DPC / PADI What's new in digital preservation

Issue no. 16 (April - August 2007)

Compiled by Najla Semple for the Digital Preservation Coalition (DPC) and Gerard Clifton (National Library of Australia)

19 September 2007

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.

Contents:

1. News from organisations and initiatives
   1.1 Digital Preservation Coalition (DPC) (UK)
   1.2 Digital Curation Centre (DCC) (UK)
   1.3 JISC (UK)
   1.4 The National Archives, UK
   1.5 The National Library of Scotland
   1.6 PLANETS
   1.7 nestor (Network of Expertise in Long-Term Storage of Digital Resources)
   1.8 DigitalPreservationEurope (DPE)
   1.9 CASPAR
   1.10 National Library the Netherlands (Koninklijke Bibliotheek)
   1.11 US National Digital Information Infrastructure and Preservation Program (NDIIPP)
   1.12 FEDORA Commons

2. Specific themes
   2.1 Digital Preservation – definitions, frameworks and directions
   2.2 Copyright
   2.3 Repositories
   2.4 Institutional Repositories
   2.5 Electronic Records Management
   2.6 Web Archiving
   2.7 Digital Objects
   2.8 Preservation Metadata
   2.9 Storage Media
   2.10 Tools

3. Other publications
1. News from organisations and initiatives

1.1 Digital Preservation Coalition (DPC)

Five projects have recently been selected for the shortlist for the 2007 Digital Preservation Award. The Award, which recognises achievement and encourages innovation in digital preservation, will be announced at the Conservation Awards Presentation evening at the British Museum on 27 September, 2007. The Shortlist of projects is:

- **LIFE**: *The British Library*. LIFE (Lifecycle Information for E-Literature) models the digital lifecycle and calculates the costs of preserving digital material, allowing organisations to understand costs and focus resources where most needed. Retrieved 19 September 2007 from: [http://www.life.ac.uk/](http://www.life.ac.uk/)


- **Active Preservation at The National Archives - PRONOM Technical Registry and DROID file format identification tool**: *The National Archives of the UK*. PRONOM is an online service which includes a knowledge base of technical information about over 600 file formats and 250 software tools, and can be used with the DROID file format identification tool to understand the technical dependencies of digital collections. Retrieved 19 September 2007 from: [http://www.nationalarchives.gov.uk/pronom/](http://www.nationalarchives.gov.uk/pronom/)

- **PARADIGM (The Personal Archives Accessible in Digital Media)**: *Bodleian Library, University of Oxford, & John Rylands University Library, University of Manchester*. The Paradigm project has worked with politicians, archivists and researchers to investigate the challenges of preserving personal archives of born-digital materials such as emails, photographs and blogs. Retrieved 19 September 2007 from: [http://www.paradigm.ac.uk/](http://www.paradigm.ac.uk/)

- **Digital Repository Audit and Certification**: *CRL, RLG-OCLC, NARA, the DCC, DPE and nestor*. The Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist (maintained by the US Center for Research Libraries), the nestor project's Criteria Catalogue and the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) published by the Digital Curation Centre and DigitalPreservationEurope present complementary methods for the self assessment, audit and certification of digital repository infrastructures.
Frances Boyle has been appointed as the new DPC Executive Director and will begin work in September. Frances has previously worked as IT Development & Strategy Manager at Oxford University Libraries and has been involved in numerous digital preservation projects. She looks forward to working with colleagues and contributing to the work of the DPC.

1.2 Digital Curation Centre (DCC)

The second International Journal of Digital Curation (IJDC) has been published. Policy and practical experience are the two main themes of this issue, which includes papers on policies for trusted digital repositories, OAIS compliance and practical approaches to preservation of records, as well as articles from CASPAR, the Digital Curation Centre and a report on the DigCCurr 2007 conference. The Journal is available from the IJDC website. Retrieved 19 September 2007 from: http://www.ijdc.net/ijdc/issue/current

Another in the DCC Standards Watch series of papers is now available, this time covering records management.

