



DPC/PADI What's new in digital preservation

Issue no. 10 (January - May 2005)

Compiled by Michael Day ([UKOLN](#), University of Bath), Marian Hanley and Gerard Clifton ([National Library of Australia](#))

27 May 2005

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	2
1.1 US National Science Board	2
1.2 US Library of Congress and the National Digital Information Infrastructure and Preservation Program (NDIIPP)	2
1.3 PREMIS Working Group	3
1.4 DigiCULT	3
1.5 UK Digital Preservation Coalition (DPC)	3
1.6 UK Joint Information Systems Committee (JISC)	4
1.7 International Council for Science (ICSU)	4
2. Specific themes	5
2.1 Peer-to-peer preservation systems	5
2.2 The management of scientific data	6
2.3 Metadata and object packaging	9
2.4 Personal archives	10
2.5 Storage media and computer forensics	10
2.6 The evaluation and cost of digital preservation strategies	11
2.7 Web archiving	12
3. Other publications	13
4. Events	15
4.1 Recent events	15
4.2 Forthcoming events	16

1. News from Organisations and Initiatives

1.1 US National Science Board

At the end of March 2005, the US National Science Board - an independent policy body that provides advice on science and engineering matters for the President and Congress as well as overseeing the National Science Foundation - issued a draft report for public comment entitled "Long-lived digital data collections: enabling research and education in the 21st century". The report specifically tries to address the policy and strategy issues relevant to long-lived digital data collections funded by the NSF, and to make recommendations towards harmonising strategy, policies, processes and budgets for maintaining these collections. The comment period expired at the end of April.

Comments on the draft by the UK Digital Curation Centre are available on the DCC Web site (Retrieved 27 May, 2005 from: <http://www.dcc.ac.uk/research/reports/nsbreport.pdf>)

National Science Board. (2005). *Long-lived digital data collections: enabling research and education in the 21st century. National Science Board draft report NSB-05-40*, 30 March. Retrieved 27 May, 2005 from: http://www.nsf.gov/nsb/meetings/2005/LLDDC_draftreport.pdf

1.2 US Library of Congress and the National Digital Information Infrastructure and Preservation Program (NDIIPP)

At the beginning of May, the Library of Congress National Digital Information Infrastructure and Preservation Program and the National Science Foundation announced the funding of 11 research projects designed to support the long-term management of digital information. The titles of the projects funded under the research programme are the following:

- Multi-Institution testbed for scalable digital archiving (University of California San Diego, Scripps Institute of Oceanography, San Diego Supercomputer Center; Woods Hole Oceanographic Institution)
- Robust technologies for automated ingestion and long-term preservation of digital information (University of Maryland)
- Digital engineering archives (Drexel University)
- Digital preservation lifecycle management: building a demonstration prototype for the preservation of large scale multimedia collections (University of California San Diego, San Diego Supercomputer Center)
- Investigating data provenance in the context of new product design and development (University of Arizona)
- Incentives for data producers to create archive-ready data sets (University of Michigan)
- Shared infrastructure preservation models (Old Dominion University)
- Planning a globally accessible archive of MODIS data (University of Tennessee at Knoxville)
- Preserving video objects and context: a demonstration project (University of North Carolina at Chapel Hill)
- Securely managing the lifetime of versions in digital archives (Johns Hopkins University)

Short project summaries are available in the press release issued by the Library of Congress: *Library of Congress and National Science Foundation announce research awards of \$3 million to advance digital preservation* [press release]. Retrieved 27 May, 2005 from: <http://www.loc.gov/today/pr/2005/05-118.html>

1.3 PREMIS Working Group

In May 2005, OCLC Online Computer Library Center and the Research Libraries Group published the final report of the PREMIS (Preservation Metadata: Implementation Strategies) Working Group. This highly significant report includes the PREMIS Data Model, version 1.0 of the PREMIS Data Dictionary, as well as some examples and implementation guidance. A set of XML Bindings is available separately.

PREMIS Working Group. (2005). *Data dictionary for preservation metadata: final report of the PREMIS Working Group*. Dublin, Ohio: OCLC Online Computer Library Center; Mountain View, Calif.: Research Libraries Group, May. Retrieved 27 May, 2005 from: <http://www.oclc.org/research/projects/pmwg/>

1.4 DigiCULT

At the end of 2004, the DigiCULT consortium produced a report entitled *The future digital heritage space: an expedition report*, edited by Guntram Geser and John Pereira of Salzburg Research. Attractively illustrated with images from the German Maritime Museum and using the analogy of a scientific expedition, this important report explores the future landscape for information and communications technology in cultural heritage applications and is based on the contributions of 62 researchers and professionals to a DigiCULT online consultation forum. The report covers a number of key themes in detail, including: knowledge representation and semantics, the understanding of objects in their appropriate contexts, user interaction and interfaces, three-dimensional objects and virtual reality, distributed information systems, and persistent and perpetual access. The report concludes with a summary and recommendations.

Geser, G., & Pereira, J., eds. (2004). *The future digital heritage space: an expedition report*. DigiCULT thematic issue 7, December (88 pp.). ISBN 3-902448-04-0. Retrieved 27 May, 2005 from: <http://www.digicult.info/pages/themiss.php>

DigiCULT has also published a technology watch report entitled "Core technologies for the cultural and scientific heritage sector." Individual chapters cover open-source software and standards, natural language processing, information retrieval, location-based systems (e.g. using geographical information systems (GIS) or Global Positioning System (GPS) technologies), data visualisation, and the technologies of telepresence, haptics and robotics.

