_low_resolution.pdf

3. Specific Areas of Activity

There are several areas of activity in which a number of papers have been released recently.

3.1 Web Archiving

Two articles in the December issue of *D-Lib Magazine* have highlighted innovative methods in harvesting and analysing web-based materials.

The result of experiments at the **Bibliothèque nationale de France** (BnF) to improve the crawler programs used for the automatic collection of Web pages are described in one article. The automatic ranking of a sample set of web sites by sophisticated crawler programs that use the link structure of the web to provide an estimation of a web page's importance was compared with manual evaluations of the same sites by librarians. There was a sufficient level of correlation from these to suggest that automatic importance evaluation could help national

libraries build large-scale collections of Web resources.

Julien Masanès, "Towards Continuous Web Archiving: First Results and an Agenda for the Future" *D-Lib Magazine* 8(12), December 2002.
http://www.dlib.org/dlib/december02/masanes/12masanes.html

Researchers at the **Austrian On-Line Archive** (AOLA) established a data warehouse in the context of their archive. Using on-line analytical processing (OLAP) techniques, they were able to analyse and investigate various technical aspects of the archive, such as operating systems and servers used, the variety of file types, forms and scripting languages and link structures within domains. Such analyses may be useful in drawing inferences about technology maturation and impacts on communities.

Andreas Rauber, Andreas Aschenbrenner, Oliver Witvoet, Robert M. Bruckner, Max Kaiser, "Uncovering Information Hidden in Web Archives: A Glimpse at Web Analysis Building on Data Warehouses" *D-Lib Magazine* 8(12), December 2002.
http://www.dlib.org/dlib/december02/rauber/12rauber.html

The Andrew W. Mellon Foundation has also awarded a grant to the Center for Research Libraries (CRL) for a project entitled **Political Communications Web Archiving**. Cornell University Library will be one of four universities subcontracting with CRL to explore the curatorship, technology, and long-term resource management components of the project.

Further information is available (in PDF) from the Cornell University Library Web site at:
http://www.library.cornell.edu/iris/research/WebPolCom.pdf

3.2 Preservation Metadata

Preservation Metadata at the National Library of New Zealand - Te Puna Matauranga o Aotearoa

The National Library of New Zealand has developed a preservation metadata schema to help support the preservation of digital objects by the library. It will help form the basis for the design of a metadata repository and associated input systems. The schema will help to identify digital objects for which preservation responsibility has been assumed, will store information about decisions and actions undertaken, will document information about preservation processes and record their effects and will help ensure their authenticity. The schema also includes those elements needed to manage the metadata record itself.

The schema has been developed in the light of wider preservation metadata activities, including the draft schemas produced by the National Library of Australia and the UK Cedars project, the work of the OCLC/RLG Preservation Metadata Working Group and the development of the OAIS Model. Appendices compare the NLNZ schema with the OCLC/RLG framework, the NISO Z39.87 draft standard and the NLA's draft schema.

Metadata Standards Framework - Preservation Metadata. Wellington, National Library of New Zealand, November 2002.
Available in PDF format at: http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf

Steve Knight, Preservation metadata and digital continuity, *DigiCULT.info*, Issue 3, February 2003, 18-20.
Available in PDF at: http://data.digicult.info/download/digicult_info3_low.pdf

Tsinghua University Library Metadata Framework

An article in the November issue of *D-Lib Magazine* describes a metadata framework developed at Tsinghua University, China, to aid in the preservation of digital resources. The framework was developed as part of the EMANI project with Goettingen State and University Library, Germany, Cornell University Library in the United States and Orsyell Library, France. The framework includes modules for rights, technical, descriptive, source and process metadata, allows description of resource structure and utilises the Metadata Transmission and Encoding Standard (METS).

Jinfang Niu, "A Metadata Framework Developed at the Tsinghua University Library to Aid in the Preservation of Digital Resources." *D-Lib Magazine* 8(11), November 2002.
http://www.dlib.org/dlib/november02/niu/11niu.html

Review of RLG/OCLC Preservation Metadata

Hans Hofman of ERPANET provides a review and critique of the final report by the OCLC/RLG Working Group on Preservation Metadata, *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects*. The review appears in the October 2002 issue of the *DigiCULT.info* newsletter and presents some perceived limitations of the framework, including the need for greater attention to both the relationship between preservation processes and metadata and the requirements of the records and archival community in terms of record authenticity, context and relationships. The article is available by accessing or downloading the entire *DigiCULT.info 2* issue (PDF format).