Standards watch paper:


A new digital preservation blog has also been set up by the Digital Curation Centre, retrieved 19 September 2007 from: http://digitalcuration.blogspot.com/

1.3 JISC

A number of studies and projects are being undertaken with the support of JISC funding:

- **InSPECT**: JISC has funded five studies into the significant properties of digital objects in relation to preservation. ‘Investigating the Significant Properties of Electronic Content Over Time’ (InSPECT), led by the Arts and Humanities Data Centre (AHDS) in association with The National Archives (TNA), will explore the concept of ‘significant properties’ and will analyse a range of digital objects to develop a generalised methodology for determining the significant properties of digital object types. Retrieved 19 September 2007 from: http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/INSPECT

Four further studies investigate the significant properties of four different types of digital object for digital preservation purposes: e-learning objects (led by ULCC), software (led by the eScience Centre), and vector images and moving images (led by System Simulation Ltd).
- **Preserv2**: Building on the PRONOM-ROAR service developed in Preserv1, which identified format profiles for more than 200 repositories, Preserv2 will investigate structured processes for active preservation aimed at repository content, including characterisation, preservation planning and preservation action. Retrieved 19 September 2007 from: [http://preserv.eprints.org/](http://preserv.eprints.org/)

- **Sherpa DP2** will extend the collaborative, shared preservation environment developed by the Sherpa DP project. The project will build on that work by extending the implementation model to interact with repositories holding different and varied types of digital content and by using a more diverse range of content management systems. A tool for creating METS packages will be developed, and automation of digital object preservation will be investigated. Retrieved 19 September 2007 from: [http://www.sherpadp.org.uk/sherpadp2.html](http://www.sherpadp.org.uk/sherpadp2.html)

- **SOAPI** (Service-Oriented Architecture for Preservation and Ingest of digital objects) will develop an architecture and toolkit for (partially) automating preservation and ingest workflows in digital repositories, based on a set of atomic Web Services, each encapsulating a unit of preservation functionality, combined with Web-based forms when human input is required. Retrieved 19 September 2007 from: [http://ahds.ac.uk/about/projects/soapi/index.htm](http://ahds.ac.uk/about/projects/soapi/index.htm)

- **REMAP** will develop a model and a tool to embed records management and preservation within the repository workflow, using appropriate Web services and fulfilling a range of institutional needs from ingest to export. Retrieved 19 September 2007 from: [http://www.hull.ac.uk/remap/](http://www.hull.ac.uk/remap/)

- **Preserving and Accessing Software Research Outputs**: Research data may often be of little use without the specific software developed to process it, or a piece of software may itself be the primary output of research. This project will investigate issues around the repository deposition and preservation of software artefacts, developing guidelines for the preservation of software research outputs. Further information is available via the JISC website. Retrieved 19 September 2007 from: [http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/tools/software.aspx](http://www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres/tools/software.aspx)

- **LIFE2**: The LIFE (Lifecycle Information for E-Literature) Project has developed a methodology to model the digital lifecycle and calculate the costs of preserving digital information for the next 5, 10 or 100 years. LIFE2 is rigorously testing the LIFE methodology via an independent review. The project is also re-examining the LIFE Preservation Model in light of the findings from LIFE1, as well adopting four new exemplar case studies which will expand the applicability of the model. The new case studies will apply the model to institutional repositories, primary data, and a digitised newspaper collection. The LIFE team is also working with other institutions (for example the Royal Library of Denmark) to help in the adoption of the LIFE methodology. Retrieved 19 September 2007 from: [http://www.life.ac.uk/](http://www.life.ac.uk/)
1.4 The National Archives, UK

The National Archives [TNA] and Microsoft have recently signed an agreement to help solve the problems of legacy file formats. Microsoft will provide the TNA with a system that incorporates past versions of software, powered by Microsoft Virtual PC 2007. More information is on the National Archives Web site. Retrieved 19 September 2007 from: http://www.nationalarchives.gov.uk/news/stories/164.htm

There is also further press coverage of the agreement:


A recent UK government petition called for open source to be more widely implemented, with the Open Document Format being put forward as a standard to be adopted by the UK government. The petition and response, noting the role of the UK National Archives, is available via the UK Government Web site, retrieved 19 September 2007 from: http://petitions.pm.gov.uk/OpenDocument/

The government also announced plans to seriously tackle digital preservation, the Cabinet Secretary releasing a statement about the vulnerability of government information. See the press release via the National Archives Web site, retrieved 19 September 2007 from: http://www.nationalarchives.gov.uk/news/stories/161.htm

1.5 The National Library of Scotland

The National Library of Scotland (NLS) has begun work on its Scottish Executive funded Trusted Digital Repository (TDR) project, which it plans to complete by September 2008. There are four key milestones:

- **Milestone 1:** Implementation of a web archiving system based on open source toolkits in order to commence selective archiving and to develop an effective set of operational processes to manage its growing web collections.

- **Milestone 2:** Development of an ingest and metadata management module to allow for descriptive and preservation metadata management in the NLS Fedora repository installation.

- **Milestone 3:** Development and roll out of a workflow management application to control and automate the management tasks associated with the web archive.

- **Milestone 4:** Implementation of a resource discovery and delivery mechanism for items held within NLS digital collections, taken from within the repository.