Ross, S., Donnelly, M., Dobрева, M., Abbott, D., McHugh, A., & Rusbridge, A. (2005). *Core technologies for the cultural and scientific heritage sector*. DigiCULT technology watch report 3, January (296 pp.) ISBN 92-894-5277-3. Retrieved 27 May, 2005 from: <http://www.digicult.info/pages/techwatch.php>

1.5 UK Digital Preservation Coalition (DPC)

In February 2005 the Digital Preservation Coalition published a technology watch report on the large-scale storage of digital objects. The report was based on the practical experience in developing a Digital Object Management system for the British Library. The report includes an overview of general issues related to the building of large secure data storage systems, including storage management, emerging technologies and software. A report from a related DPC meeting on large-scale archival storage has also been published on the DPC Web site (retrieved May 27, 2005, from: <http://www.dpconline.org/graphics/events/050422meeting.html>).

Linden, J., Martin, S., Masters, R., & Parker, R. (2005). *The large-scale archival storage of digital objects*. DPC Technology Watch Series Report 04-03, February (20 pp). Retrieved 27 May, 2005 from: <http://www.dpconline.org/docs/dpctw04-03.pdf>

One of the many useful resources maintained by the DPC is the Preservation Management of Digital Materials handbook, first published by the British Library in 2001. Neil Beagrie outlines the handbook's current use in training contexts and looks forward to its integration into a modular training course on digital preservation being developed by the University of London Computer Centre, Cornell University, the British Library and the DPC.

Beagrie, N. (2005). "Digital preservation: best practice and its dissemination." *Ariadne*, 43, April. Retrieved 27 May, 2005 from: <http://www.ariadne.ac.uk/issue43/beagrie/>

1.6 UK Joint Information Systems Committee (JISC)

The previous edition of this bulletin reported on the JISC's Digital Preservation and Asset Management in Institutions programme. Leona Carpenter provides a brief overview of the projects funded under this programme in the April issue of *Ariadne*. One of the other projects in this programme, the *espida* project led by the University of Glasgow, has published a report of its inaugural event, a workshop on the "sustainable preservation of digital assets in a university" held in February 2005 (retrieved 27 May, 2005 from: <http://www.gla.ac.uk/espida/>).

Carpenter, L. (2005). "Supporting digital preservation and asset management in institutions." *Ariadne*, 43, April. Retrieved 27 May, 2005 from: <http://www.ariadne.ac.uk/issue43/carpenter/>

In February 2005, supported by a review commissioned from Heery and Anderson (2005), the JISC issued a call for proposals for funding in the area of digital repositories. The deadline for completed proposals was the 7 April 2005, so decisions on the projects funded will be published shortly.

Heery, R., & Anderson, S. (2005). *Digital repositories review*. JISC, February. Retrieved 27 May, 2005 from: (Link updated 05 Oct 2005) http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf

JISC Digital Repositories programme, retrieved 27 May, 2005 from: http://www.jisc.ac.uk/index.cfm?name=programme_digital_repositories

1.7 International Council for Science (ICSU)

At the end of 2004, the International Council for Science published the report of the Committee on Scientific Planning and Review (CSPR) Assessment Panel on Scientific Data and Information. Most fundamentally, this concluded that the research community should assume responsibility for building a robust data and information infrastructure for the future and produced a number of recommendations to support this objective.

International Council for Science (2004). *Scientific data and information: report of the CSPR Assessment Panel*, December. ISBN 0-930357-60-4. Retrieved 27 May, 2005 from: http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf

2. Specific themes

2.1 Peer-to-peer preservation systems

Some preservation strategies advocate the use of peer-to-peer systems for replicating and preserving the integrity of digital resources. A good example is the LOCKSS (Lots of Copies Keep Stuff Safe) system developed to preserve electronic journals or other Web based information. Maniatis, *et al.* (2005) provide an introduction to the LOCKSS system and present a protocol that can be used to detect and repair damage within peer-to-peer networks. Parno and Roussopoulos (2004) simulate the LOCKSS system's behaviour using dynamic models, proposing and evaluating countermeasures to network subversion. Rosenthal, *et al.* (2005) report on a implementation of format migration for Web content in LOCKSS.

Researchers based at Stanford University describe a similar peer-to-peer information preservation and exchange network (PIPE). Cooper, *et al.* (2005) outline the services required in these networks, specifically to deal with malicious sites that may delete or alter data, or refuse to serve it. In a second article, Cooper and Garcia-Molina (2005) look at 'bid trading' between sites in these networks, defining a range of trading scenarios. A more general introduction to peer-to-peer preservation systems can be found in Bungale, Goodell and Roussopoulos (2005).

References:

Bungale, P. P., Goodell, G., & Roussopoulos, M. (2005). "Conservation vs. consensus in peer-to-peer preservation systems." 4th International Workshop on Peer-to-Peer Systems (IPTPS '05), Ithaca, N.Y., USA, February 24-25, 2005. Retrieved 27 May, 2005 from: http://iptps05.cs.cornell.edu/PDFs/CameraReady_214.pdf

Preprint retrieved 27 May, 2005 from: <http://www.eecs.harvard.edu/~mema/publications/iptps2005-Conservation.pdf>

Cooper, B.F., Bawa, M., Daswani, N., Marti, S., & Garcia-Molina, H. (2005). "Authenticity and availability in PIPE networks." *Future Generation Computer Systems*, 21(3), 391-400. Available (to subscribers) from ScienceDirect: doi:10.1016/j.future.2004.04.017
Preprint retrieved 27 May, 2005 from: <http://www.cc.gatech.edu/~cooperb/pubs/pipe.pdf>

Cooper B.F., & Garcia-Molina, H. (2005). "Peer-to-peer data preservation through storage auctions" *IEEE Transactions on Parallel and Distributed Systems*, 16(3), 246-257. Available (to subscribers) from IEEE Xplore: DOI: 10.1109/TPDS.2005.34
Preprint retrieved 27 May, 2005 from <http://www.cc.gatech.edu/~cooperb/pubs/bidtradingtpds.pdf>

Maniatis, P., Roussopoulos, M., Giuli, T.J., Rosenthal, D.S.H., & Baker, M. (2005). "The LOCKSS peer-to-peer digital preservation system." *ACM Transactions on Computer Systems*, 23(1), 2-50. Available (to subscribers) from the ACM Digital Library, retrieved 27 May, 2005 from: <http://doi.acm.org/10.1145/1047915.1047917>
Preprint retrieved 27 May, 2005 from: <http://berkeley.intel-research.net/maniatis/publications/tocs2004.pdf>

Parno, B., & Roussopoulos, M. (2004). "Defending a P2P digital preservation system." *IEEE Transactions on Dependable and Secure Computing*, 1(4), 209-222. Available (to subscribers) from: IEEE Xplore: DOI: 10.1109/TDSC.2004.39
Preprint retrieved 27 May, 2005 from: <http://www.eecs.harvard.edu/~mema/publications/defendingLOCKSS.pdf>

Rosenthal, D. S. H., Lipkis, T., Robertson, T. S., & Morabito, S. (2005). "Transparent format migration of preserved Web content." *D-Lib Magazine*, 11(1), January 2005.