Hans Hofman, "Review: Some Comments on Preservation Metadata and the OAIS Model." *DigiCULT.info Newsletter*, Issue 2, October 2002, 15-20.
DigiCULT publications page: http://www.digicult.info/pages/publications.php
DigiCULT.info, Issue 2:
http://www.digicult.info/downloads/digicult_info2.pdf

3.3 Institutional Repositories

Several papers are available describing the rationale, advantages and practical implementations of institutional repositories for capturing and preserving scholarly output. A report produced by the Scholarly Publishing and Academic Resources Coalition (SPARC) examines the benefits of developing institutional repositories and how the trend towards the self-publishing of primary research materials affects the roles of traditional scholarly publishers, academic libraries and specific faculties. An article by Roy Tennant in *Library Journal* provides an overview of several institutional initiatives, touching on software, implementations and metadata, and a new article about DSpace, a collaborative, open source repository project between MIT Libraries and Hewlett Packard Laboratories, features in the January 2003 issue of *D-Lib Magazine*.

Raym Crow, *The Case for Institutional Repositories: A SPARC Position Paper*. Washington

D.C., The Scholarly Publishing & Academic Resources Coalition, 2002.
Available in PDF format at: http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf

Roy Tennant, "Institutional Repositories", *Library Journal*, 15 September 2002
http://libraryjournal.reviewsnews.com/index.asp?layout=articleArchive&articleid=CA242297

MacKenzie Smith, Mick Bass, Greg McClellan, Robert Tansley, Mary Barton, Margret Branschofsky, Dave Stuve and Julie Harford Walker, "DSpace: An Open Source Dynamic Digital Repository", *D-Lib Magazine*, 9(1), January 2003.
http://www.dlib.org/dlib/january03/smith/01smith.html
The DSpace project web site is at: http://dspace.org/

A press release from the **University of Cambridge** announced a joint project between the Cambridge University Library and the Massachusetts Institute of Technology (MIT) Libraries to establish a digital repository based on the DSpace software. The press release says that the Cambridge system - to be known as 'DSpace@Cambridge' - will have two main roles. Firstly, it "has the ability to capture, index, store, disseminate and preserve digital material created by the academic community, including sdholarly [sic] articles and pre-prints, theses, technical reports, archives and other textual material, together with different formats such as multimedia clips, interactive teaching programmes, datasets and databases. Secondly, it will provide a home for the increasing amount of material that is being digitised from the University Library's collections." University of Cambridge, *Preserving our Past: a joint digital repository project for University and MIT*, Press Release (10 February 2003):
http://www.admin.cam.ac.uk/news/press/dpp/2003021001

Another important development was the launch of **DARE (Digital Academic Repositories)**, a collective initiative by universities in the Netherlands to make all of their research results digitally accessible. The initiative is a collaboration that includes the Koninklijke Bibliotheek [National Library of the Netherlands], the Koninklijke Nederlandse Academie van Wetenschappen (KNAW) [Royal Netherlands Academy of Arts and Sciences] and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) [Netherlands Organisation for Scientific Research]. Co-ordination is provided by the SURF Foundation, the ICT partnership organisation for higher education and research in the Netherlands.

More information on DARE is available (in Dutch) on the SURF Web site:
http://www.surf.nl/themas/index2.php?oid=18&inhoud_oid=107

Information in English: http://www.surf.nl/en/themas/index2.php?oid=7

4. Events

4.1 Recent Events

Reports, papers or presentations are available for several recent events:

Leonid A. Kalinichenko, Vladimir V. Korenkov, Vladislav P. Shirikov, Alexey N. Sissakian and Oleg V. Sunturenko, "Digital Libraries: Advanced Methods and Technologies, Digital Collections - Report on RCDL'2002 - The 4th All-Russian Scientific Conference, Dubna, 15 - 17 October 2002", *D-Lib Magazine*, 9(1), January 2003.

http://www.dlib.org/dlib/january03/kalinichenko/01kalinichenko.html

Sally Jo Cunningham, "Report on the Fifth International Conference on Asian Digital Libraries (ICADL 2002) - 11-14 December 2002, Singapore", *D-Lib Magazine,* 9(1), January 2003.
http://www.dlib.org/dlib/january03/cunningham/01cunningham.html

Presentations (in Portuguese and English) from *Workshop on Digital Preservation: Experiences and Strategies - Encontro sobre Preservação Digital: Experiências e Estratégias*, Biblioteca Nacional, Lisboa, Portugal, 25 November 2002.
Available at: http://www.bn.pt/agenda/ecpa/ecpa-programa.html

Presentations from ERPANET Training Seminars:

> *OAIS Training Seminar, København, Thursday 28th - Friday 29th November, 2002, The Black Diamond, The Royal Library, Denmark*, ERPANET, 2002.
> http://www.erpanet.org/

> *Policies for Digital Preservation, Wednesday 29th - Thursday 30th January, 2003, Fontainebleau, Paris, France*, ERPANET, 2003.
> http://www.erpanet.org/