In parallel with its technical implementation plan, the project is actively involved in
producing a strategic long-term plan for the operation and integration of the repository systems as part of day to day life at NLS. The TDR project team are currently undertaking detailed investigations of digital repository operations across the globe with the aim of producing a trustworthy business model to support a ‘best in class’ digital preservation repository.

1.6 PLANETS (Preservation and Long-term Access through Networked Services)

The Planets Year 1 review took place on 23-24 July 2007. Planets technical achievements from the first year include:

- Significant advances in an infrastructure for storing and accessing information about file formats
- The definition of new XML-based description languages for file format characterization and content representation
- Implementation of file migration tools as Web Services, allowing them to be included in distributed preservation workflows
- Conceptual and technical requirements and design for the Planets Preservation Planning Tool and the Planets Testbed, applications that will support the librarian or archivist in selecting digital preservation strategies
- Prototypical implementation of the Planets Interoperability Framework, a software bundle that supports the administration, configuration, and deployment of preservation services and workflows

The focus in the project’s second year will be on completing the initial design and development, taking the establishment of the interoperability framework into the next phase, starting to extend access, formatting range and scope of services, and on increasing the active engagement with the user and vendor communities. Key deliverables from the project’s first year will be published shortly on the Planets Web site. Retrieved 19 September 2007 from: [http://www.planets-project.eu](http://www.planets-project.eu).

1.7 nestor (Network of Expertise in Long-Term Storage of Digital Resources)

Nestor, in conjunction with the Digital Curation Centre, DigitalPreservationEurope and the Centre for Research Libraries have recently published an agreed set of ten core principles for trusted digital repositories. An English version of the Core Requirements is available from the Centre for Research Libraries website.


The German National Library and SUB Göttingen have released the kopal Library for Retrieval and Ingest (koLibRI). Kopal was a recently completed project to find a solution for long-term storage and availability of digital materials and the software tools included in koLibRI form part of the successful outcome. koLibRI Version 1.0 is
an open-source framework to assist those setting up preservation systems, and supports creation and retrieval of Archival Information Packages. See the kopal Web site for more information. Retrieved 19 September 2007 from: http://kopal.langzeitarchivierung.de/index_koLibRI.php.en

The nestor partner responsible for Work Package 5 (Training & Education) and a range of educational institutions have recently signed a Memorandum of Understanding, agreeing to support instruction in the area of long-term digital archiving and to co-operatively develop curricular modules which are available to all participants for education, training and continuing education. It is intended that the courses and the corresponding ECTS points earned be mutually recognized. The partners are: Göttingen State and University Library; Archivschule Marburg; Fachhochschule Köln; Fachhochschule Potsdam; Hochschule für Technik, Wirtschaft und Kultur Leipzig (FH); Hochschule für Technik und Wirtschaft Chur; Institut für Bibliotheks- und Informationswissenschaft (IBI) an der Humboldt-Universität zu Berlin and the Technische Universität Wien.

1.8 DigitalPreservationEurope (DPE)

DPE is making progress in a number of areas of activity:

- **DPE Research and Industrial Exchange Programme [DPEX].** DPEX aims to encourage innovative practice and cross-institutional synergies by establishing a series of twenty-five exchanges among research institutions. The deadline for submissions for the first exchange was 1 August 2007, with additional rounds scheduled throughout the year. More information and application details, retrieved 19 September 2007, are available via from: http://www.digitalpreservationeurope.eu/exchange/.

- The first **Digital Preservation Challenge** was announced in April. Entrants were asked to submit a solution to one of six real-life digital preservation challenges, including file format identification, emulation or preservation of multimedia art, with the results to be announced during the European Conference on Research and Advanced Technology for Digital Libraries (ECDL). Information on the Challenge, retrieved 19 September 2007, is available from: http://www.digitalpreservationeurope.eu/challenge/

- **Competence Centres: State of the Art Review:** DPE has released a report reviewing the current availability of digital curation and preservation expertise in the EU and beyond. DPE developed a benchmarking model (the 7 ‘C’s) to allow comparison of competence centres in different sectors in the review, consequently recommending a ‘virtuous circle’ approach to curation covering user needs, research, development, service provision, dissemination and practice. The full report, retrieved 19 September 2007, is available at: http://www.digitalpreservationeurope.eu/publications/competence_centre_SoAR_1.pdf

- **DRAMBORA Toolkit:** The DCC and DPE released the *Digital Repository Audit Method Based on Risk Assessment* (DRAMBORA) in March 2007. DRAMBORA characterises digital curation as a risk-management activity, and provides a metric and a methodology for repository self-assessment,
allowing measurement of risks which can be managed and compared. DRAMBORA can be downloaded from the joint DCC/DPE Repository Audit Web site, retrieved 19 September 2007 at: http://www.repositoryaudit.eu/download.