Retrieved 27 May, 2005 from:
<http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html>

2.2 The management of scientific data

A number of conference and journal papers have reported on research work in the area of the management of scientific data. Some of this work is not directly related to long-term preservation, but they do point to some current directions in scientific data management research.

Jim Gray of Microsoft Research and his various collaborators have produced a series of recent papers looking at petabyte-scale data management and related issues. The first of these looked at the lessons learned in the development of management tools for data-intensive services like the World Wide Telescope (Gray & Szalay, 2004). A more specialised paper given at the 2005 Conference on Innovative Data Systems Research (CIDR) looked at the interfaces between database systems and Grid technologies, noting that databases are currently rarely used in Grid projects to perform analytic tasks (Nieto-Santisteban, *et al.*, 2005). An third paper (Gray, *et al.*, 2005) looked at prospects for the coming decade, including the need for new data-analysis methods, the growing importance of scientific data centres, and the use of metadata for facilitating access and re-use.

Moore, Rajasekar & Wan (2005) of the San Diego Supercomputing Center (SDSC) have produced an important paper looking at the potential for data grids, digital library technologies and persistent archives to combine to produce an integrated way of publishing, sharing and preserving research data.

SDSC have also been collaborating with the University of Maryland and the US National Archives and Records Administration (NARA) on a NSF-funded research project on Persistent Digital Archives. This has set up a distributed pilot archive linked through the SDSC Storage Resource Broker (SRB) software. As part of this project, the University of Maryland's Institute for Advanced Computer Studies has been developing a system for ingest called the Producer - Archive Workflow Network (PAWN), details of which have recently been published (JaJa, McCall & Smorul, 2004; Smorul, *et al.*, 2004). An update on the project focusing on the issues of authenticity, cost and scalability was recently presented at the annual Conference on Mass Storage Systems and Technologies (MSST 2005) hosted by NASA Goddard (Moore, JaJa & Chadduck, 2005).

The 'data deluge' currently underway in many scientific disciplines means that there is an increased interest in the storage and management of data at the petabyte-scale. In the particle physics domain, the international BaBar experiment based at the Stanford Linear Accelerator Center (SLAC) has been generating vast amounts of data since around 1999. A recent conference paper has drawn on experience with SLAC's BaBar data store to investigate some of the wider issues surrounding the management of petabytes of data (Becla & Wang, 2005). Some papers at the 2005 Mass Storage Systems and Technologies conference also dealt with dealing with the problems of petabyte-scale storage; links to all papers and presentation material can be found on the conference Web site (Retrieved 27 May, 2005 from: <http://www.storageconference.org/2005/>).

Knowing the provenance (or lineage) of scientific datasets is essential if they are to be understood or accurately re-used by other scientists. Bose and Frew (2005) have produced a survey of data lineage research and propose a metamodel that can be used for lineage retrieval. Other recent papers on the provenance of scientific data include two technical reports issued by the PASOA project led by the University of Southampton (Miles, *et al.*, 2005; Moreau, *et al.*,

2005). Paskin (2005) also explores the potential application of Digital Object Identifiers (DOIs) to scientific data.

Zeller, Froese & Pauly (2005) highlight the problem of data loss in the marine science domain, noting the importance of historical data for investigating changes in biodiversity over time. They look in detail at the creation of a database from data recorded (on paper) by the Guinean Trawling Survey in the early 1960s, noting that such 'data recovery' was vastly cheaper than the cost of the original survey. They recommend that research institutions and funding agencies should make digital data globally available, e.g. through services like the Global Biodiversity Information Facility (GBIF). However, the authors do not really address the problem of curating the databases resulting from the 'data recovery' processes they advocate.

A recent issue of CODATA's *Data Science Journal* (retrieved 27 May, 2005, from: <http://www.datasciencejournal.org/>) contains a special section on the selection, appraisal and retention of scientific and technical data. This contains the full-text of four papers (Anderson, 2004; Eastwood, 2004; Fusco & van Bemmelen, 2004; Gutman, *et al.*, 2004), first presented at a joint CODATA and ERPANET workshop held in Lisbon in December 2003. A report of the workshop by Esanu, *et al* (2004) was also published in the same issue.

References:

Anderson, W. L. (2004). "Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data." *Data Science Journal*, 3, 191-202. Retrieved 27 May, 2005 from: http://journals.eecs.qub.ac.uk/codata/Journal/contents/3_04/3_04pdfs/DS389.pdf

Becla, J., & Wang, D. L. (2005). "Lessons learned from managing a petabyte." *Proceedings of the 2005 Conference on Innovative Data Systems Research (CIDR)*, Asilomar, Calif, USA, January 4-7, 2005. Retrieved 27 May, 2005 from: <http://www-db.cs.wisc.edu/cidr/cidr2005/papers/P06.pdf>

Bose, R., & Frew, J. (2005). "Lineage retrieval for scientific data processing: a survey." *ACM Computing Surveys*, 37(1), 1-28. Available (to subscribers) from ACM Portal, retrieved 27 May, 2005 from: <http://doi.acm.org/10.1145/1057977.1057978>. Also available (retrieved 27 May, 2005) from: http://homepages.inf.ed.ac.uk/rbose/pubs/bose_2005_ACM_CS.pdf