Presentations and Further Proceedings from the Workshop on Digital Preservation in Copenhagen, December 11, 2002. Danish Ministry of Culture, Copenhagen, December 2002.
Available at: http://www.kum.dk/sw5312.asp

Digital Preservation: Technology and Policy - Invitational Workshop, Koninklijke Bibliotheek, The Hague, Netherlands, 13 December, 2002.
http://www.kb.nl/kb/ict/dea/workshop2002/workshop.html

> Several reports from this workshop are of particular interest, including Raymond Lorie's proof-of-concept report for the Universal Virtual Computer (UVC) (PDF format): Raymond Lorie, *The UVC: a Method for Preserving Digital Documents - Proof of Concept*. The Hague, IBM and Koninklijke Bibliotheek, December 2002.
> http://www.kb.nl/kb/ict/dea/ltp/reports/4-uvc.pdf

The meeting report and presentations from the DPC Forum on Preservation of e-Learning Materials and Cost Models for Digital Preservation held on 15 October 2002 are now available on the DPC Web site:
http://www.dpconline.org/graphics/events/index.html#elearning

The 9th ALPSP Publishing and the Internet Seminar on Archiving - was held on 29 November 2002 and was jointly organised with the DPC. The programme and presentations are online at:
http://www.alpsp.org/default.htm

4.2 Forthcoming Events

February

"Web-Wise 2003: Sustaining Digital Resources", 26-28 February, 2003, Washington DC, USA
http://webwise.mse.jhu.edu/

Society of American Archivists Seminar - "Archival Perspectives in Digital Preservation", 27-28 February, 2003, University of Arizona Library, Tucson, Arizona, USA
http://www.archivists.org/prof-education/workshop-detail.asp?id=499

March

DLib Center Workshop - "Permanent Access to On-line Journals", 4-5 March, 2003, Pisa, Italy
http://dlibcenter.iei.pi.cnr.it/en/courses4.shtml

RLG-JISC Joint Symposium - "Selection and Collaboration in Digital Preservation Symposium", 24-25 March, 2003, Library of Congress, Washington DC, USA
http://www.rlg.org/events/rlgjisc2003-agenda.html

April

PRO Conference - "Practical Experiences in Digital Preservation", 2-4 April, 2003, Public Record Office, Kew, UK
http://www.pro.gov.uk/about/preservation/digital/conference/default.htm

ERPANET Workshop - "Long Term Preservation of Databases", 9-11 April, 2003, Swiss Federal Archives, Berne, Switzerland
http://www.erpanet.org/

May

"JCDL 2003 - The Third ACM+IEEE Joint Conference on Digital Libraries", 27-31 May, 2003, Houston, Texas, USA
http://www.jcdl.info/jcdl03/

5. Other recent publications

Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH), *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, Version 1.0, New York, NINCH, December 2002.
http://www.nyu.edu/its/humanities/ninchguide/index.html

> A guide developed by an expert working group on current best practice on the capture, preservation and management of digitised and networked digital resources. The guide provides a useful chapter on digital preservation.

Henry M. Gladney, "Digital Preservation Technical Solution" *Digital Document Quarterly*, 1(4). Saratoga, HMG Consulting, 2002
Available at: http://home.pacbell.net/hgladney/ddq_1_4.htm

> Henry Gladney comments on the perception that digital resources present more difficult preservation problems than analogue media and considers requirements analyses and the

characteristics of preservation solutions. A series of papers covering in detail the Universal Virtual Computer (UVC) approach as a proposed technical solution to the problems of digital preservation is announced. "Trustworthy 100-Year Digital Objects: Evidence Even After Every Witness is Dead" is the first in the series of papers and is available via email request from the author. Refer to the article for the contact address

Daniel Lee, "JPEG 2000" *RLG DigiNews*, 6(6), December 2002
http://www.rlg.org/preserv/diginews/diginews6-6.html#interview

> An interview with Dr Daniel Lee, convenor of ISO's JPEG group, about JPEG 2000, an emerging alternative image encoding format offering advantages over the current JPEG standard, including lossless compression and support for multiple resolutions.

John Helly, Hubert Staudigel and Anthony Koppers, "Scalable models of data sharing in earth sciences." *Geochemistry, Geophysics, Geosystems*, Vol. 4, Article No. 1010 (25 January 2003).
Abstract available at: http://www.agu.org/pubs/crossref/2003/2002GC000318.shtml
E-print (PDF) available at: http://www.sdsc.edu/~hellyj/papers/datasharing.pdf

> Discusses options for the publication, archiving and exchange of scholarly data in earth science disciplines. Proposes a Metadata Interchange Format (.mif) that could be used for the effective sharing of data and metadata across digital libraries.