- **DPE Research Roadmap:** The Research Roadmap aids digital preservation research planning by providing analysis on the state-of-the-art in global digital preservation research, computer science and ICT markets, and the needs of digital preservation user communities. DPE has released a draft Research Roadmap identifying ten core domains for preservation research: Restoration, Conservation, Management, Risk, Significant Properties of Digital Objects, Interoperability, Automation, Context, Storage, and Experimentation. For further information please contact Holger Brocks: Holger.Brocks@FernUni-Hagen.de (remove .nospam from address before sending.)

- **DPE Training activities:** DPE seeks to implement a Pan-European framework of training and continuing professional development in the area of digital preservation, and is planning five two-week training events in Lithuania, Austria, Spain, Germany and Italy. The first joint DPE/Planets/nestor training event, ‘Principles of Digital Preservation: a hands-on approach’, will be held in Vilnius, Lithuania, 1-5 October 2007. For the course program and further details, retrieved 19 September 2007, visit: http://www.wepreserve.eu/events/dpe-planets-vilnius-2007

---

### 1.9 CASPAR

In recent months, CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) has published a number of key documents, available from the CASPAR web site.

The *Conceptual Model* provides the fundamental ideas which build on the OAIS Reference Model (ISO 14721), adding the concepts needed for a support infrastructure which is itself able to be preserved. This infrastructure is detailed in the *Overall Component Architecture* which provides the high level components and their interfaces. The testing of these ideas will be based on the collection of *User Requirements and Scenario Specifications*, the first instalment of which is also available, supported by further detailed case studies from the domains of science, cultural heritage and contemporary arts. A ‘State of the Art’ review is also available, providing a new cross-cutting approach based on OAIS to allow the comparison of preservation techniques. This document will be extended throughout the project.

http://www.casparpreserves.eu/publications/deliverables

http://www.casparpreserves.eu/publications/deliverables

*D4101 User Requirements and Scenario Specifications, CASPAR, 21 December 2006. Retrieved 19 September 2007 from:*
David Giaretta describes the CASPAR approach to digital preservation in the second issue of the *International Journal of Digital Curation* (IJDC).


CASPAR partners have also been looking beyond the CASPAR project, working with other European and global partners on a number of initiatives including the production of an ISO standard for audit and certification of digital repositories, and an alliance of major players in the field of permanent access to the records of science. Further information on these initiatives is available via their respective Web sites.


In addition, CASPAR has been one of the leaders of an effort to integrate the EU funded projects. One of the latest developments is the WePreserve collaborative platform, highlighting activities of DigitalPreservationEurope (DPE), Preservation and Long-term Access through NETworked Services (PLANETS), and CASPAR. WePreserve has been designed to provide a common entry point to digital preservation projects and provide common services. More information is available on the WePreserve Web site, retrieved 19 September 2007 from http://www.wepreserve.eu/

---

### 1.10 National Library of the Netherlands (Koninklijke Bibliotheek)

Since 2005 the Koninklijke Bibliotheek and the Nationaal Archief of the Netherlands have been working together to develop a new preservation action based on emulation. In July 2007 the project presented its result: *Dioscuri*, a modular emulator for digital preservation. The emulator is designed to ensure that documents and programs from the past can still be accessed in the future. Being Java-based, *Dioscuri* can run on almost any computer platform, promoting durability, while the component-based approach, which allows individual hardware components to be emulated as modules, means that the emulator remains flexible.

Further information is available from the *Dioscuri* project Web site, retrieved 19 September 2007 from: http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatieproject-en.html

The latest version of the software is also available for download, retrieved 19 September 2007 from: http://dioscuri.sourceforge.net
1.11 US National Digital Information Infrastructure and Preservation Program (NDIIPP).

The Library of Congress, through the NDIIPP, recently announced eight partnerships and awarded $2.15 million USD in funding for the Preserving Creative America projects. The projects will “target preservation issues across a broad range of creative works, including digital photographs, cartoons, motion pictures, sound recordings and even video games” and will involve development of “standardized approaches to content formats and metadata.”

There are eight lead entities involved:

- **Academy of Motion Picture Arts and Sciences (AMPAS)** will lead the Digital Motion Picture Archive Framework Project, which will investigate archival strategies for digital motion pictures and recommend specifications for image data formats across the production chain.