Eastwood, T. (2004). "Appraising digital records for long-term preservation." *Data Science Journal*, 3, 202-208. Retrieved 27 May, 2005 from: http://journals.eecs.qub.ac.uk/codata/Journal/contents/3_04/3_04pdfs/DS387.pdf

Esanu, J., Davidson, J., Ross, S., & Anderson, W. (2004). "Selection, appraisal, and retention of digital scientific data: highlights of an ERPANET/CODATA workshop." *Data Science Journal*, 3, 226-232. Retrieved 27 May, 2005 from: http://journals.eecs.qub.ac.uk/codata/Journal/contents/3_04/3_04pdfs/DS390.pdf

Fusco, L., & van Bemmelen, J. (2004). "Earth observation archives in digital library and grid infrastructures." *Data Science Journal*, 3, 222-226. Retrieved 27 May, 2005 from: http://journals.eecs.qub.ac.uk/codata/Journal/contents/3_04/3_04pdfs/DS388.pdf

Gray, J., & Szalay, A. S. (2004). "Where the rubber meets the sky: bridging the gap between databases and science." *IEEE Data Engineering Bulletin*, 27(4), 3-11. Retrieved 27 May, 2005 from: http://research.microsoft.com/research/pubs/view.aspx?tr_id=815 Also available (retrieved 27 May, 2005) from: <http://arxiv.org/abs/cs.DB/0502011>

Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A. S., DeWitt, D., & Heber, G. (2005). "Scientific data management in the coming decade." *Technical Report MSR-TR-2005-10*, Redmond, Wa.: Microsoft Research. Retrieved 27 May, 2005 from: http://research.microsoft.com/research/pubs/view.aspx?tr_id=860. Also available (retrieved 27 May, 2005) from: <http://arxiv.org/abs/cs.DB/0502008>

Gutmann, M., Schüer, K., Donakowski, D., & Beedham, H. (2004). "The selection, appraisal, and retention of social science data." *Data Science Journal*, 3, 209-221. Retrieved 27 May, 2005 from: http://journals.eecs.qub.ac.uk/codata/Journal/contents/3_04/3_04pdfs/DS386.pdf

JaJa, J., McCall, F., Smorul, M., Moore, R., & Chadduck, R. (2004). *Digital archiving and long-term preservation: an early experience with Grid and digital library technologies*. THIC Meeting, Upper Marlboro, Md., USA, October 26-27, 2004. Retrieved 27 May, 2005 from: (link updated 05 Oct 2005) <http://www.archives.gov/era/papers/thic-04.html>

JaJa, J., Smorul, M., McCall, F., & Yang, W. (2005). "Scalable, reliable marshalling and organization of distributed large scale data onto enterprise storage environments." *Proceedings of the 22nd IEEE / 13th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2005)*, Monterey, Calif., USA, April 11-14, 2005, Los Alamitos, Calif.: IEEE Computer Society, 197-201. Available (to subscribers) from IEEE Xplore: doi:10.1109/MSST.2005.29 Retrieved 27 May, 2005 from: http://www.storageconference.org/2005/papers/18_jajaj_marshalling.pdf

Miles, S., Groth, P., Branco, M., & Moreau, L. (2005). "The requirements of recording and using provenance in e-science experiments." *University of Southampton Electronics and Computer Science Technical Report*. Retrieved 27 May, 2005 from: <http://eprints.ecs.soton.ac.uk/10269/>

Moore, R. W., JaJa, J. F., & Chadduck, R. (2005). "Mitigating risk of data loss in preservation environments." *Proceedings of the 22nd IEEE / 13th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST 2005)*, Monterey, Calif., USA, April 11-14, 2005, Los Alamitos, Calif.: IEEE Computer Society, 39-48. Available (to subscribers) from IEEE Xplore: doi:10.1109/MSST.2005.20 Retrieved 27 May, 2005 from: http://www.storageconference.org/2005/papers/04_moorer_risk.pdf

Moore, R.W., Rajasekar, A., & Wan, M. (2005). "Data grids, digital libraries, and persistent archives: an integrated approach to sharing, publishing, and archiving data." *Proceedings of the IEEE*, 2005, 93(3), 578-588. Available (to subscribers) from IEEE Xplore: doi:10.1109/JPROC.2004.842761 Preprint retrieved 27 May 2005, from: http://www.sdsc.edu/dice/Pubs/IEEE_Moore.doc

Moreau, L., Chen, L., Groth, P., Ibbotson, J., Luck, M., Miles, S., Rana, O., Tan, V., Willmott, S., & Xu, F. (2005). "Logical architecture strawman for provenance systems." *University of Southampton Electronics and Computer Science Technical Report*. Retrieved 27 May, 2005 from: <http://eprints.ecs.soton.ac.uk/10796/>

Nieto-Santisteban, M. A., Gray, J., Szalay, A. S., Annis, J., Thakar, A. R., & O'Mullane, W. J. (2005). "When database systems meet the grid." *Proceedings of the 2005 Conference on Innovative Data Systems Research (CIDR)*, Asilomar, Calif, USA, January 4-7, 2005. Retrieved 27 May, 2005 from: <http://www-db.cs.wisc.edu/cidr/cidr2005/papers/P13.pdf>

Paskin, N.(2005). "Digital object identifiers for scientific data." *Data Science Journal*, 4, 12-20. Retrieved 27 May, 2005 from: http://journals.eecs.qub.ac.uk/codata/Journal/contents/4_05/4_05pdfs/DS392.pdf

Smorul, M., JaJa, J., Wang, Y., & McCall, F. (2004). "PAWN: Producer - Archive Workflow Network in support of digital preservation." *Technical Report CS-TR-4607 UMIACS-TR-2004-49*. College Park, Md.: University of Maryland. Retrieved 27 May, 2005 from: <http://www.umiacs.umd.edu/research/adapt/papers/UMIACS-TR-2004-49.pdf>

Zeller, D., Froese, R., & Pauly, D. (2005). "On losing and recovering fisheries and marine science data." *Marine Policy*, 29(1), 69-73. Available (to subscribers) from ScienceDirect: doi:10.1016/j.marpol.2004.02.003

2.3 Metadata and object packaging

The publication in May 2005 of the PREMIS Working Group's Data Dictionary for Preservation Metadata is a highly significant step (retrieved 27 May, 2005 from: <http://www.oclc.org/research/projects/pmwg/>), as noted in the section on the PREMIS Working Group, [above](#).

Amongst other publications, Brindley, Muir and Probets (2004) investigate the possibility of using the ONIX metadata schema used by publishers for generating preservation metadata. Brown (2005) provides an update on the PRONOM format registry system developed for the UK National Archives, describing enhancements to the technical registry, unique identifiers for file formats, and development of a tool for automatic file format identification. The US Government Printing Office (GPO) published a report of a *Meeting of Experts on Digital Preservation: Metadata Specifications* on their Web site (retrieved 27 May, 2005 from: <http://www.gpoaccess.gov/about/reports/metadata.html>).

The packaging of complex digital objects is another area of relevance to those interested in digital preservation, with a great deal of current interest in the Metadata Encoding and Transmission Standard (METS). Researchers at the University of Ghent have looked at the MPEG-21 Digital Item Declaration (DID) as a promising new packaging format, making comparisons with METS, the Sharable Content Object Reference Model (SCORM), and the Content Packaging XML Binding of the Instructional Management System (IMS) project (Bekaert, De Kooning, & Van de Walle, 2005).

Regarding specific formats, the JPEG2000 in Libraries and Archives Web site (retrieved 27 May, 2005 from: <http://j2karclib.info/>) was launched in early 2005, and allows users to register and post their own information on best practice, articles, projects, and products that use the standard. The Web site also provides access to a report of a 2004 Symposium on JPEG2000 and other resources. The Web site has a companion mailing list discussing the application of JPEG2000 in libraries and archives.

References:

Bekaert, J., De Kooning, E., & Van de Walle, R. (2005). "Packaging models for the storage and distribution of complex digital objects in archival information systems: a review of MPEG-21 DID principles." *Multimedia Systems*, 10(4), 286-301. Available (to subscribers) from SpringerLink: DOI: 10.1007/s00530-005-0163-x

Brindley, G., Muir, A., & Probets, S. (2004). "Provision of digital preservation metadata: a role for ONIX?" *Program*, 38(4), 240-250.

Brown, A. (2005). "Automating preservation: new developments in the PRONOM service." *RLG DigiNews*, 9(2), 15 April. Retrieved 27 May, 2005 from: http://www.rlg.org/en/page.php?Page_ID=20571#article1

2.4 Personal archives

Interest in the preservation of personal archives is increasing, as is reflected in the development of new initiatives and projects, including the Paradigm project and Ourmedia.

JISC has funded the Paradigm (Personal Archives Accessible in Digital Media) project as part of its Supporting Institutional Digital Preservation and Asset Management programme. It is a two-year exemplar project exploring the issues involved in the long-term preservation of digital private papers. Using the digital papers of contemporary UK politicians, the project will test digital preservation tools and digital repository software, and produce an online "Digital Private Papers Workbook" that will include template policies and procedures (Paradigm project Web site retrieved 27 May, 2005 from: <http://www.paradigm.ac.uk/>).

Launched in 2005, Ourmedia (retrieved 27 May, 2005 from: <http://ourmedia.org/>) is a project that allows third parties to store personal files (e.g., images, audio, video and games) in its managed digital repository. The project's FAQ page says that the site will act as "a central gathering spot where professionals and amateurs come together to share works, offer tips and tutorials, and interact in a combination community space and virtual library that will preserve these works for future generations." Partners in Ourmedia include the Internet Archive, which has undertaken to provide storage space for submitted files for "generations to come."

2.5 Storage media and computer forensics

While they are only one part of a long-term digital preservation strategy, there is still a keen interest in the reliability and longevity of storage media.

The United States government has developed a survey that seeks opinions on the longevity of optical disks. The survey, developed by the Government Information Preservation Working Group (GIPWoG) is working with the National Institute of Standards and Technology (NIST) to establish a "long-term, or archival, standard measurement for recordable CD and DVD media". The deadline for submission is May 31 2005.

Government survey: how long do you want digital storage media to last?
Retrieved 27 May, 2005 from: http://www.dvda.org/html/nist_survey.php

In other publications, Slattery, *et al.* (2004) of the US National Institute of Standards and Technology (NIST) has produced a study of commercially available recordable CD and DVD media that shows wide differences in the stability of the products of different manufacturers. Wirth, *et al.* (2005) have compared the use of different media types (hard disk arrays, magneto-optical disks, magnetic tape) in medical picture archiving and communication systems (PACS), but their recommendations are based on retrieval times rather than media stability.

Forensic computing is another area of growing interest, with new journals like *Digital Investigation* providing a focus for new research and discussion. A general introduction to the subject is the paper by Wang, Cannady & Rosenbluth (2005). Articles frequently touch on media issues: for example, Nikkel (2005) looks at the recovery of evidential data from magnetic tapes. An article by Fernandez, *et al.* (2005) argues that there is a growing need to include courses in forensics in computer science education. BouHaidar (2005) provides a short introduction to Web sites of relevance to forensic computing.