- **American Society of Media Photographers** (with partners PhotoDistrict News and the ASMP Foundation) will seek to expand and refine the existing Universal Photographic Digital Imaging Guidelines and to promote the use of the guidelines within the professional photographer community.

- **ARTstor** (with partners Art on File, Artesia, the Josef and Anni Albers Foundation, and Northwestern University) will develop training and tools to enable photographers to submit archive-ready images and metadata to repositories such as ARTstor.

- **BMS/Chace** (with partners Sony BMG Music, Universal Music Group, EMI Group, Disney Music Group and the Producers and Engineers Wing of the The Recording Academy) will create a standardized approach and software models for gathering and managing metadata for recorded music, allowing effective documentation, archiving and management of "born digital" recordings.

- **Stock Artists Alliance (SAA)** will promote, through online resources and educational seminars, the importance of metadata for long-term usability of digital photographs.

- **Universal Press Syndicate** will use a collection of comic strips and editorial cartoons to model and test the transfer of digital content to the Library of Congress, providing a case study for metadata practices, transfer procedures and digital collection management.

- **UCLA Film & Television Archive** (with partner the Sundance Institute), will support awareness and education on long-term sustainability of digital works within the independent film community.

- **University of Illinois at Urbana-Champaign** (with partners the University of Maryland, Stanford University, Rochester Institute of Technology and Linden Lab) will, through the Preserving Virtual Worlds project, explore methods for preserving digital games, electronic literature and interactive multiplayer games, such as Second Life.
Further information is available in the press release from the NDIIPP Web site:


---

**1.12 FEDORA Commons**

Fedora Commons have announced a $4.9 million USD grant to build organisational and technical frameworks for digital preservation. Fedora Commons is a continuation of the open-source Fedora Project, and this grant will promote the building and use of open-source repositories to support the digital output of online scholars.


---

**2. Specific themes**

**2.1 Digital Preservation - definitions, frameworks & directions**

A recent article by Nancy McGovern in *RLG DigiNews* reflects on a decade in digital preservation and the standards that define today’s preservation practices. It takes a ‘gap analysis’ approach and highlights the need for integrating policies into action, addressing skills shortages, developing suites of tools and securing digital preservation funding.


At the Midwinter ALA meeting it was decided to develop a definition of digital preservation for the benefit of future ALA meetings and policy documents. Short, medium and long definitions and their series of revisions can be found via the blog on the ALA Web site.


The Netherlands National Commission for UNESCO recently published the four keynote presentations and synthesis report from the conference *Preserving the Digital Heritage*, held in The Hague on 4-5 November 2005. The conference, which was part of an action plan to promote and disseminate the UNESCO Charter on the Preservation of the Digital Heritage, focussed on selection of material to be
preserved and the division of responsibility between institutions.


A draft paper recently published by the Research Information Network “sets out a policy framework of five principles, with associated guidelines” to ensure appropriate management of research data, and includes a section on preservation and sustainability, posing questions for consideration by data managers. The current version is a draft for consultation.


A recent report from the Storage Networking Industry Association [SNIA] Task Force, a group established to define best practices and storage standards for digital retention, is a culmination of a requirements survey for potential technology solutions. The study surveyed a range of practitioners to find out the need for long-term retention solutions. Registration is required to access this report.


### 2.2 Copyright

The EU’s High Level Expert Group on Digital Libraries presented a report on copyright issues to the EU Commission. The report sets out some high level principles, recognising both the importance of respecting copyright and related rights and the importance of digital preservation, including the needs of institutions to make copies of materials for preservation purposes and to have legal certainty in their actions.


An FAQ in the latest edition of *RLG DigiNews* seeks to answer questions asked by readers about the difference in repositories such as open access repositories and using a ‘green’ publisher. It explores the role of preservation in the open archive as well as access rights for digital repositories.

2.3 Repositories

Provision of relatively simple, lower cost solutions for digital archival and preservation systems to meet the needs of less well-funded communities is the focus of a recent UNESCO Memory of the World report prepared by Bradley et al., released in June 2007. The report looks at existing software, highlights the gaps and puts forward requirements for the development of open source preservation systems, recommending that UNESCO support the “aggregation and development of an open source archival system, building on, and drawing together existing open source programs.”


An article by Hitchcock et al. in D-Lib Magazine describes three models for provision of preservation services for digital repositories and illustrates how use of the UK National Archives’ PRONOM/DROID service for file format identification across repositories as part of the ROAR (Registry of Open Access Repositories) project produced a shift in thinking towards a model of distributed preservation services.

The evolution of repository models during the inter-related development of the ARROW, DART and ARCHER projects is described by Treloar and Groenewegen in the April 2007 issue of Ariadne.