References:

- BouHaidar, R. (2005). "Forensic webwatch: forensic computing." *Journal of Clinical Forensic Medicine*, 12(1), 47-49. Available (to subscribers) from ScienceDirect: doi:10.1016/j.jcfm.2005.01.001
- Fernandez, J. D., Smith, S., Garcia, M., & Kar, D. (2005). "Computer forensics: a critical need in computer science programs." *Journal of Computing Sciences in Colleges*, 20(4), 315-322. Available (to subscribers) from the ACM Portal, Retrieved 27 May, 2005 from: <http://portal.acm.org/citation.cfm?id=1047846.1047894>
- Nikkel, B. J. (2005). "Forensic acquisition and analysis of magnetic tapes." *Digital Investigation*, 2(1), 8-18. Available (to subscribers) from ScienceDirect: doi:10.1016/j.diin.2005.01.007
- Slattery, O., Lu, R., Zheng, J., Byers, F., & Tang, X. (2004) "Stability comparison of recordable optical discs - A study of error rates in harsh conditions." *Journal of Research of the National Institute of Standards and Technology*, 109(5), 517-524. Retrieved 27 May, 2005 from: <http://www.itl.nist.gov/div895/gipwg/StabilityStudy.pdf>
- Wang, Y., Cannady, J., & Rosenbluth, J. (2005). "Foundations of computer forensics: a technology for the fight against computer crime." *Computer Law & Security Report*, 21, 119-127. Available (to subscribers) from ScienceDirect: doi:10.1016/j.clsr.2005.02.007
- Wirth, S., Treitl, M., Lucke, A., Mittermaier, I., Nissen-Meyer, S., Villain, S., Pfeifer, K. -J., & Reiser, M. (2005). "Empfehlungen zur Medienwahl für die Archivierung radiologischer Bilddaten auf der Basis einer Analyse der Zugriffszeiten auf verschiedene PACS-Archivebenen" [Recommendations for the selection of storage media for archiving digital radiological image data based on the comparisons of retrieval times at different PACS archive levels]. *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, 177(2), 250-257.

2.6 The evaluation and cost of digital preservation strategies

Oltmans & Kol (2005) make cost comparisons of migration and emulation strategies. They use the example of the Koninklijke Bibliotheek's e-Depot and predict that an emulation strategy will be more cost effective in institutions which have larger collections.

A key issue for preservation services is the need to make the links between the significant properties of resources, file formats, and preservation strategies. Rauch and Rauber (2004) of Vienna University of Technology present an interesting approach to the selection of preservation strategies based on an adapted version of Utility Analysis, used to gather requirements for the preservation process.

Van der Hoeven, van Diessen & van der Meer (2005) introduce preservation strategies based on emulation and describe the development of an enhanced Universal Virtual Computer (UVC) developed for the Koninklijke Bibliotheek (National Library of the Netherlands) for emulating the JPEG and GIF97a formats. The UVC tool is available on the Alphaworks Web site for testing.

References:

- Oltmans, E., & Kol, N. (2005). "A comparison between migration and emulation in terms of cost." *RLG DigiNews*, 9(2), 15 April. Retrieved 27 May, 2005 from: http://www.rlg.org/en/page.php?Page_ID=20571#article0

Rauch, C., & Rauber, A. (2004). "Preserving digital media: towards a preservation solution evaluation metric." In: Chen, Z., et al. (eds.), *Digital libraries: international collaboration and cross-fertilization: 7th International Conference on Asian Digital Libraries, ICADL 2004*, Shanghai, China, December 13-17, 2004: proceedings, *Lecture Notes in Computer Science*, 3334, Berlin: Springer Verlag, 203-212. Available (to subscribers) from SpringerLink: DOI: 10.1007/b104284 Also available (retrieved 27 May, 2005) from: http://www.ifs.tuwien.ac.at/~andi/publications/pdf/rau_icadl04.pdf

Van der Hoeven, J.R., van Diessen, R.J., van der Meer, K. (2005). "Development of a Universal Virtual Computer (UVC) for long-term preservation of digital objects." *Journal of Information Science*, 31(3), 196-208.

2.7 Web archiving

The first phase of the UK Web Archive - a searchable collection of Web sites selected for their scholarly, cultural and scientific value - was made available in early May 2005 (retrieved 27 May, 2005 from: <http://www.webarchive.org.uk/>)

Reports of the Archiving Web Resources conference held at the National Library of Australia in November 2004 are now available on the NLA Web site (retrieved 27 May, 2005 from: <http://www.nla.gov.au/webarchiving/>) and in *RLG DigiNews* (Phillips, 2005).

Some publications from the Centre for Internet Research at the University of Aarhus deal with what is called the micro archiving of Web sites, that is, the storage of sites by researchers who want to keep them for further study. Brügger (2005) provides an introduction and an outline of methodological issues, plus a step-by-step guide to archiving Web sites. The related paper by Thomasen (2004) tests some of the available software.

Hiragi, et al. (2004) present a Web archiving system that is designed to collect resources in accordance with separately defined policies.

References:

Brügger, N. (2005). *Archiving Websites: general considerations and strategies*. University of Aarhus, Centre for Internet Research (78 pp). ISBN 87-990507-0-6. Retrieved 27 May, 2005 from: <http://cfi.imv.au.dk/eng/pub/webarc>

Hiragi, W., Sakaguchi, T., Sugimoto, S., & Tabata, K. (2004). "A policy-based system for institutional web archiving." In: Chen, Z., et al. (eds.), *Digital libraries: international collaboration and cross-fertilization: 7th International Conference on Asian Digital Libraries, ICADL 2004*, Shanghai, China, December 13-17, 2004: proceedings, *Lecture Notes in Computer Science*, 3334, Berlin: Springer Verlag, 144-154. Available (to subscribers) from SpringerLink: DOI: 10.1007/b104284

Phillips, M. E. (2005). "Archiving Web Resources International Conference: issues for cultural heritage organisations." *RLG DigiNews*, 9(1), 15 February. Retrieved 27 May, 2005 from: http://www.rlg.org/en/page.php?Page_ID=20522#article2

Thomasen, B. H. (2004). "Tests of software and strategies for micro-archiving Websites." University of Aarhus, Centre for Internet Research, December. Retrieved 27 May, 2005 from: http://cfi.imv.au.dk/eng/pub/webarc/test_premises_thomasen.pdf

3. Other publications

Bradley, R. (2005). "Digital authenticity and integrity: digital cultural heritage documents as research resources." *Portal: Libraries and the Academy*, 5(2), pp 165-175.

Presents the results of a North American survey investigating the authenticity and integrity of digital content.

Chen, S. S. (2004). "Digital preservation and workflow process." In: Chen, Z., et al. (eds.), *Digital libraries: international collaboration and cross-fertilization: 7th International Conference on Asian Digital Libraries, ICADL 2004*, Shanghai, China, December 13-17, 2004: proceedings, *Lecture Notes in Computer Science*, 3334, Berlin: Springer Verlag, 61-72. Available (to subscribers) from SpringerLink: DOI: 10.1007/b104284

Discusses three factors (archival stability, organizational process, and technology continuity) that are essential for digital preservation strategies to succeed.

Cloonan, M.V., & Sanett, S. (2005). "The preservation of digital content." *Portal: Libraries and the Academy*, 5(2), 213-237.

Presents work undertaken in the Preservation Task Force of the InterPARES (International Research on Permanent Authentic Records in Electronic Systems) project, incorporating the results from surveys and interviews undertaken between 2001 and 2003.

Danner R.A. (2004). "Issues in the preservation of born-digital scholarly communications in law." *Law Library Journal*, 96(4), 591-604. Retrieved 27 May, 2005 from: <http://eprints.law.duke.edu/archive/00000878/>

A general introduction to the influence that 'born-digital' information has on scholarly communication in the law domain.

Honey, S.L. (2005). "Preservation of electronic scholarly publishing: an analysis of three approaches." *Portal: Libraries and the Academy*, 5(1), 59-75.

Looks at three approaches to the preservation of scholarly communication in digital form: dark archives, moving wall and caching.

Keller, M. A. (2004). "Gold at the end of the digital library rainbow: forecasting the consequences of truly effective digital libraries." In: Chen, Z., et al. (eds.), *Digital libraries: international collaboration and cross-fertilization: 7th International Conference on Asian Digital Libraries, ICADL 2004*, Shanghai, China, December 13-17, 2004: proceedings, *Lecture Notes in Computer Science*, 3334, Berlin: Springer Verlag 84-94. Available (to subscribers) from SpringerLink: DOI: 10.1007/b104284. Also available (retrieved 27 May, 2005) from: http://library.stanford.edu/staff/pubs/Forecasting_Effective_Digital_Libs.pdf

Looks at the requirements of effective digital libraries, including long-term preservation.

Lynch, C. (2004). "Preserving digital documents: choices, approaches, and standards." *Law Library Journal*, 96(4), 609-617. Retrieved 27 May, 2005 from: <http://www.aallnet.org/products/2004-40.pdf>

Surveys technology approaches to preserving born-digital objects and the use of standards to aid preservation.

Marcum, D.B. (2004). "Landscape of digital archiving." *Law Library Journal*, 96(4), 605-608. Retrieved 27 May, 2005 from: <http://www.aallnet.org/products/2004-39.pdf>

An overview projects and initiatives in the area of digital archiving within government, non-profit and university sectors, and internationally, emphasising the importance of collaboration.

Prudlo, M. (2005). "E-archiving: an overview of some repository management software tools." *Ariadne*, 43, April. Retrieved 27 May, 2005 from: <http://www.ariadne.ac.uk/issue43/prudlo/>

Examines three of the more popular repository management systems - LOCKSS, EPrints and DSpace - in terms of functionality, costs, users and required skills.

Van Nispen, A., Kramer, R., & van Horik, R. (2005). "The eXtensible past: the relevance of the XML data format for access to historical datasets and a strategy for digital preservation." *D-Lib Magazine*, 11(2), February. Retrieved 27 May, 2005 from: <http://www.dlib.org/dlib/february05/vannispen/02vannispen.html>

Describes the use of XML for historical datasets in the X-past project undertaken by the Netherlands Historical Data Archive

4. Events

4.1 Recent events

7th International Conference on Asian Digital Libraries (ICADL 2004), Shanghai, China

ICADL 2004 was held in Shanghai, China, on 13 -17 December 2004, under the theme "Digital Libraries: International Collaboration and Cross-Fertilization". Topics included the information society, digital library infrastructure, data semantics, knowledge representation and discovery, digital rights management, web archiving and multimedia libraries.

All peer-reviewed papers accepted for the conference were published by Springer, in the series *Lecture Notes in Computer Science* (Vol. 3334, ISBN 3-540-24030-6, book details retrieved 27 May, 2005, from: <http://www.springeronline.com/sgw/cda/frontpage/0,11855,5-40109-22-37162301-0,00.html>).

22nd IEEE - 13th NASA Goddard Conference on Mass Storage Systems and Technologies (MSST2005), Monterey, California USA

Papers and presentations are available from MSST2005, which was held on 11 - 14 April 2005 at the Monterey Marriott and the Monterey Convention Center, Monterey, California. With "Emerging Trends in Scalable Storage" as the theme, presented papers covered topics such as mass storage system architecture, lifecycle management, file systems and preservation.

Papers are available on the conference Web site, retrieved 27 May, 2005 from: <http://www.storageconference.org/2005/papers-presentations.html>

Conference proceedings are also available from IEEE in print (ISBN 0-7803-9228-0) or CD-ROM (ISBN 0-7803-9229-9), or via online subscription to IEEE Xplore, retrieved 27 May, 2004, from: <http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=30571&isYear=2005>

Selected papers are noted in the section on management of scientific data, [above](#).

DPC Meeting on the Large-scale Archival Storage of Digital Objects, York, United Kingdom

A report of the DPC meeting held at BioCentre, York Science Park, York, on 22 April, 2005 is available from the Digital Preservation Coalition Web site, retrieved 27 May, 2005 from: <http://www.dpconline.org/graphics/events/050422meeting.html>

Society for Imaging Science and Technology (IS&T) Archiving Conference 2005, Washington, D.C., USA

Proceedings are available for IS&T's 2005 Archiving Conference, which was held in Washington, D.C., on 26 - 29 April, 2005. Papers presented at the conference covered topics such as digital repositories and lifecycles, formats and metadata for digital archiving, and strategies and tools for preservation, and included several case studies.