A policy-driven, novel software environment for “capturing the interactions between distributed producers and an archive” is described in a number of articles by Smorul et al. about the prototype Producer – Archive Workflow Network (PAWN). PAWN provides a common infrastructure for management of the ingest process, assembling OAIS Submission Information Packages (SIPs) using METS as the default packaging method.


Further information on PAWN and other components is available from the ADAPT project Web site, retrieved 19 September 2007 from: http://adaptwiki.umiacs.umd.edu/twiki/bin/view/Main/PAWN

Presentations and reports are available from the recent JISC event, Digital Repositories: Dealing with the Digital Deluge, held at Manchester University on 5-6 June 2007. The ‘repositories and preservation strand’ notes describe presentations on various case-studies, as well as the DRAMBORA toolkit. A brief report on the conference is also provided by Peter Cliff in the July 2007 issue of Ariadne.


DAITSS (Dark Archive in the Sunshine State) is a preservation repository application which implements format migration and normalization as preservation strategies for supported formats, of which there are around ten, including common image and media formats. The system, which can support multiple customers, provides a ‘dark archive’ and has no public user interface or real-time access to content, but could be used as a ‘back end’ to other systems. The system was developed by the Florida Center for Library Automation for use in the Florida Digital Archive, and has recently been released as open source. The software and documentation are freely available from the DAITSS Web site, retrieved 19 September 2007 from: http://daitss.fcla.edu

2.4 Institutional repositories

Considerations for setting up and maintaining institutional repositories are the focus of several recent articles.

Oya Y. Reiger “offers strategies to match specific institutional requirements with repository system features and functionalities” in a recent article in D-Lib Magazine, describing the stages of selecting a suitable repository model for an institution.

Margaret Henty describes ten major challenges for set up and sustainability of institutional repository services in the higher education sector in Australia, including defining collections, roles and responsibilities, service quality and reporting requirements.

Sanjay Kataria discusses the importance of institutional repositories and the challenges in establishing them, with particular reference to higher education institutions in India.
Some of the key issues, choices made and challenges overcome in establishing a digital repository at Loughborough University are described in a recent Program article by Joanna Barwick, which provides a case study of a year in the life of an institutional repository.


An earlier version of this article is also available from the Loughborough University Web site. Retrieved 19 September 2007 from: http://dspace.lboro.ac.uk/dspace/handle/2134/2640 or http://hdl.handle.net/2134/2640


2.5 Electronic Records Management

The Transcontinental Persistent Archives Prototype (TPAP), a project which brings together a number of US institutions including the National Archives and Records Administration [NARA] and the San Diego Supercomputer Center [SDSC], aims to address “key challenges in safeguarding, preserving, and providing access to authentic electronic records”. A brief article by Paul Tooby in D-Lib Magazine explains the project and its aims.


An article by Kenneth Thibodeau in First Monday, presented as part of the WebWise 2007 conference, describes the issues to be addressed and the expectations for the Electronic Record Archives System at NARA, setting out the main requirements and plans for implementation of the system.

The Federal Depository Library Program has accumulated a large number of legacy CD-ROMs, to which provision of access is difficult. This article highlights how the US Government Printing Office is tackling the problem by piloting migration of the material to a more stable server environment.


### 2.6 Web Archiving

Using a grant from the Mellon Foundation, the Internet Archive is conducting a project to capture two billion Web pages from around the world, in an effort to create a global snapshot. Seed URL submissions from memory and cultural institutions were sought, and the crawl was scheduled to begin on 1 June 2007. The Web page for the project, retrieved 19 September 2007, is available at: [http://wa.archive.org/aroundtheworld/](http://wa.archive.org/aroundtheworld/)

A recent article in *First Monday* describes a Web archiving project, the Web-at-Risk, and provides an introduction to the topic, as well as describing the availability of current web harvesting tools. The project is funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP), and its purpose is to build tools that will allow librarians to ‘capture, curate and preserve Web–based government and political information.’ The project partners are the California Digital Library, University of North Texas, and New York University.


The Danish *NetarchiveSuite* is a Web archiving software package developed through the netarchive.dk project. Its main function is to plan, schedule and run Web harvests of parts of the internet and it can be scaled from small thematic harvests to large national domain crawls. It is built around the *Heritrix* Web crawler. Det Kongelige Bibliotek (The Royal Library, Copenhagen) and the Statsbiblioteket (State and University Library, Århus) have released the *NetarchiveSuite* as open source. Download the system from the netarchive Web site, retrieved 19 September 2007 from: [http://netarchive.dk/suite](http://netarchive.dk/suite)

The British Library has recently undertaken performance testing of the International Internet Preservation Consortium (IIPC) Web Archiving Toolset, in anticipation of
Legal Deposit legislation covering online materials being introduced in the future. The aim of the tests was to evaluate the hardware requirements for running the Toolset in a production environment. Pope and Beresford report on the testing in a recent *Ariadne* article.


Papers have been released from the 7th International Web Archiving Workshop (IWAW), held in Vancouver, Canada, on 23 June 2007. Topics include indexing of high volumes of Web data; building interfaces to Web archives; and CRATE – a model for self-describing Web resources, as well as a number of case studies.


### 2.7 Digital Objects

A number of recent articles describe approaches to handling complex digital objects.

Saidis and Delis describe an approach to handling “variations in digital object parts at a higher level of abstraction”, seeking to reduce the requirement for custom content models and coding.

The Open Archives Initiative Object Reuse and Exchange (OAI-ORE) project aims to create extensible methods for describing relationships between objects and parts to promote use and re-use of compound information objects and their components. A discussion of the issues being addressed and possible solutions is contained in a recent white paper by Lagoze and Van de Sompel.

A model for self-describing Web resources is proposed in a paper by Smith and Nelson, presented at the 7th International Web Archiving Workshop. Labelled a ‘CRATE’, the proposed complex object consists of undifferentiated metadata and the resource bytestream.


2.8 Preservation Metadata

The National Library of New Zealand has released its metadata extraction tool as open source. The tool extracts preservation metadata from a range of file formats and outputs it in XML format. The tool is freely available from the Sourceforge project page, retrieved 19 September 2007 from: http://meta-extractor.sourceforge.net

A number of recent events and papers cover the use of preservation metadata with other metadata standards, particularly the Metadata Encoding and Transmission Standard (METS).

A range of existing digital library metadata standards and their use were discussed at the recent ALA conference. Presentations are available from the Library of Congress Web site.


A METS Open Day took place at Göttingen State and University Library in May 2007 and a report of the event is available from the nestor Web site. The event covered a range of METS subjects, including how to build a METS file from scratch. Retrieved 19 September 2007 from: http://nestor.sub.uni-goettingen.de/mets_2007/index.php?show=opd

A metadata model at the Landesarchiv Baden-Württemberg has emerged as part of their digital preservation project. The representation model for the preservation metadata is based on PREMIS. More information is included in the April 2007 *nestor Newsletter*, retrieved 19 September 2007 from: http://nestor.sub.uni-goettingen.de/newsletter/archiv.php?show=11&lang=en

A draft of best practices for use of PREMIS in METS, prepared by Rebecca Guenther, is available for comment from the PREMIS Maintenance Activity Web site, and a paper by Guenther and Xie, presented at the Archiving 2007 conference, describes the use of PREMIS in container formats. A list of tools for creating and manipulating preservation metadata, particularly PREMIS and METS, has also been compiled.


A report by Deborah Woodyard-Robinson, *Implementing the PREMIS Data Dictionary: A Survey of Approaches*, was released in June 2007. Commissioned as part of the PREMIS Maintenance Activity to aid review of the PREMIS Data Dictionary and schemas, the study surveyed sixteen repositories about their interpretation and application of the PREMIS Data Dictionary, in order to help
understand difficulties that implementers have encountered in applying PREMIS semantic units.


The list of changes and corrections to the PREMIS Data Dictionary and schemas already approved and those under discussion are available from the PREMIS Maintenance Activity Web site, retrieved 19 September 2007 from: http://www.loc.gov/standards/premis/changes.html

2.9 Storage Media

The Optical Storage Technology Association (OSTA) and Ecma International have announced the creation of an Optical Disc Archival Test Standard. This is the culmination of research to provide the industry with more reliable archival-grade optical disks. The standard, released for review as ECMA-379, will be transferred to an ISO/IEC 'fast track' process, with the first ISO/IEC version expected in December 2007.


2.10 Tools

A range of software tools designed to help overcome various challenges for digital preservation have been released in recent months. (Most of these tools have already been described elsewhere in this issue.)