Individual full-text papers are available for purchase in PDF, or the entire proceedings may be purchased in hardcopy (ISBN 0-89208-255-0). Abstracts are available online, retrieved 27 May, 2005, from: <http://www.imaging.org/store/physpub.cfm?seriesid=28&pubid=692>

14th International World Wide Web Conference (WWW 2005), Chiba City, Japan

Full papers are available from the 14th International World Wide Web Conference, which was held on 10 - 14 May, 2005, in Chiba City, Japan. Retrieved 27 May, 2005, from: <http://www2005.org/cdrom/contents.htm>

4.2 Forthcoming events

2005

June

DELOS summer school on Digital Preservation in Digital Libraries, 5-11 June 2005, INRIA, Sophia Antipolis, France.
Retrieved 27 May, 2005 from: <http://www.dpc.delos.info/>

Joint Conference on Digital Libraries 2005 : Digital Libraries Cyberinfrastructure for Research and Education, 7 - 11 June 2005 , Denver, Colorado, USA.
Retrieved 27 May, 2005 from: <http://www.jcdl2005.org/>

ELPUB 2005 : From Author to Reader : Challenges for the Digital Content Chain, 8 - 10 June 2005, Leuven-Heverlee, Belgium.
Retrieved 27 May, 2005 from : <http://www.elpub.net/>

Digital Curation Centre (DCC) Workshop on Persistent Identifiers, 30 June - 1 July 2005 , Wolfson Medical Building, University of Glasgow, Glasgow, Scotland, United Kingdom.
Retrieved 27 May, 2005 from: <http://www.dcc.ac.uk/piworkshop.html>

July

Digital Curation Centre (DCC) Workshop on Long-term Curation within Digital Repositories, 6 July 2005, University of Cambridge, England, United Kingdom.
Retrieved 27 May, 2005 from: <http://www.dcc.ac.uk/drworkshop.html>

2nd DSpace Federation User Group Meeting, 7 - 8 July 2005, University of Cambridge, England, United Kingdom.
Retrieved 27 May, 2005 from: <http://www.lib.cam.ac.uk/dspace/usergroup2005/>

Digital Preservation Management : Implementing Short Term Strategies for Long Term Problems, 17 - 22 July 2005 , Ithaca, NY, USA.
Retrieved 27 May, 2005 from: <http://www.library.cornell.edu/iris/dpworkshop/>

Joint Digital Curation Centre (DCC) and Digital Preservation Coalition (DPC) Workshop on Digital Curation Cost Models, 26 July 2005, British Library, London, United Kingdom.
Retrieved 27 May, 2005 from: <http://www.dcc.ac.uk/cmworkshop.html>

August

Digital Libraries à la Carte : Choices for the Future, 22 - 26 August 2005 , Tilburg University, Tilburg, Netherlands.
Retrieved 27 May, 2005 from: <http://www.ticer.nl/05carte/>

September

DC-2005: International Conference on Dublin Core and Metadata Applications 2005, 12 - 15 September 2005, Madrid, Spain.
Retrieved 27 May, 2005 from: <http://dc2005.uc3m.es/>

ECDL 2005: 9th European Conference on Research and Advanced Technology for Digital Libraries, 18 - 25 September 2005, Vienna, Austria.
Retrieved 27 May, 2005 from: <http://www.ecdl2005.org/>

IWAW '05: 5th International Web Archiving Workshop and Digital Preservation, 22 - 23 September 2005, Vienna, Austria.
Retrieved 27 May, 2005 from: <http://www.iwaw.net/05/>

ETD 2005 : 8th International Symposium on Electronic Theses and Dissertations, 27 - 30 September 2005, Sydney, Australia.
Retrieved 27 May, 2005 from: <http://adt.caul.edu.au/etd2005/>

Refresh : First International Conference on the Histories of Media Art, Science and Technology, 28 September - 3 October 2005, Banff, Canada.
Retrieved 27 May, 2005 from: <http://www.mediaarthistory.org/>

DCC (Digital Curation Centre) Conference 2005, 29 - 30 September 2005, Hilton Bath City, Bath, United Kingdom.
Retrieved 27 May, 2005 from: <http://www.ukoln.ac.uk/events/dcc-2005/>

October

Joint Digital Curation Centre (DCC) and ERPANET workshop on Long-term Curation of Medical Databases, 13-14 October 2005, Gulbenkian Institute, Lisbon, Portugal.
Retrieved 27 May, 2005 from: <http://www.dcc.ac.uk/mdbworkshop.html>

Digital Preservation Management : Implementing Short Term Strategies for Long Term Problems, 31 October - 4 November 2005, Ithaca, NY, USA.
Retrieved 27 May, 2005 from: <http://www.library.cornell.edu/iris/dpworkshop/>

November

Sommet Mondial sur la Société de l'Information 2005 : World Summit on the Information Society Conference 2005, 16 - 18 November 2005, Tunis, Tunisia.
Retrieved 27 May, 2005 from: <http://www.smsitunis2005.org/plateforme/home.htm>

Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data, 21 - 23 November 2005, Royal Society, Edinburgh, Scotland, United Kingdom.
Retrieved 27 May, 2005 from: <http://www.ukoln.ac.uk/events/pv-2005/>

December

8th International Conference on Asian Digital Libraries : ICADL 2005, 12 - 15 December 2005, Imperial Queen's Park Hotel, Bangkok, Thailand.
Retrieved 27 May, 2005 from: <http://www.icadl2005.ait.ac.th/>

A comprehensive and frequently updated list of forthcoming events is available from the PADI Web site: <http://www.nla.gov.au/padi/format/event.html>

Problem links last disabled or updated: 30 September 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.