- NetarchiveSuite: This Web archiving software package was developed through the Danish netarchive.dk project and has been released as open source. Its main function is to plan, schedule and run Web harvests of parts of the internet and it can be scaled from small thematic harvests to large national domain crawls. It is available from the netarchive Web site,
• **KoLibRI**: The *kopal Library for Retrival and Ingest* (koLibRI) supports creation and retrieval of Archival Information Packages, and was developed as part of the German *kopal* project. The software may be obtained from the *kopal* Web site, retrieved 19 September 2007 from: [http://kopal.langzeitarchivierung.de/index_koLibRI.php.en](http://kopal.langzeitarchivierung.de/index_koLibRI.php.en)

• **DAITSS** (Dark Archive in the Sunshine State) is a preservation repository application released as open source by the Florida Center for Library Automation. The application can be used as a ‘back end’ to other applications. Software and documentation are available from the DAITSS Web site, retrieved 19 September 2007 from: [http://daitss.fcla.edu](http://daitss.fcla.edu)

• **The National Library of New Zealand Metadata Extraction Tool**: This tool extracts preservation metadata from a range of file formats and outputs it in XML format. It has recently been released by the National Library of New Zealand as open source and is freely available from the Sourceforge project page, retrieved 19 September 2007 from: [http://meta-extractor.sourceforge.net](http://meta-extractor.sourceforge.net)

• **Tools for creating and manipulating preservation metadata**, particularly PREMIS and METS, have been listed by the PREMIS Maintenance Activity, retrieved 19 September 2007 from: [http://www.loc.gov/standards/premis/tools.html](http://www.loc.gov/standards/premis/tools.html)

• **Automatic Obsolescence Notification System (AONS) II**: AONS II is a platform-independent tool to aid repository managers in monitoring file format obsolescence risk for formats in their repositories. Drawing on the PANIC prototype by Hunter and Choudhury ([http://metadata.net/panic/](http://metadata.net/panic/), retrieved 19 September 2007) and the 2006 APSR AONS (I) application, a beta release of the application is available for download, retrieved 19 September 2007 from: [http://sourceforge.net/projects/aons](http://sourceforge.net/projects/aons).

• **Dioscuri – the modular emulator**: The Koninklijke Bibliotheek and the Nationaal Archief of the Netherlands have produced a durable and flexible Java-based modular emulator for digital preservation, which has been released as open source. Retrieved 19 September 2007 from: [http://dioscuri.sourceforge.net](http://dioscuri.sourceforge.net)

### 3. Other publications

A number of interesting papers and podcasts are available from the recent Eighth Annual Conference on Libraries and Museums in the Digital World, 'Stewardship in the Digital Age – WebWise 2007', which are included in the July 2007 issue of *First Monday*.


A recent report describes the results of the Sustainability Guidelines for Australian Repositories (SUGAR) Project. The project aimed to produce a tool for creation of navigable guidelines, flexibly drawing together information on preservation of digital materials to support higher education communities without direct access to a digital preservation program.


A short DSpace guide has been published for newcomers to the system. It concentrates on customisation tasks. The guide was a handout at a recent JCDL conference in Vancouver.


Lessons learned regarding the long-term preservation of digital content, as experienced through development of the US National Digital Newspaper Program, are described in an article by Justin Littman in the July/August issue of D-Lib Magazine.


The proceedings of the CIPA/VAST/EG/EuroMed 2006 event are now published. The papers touch on various aspects related to digital preservation and reveal some
case-studies.


A new digital preservation list has been established in the US. Retrieved 19 September 2007 from: http://lists.ala.org/wws/info/digipres

---

4. Events

4.1 Recent events

International Workshop on Database Preservation, 23 March 2007, Edinburgh, Scotland.


7th International Web Archiving Workshop, 23 June 2007, Vancouver, Canada.

Papers and presentations are now available. Retrieved 19 September 2007 from: http://www.iwaw.net/07/


The DPC and British Library held a joint workshop on the JPEG 2000 Image file format on 25 June 2007. Much work has been done with this new file format and is now recognised as an ISO standard. Presentations include the preservation aspects of the standard and how it can be implemented at a local level. The presentations are available from the DPC Web site. Retrieved 19 September 2007 from: http://www.dpconline.org/graphics/events/0706jpeg2000wkshop.html


4.2 Current and Forthcoming events

September


October


November


December


4.3 Training Events

October

Planets, DPE and CASPAR (the three projects co-funded by the European Union under the Sixth Framework Programme) have reached agreement to collaborate in a number of areas, including dissemination and training. More information can be found on the website, including details of the first joint training event which will take place on 1-5 October 2007. More information from the WePreserve Web site. Retrieved 19 September 2007 from: http://www.wepreserve.eu/events/dpe-planets-vilnius-2007/

Links last checked: 16 October 2009

Warning! Web site links tend to have very short lifetimes, as documents are frequently updated or deleted, Web sites are restructured, domains are renamed or moved, etc. The compilers of this bulletin, therefore, cannot guarantee that all of the URLs in this document will successfully resolve to the resources described here. However, in these cases, try searching for the same resource on the PADI gateway (http://www.nla.gov.au/padi/), which will provide updated URLs wherever possible.