Information Services & Use 25 (2005) 47–57
IOS Press

Preservation of and permanent access to electronic information resources: A system perspective

Gail Hodge
Information International Associates, Inc., 312 Walnut Place, Havertown, PA 19083, USA
E-mail: ghodge@iiaweb.com

Abstract. The rapid growth in the creation and dissemination of electronic information has emphasized the digital environment's speed and ease of dissemination often with little regard for its long-term preservation and access. But, electronic information is fragile in ways that traditional paper-based information is not. Many projects, worldwide, have contributed to the growing collection of best practices and standards in areas such as metadata creation and format standards. These best practices are increasingly reflected in system software that can be tailored to local needs. However, these systems are still in development and for local implementations more than one approach may be needed. These systems provide an indication of the trends and issues remaining in the area of digital preservation and permanent access to electronic information resources.

1. Introduction

Since early in the investigation of digital preservation, institutions concerned about preservation and interested in performing this function have been awaiting "off the shelf" systems or services that could be installed with limited resources but variant levels of flexibility to meet local needs. These systems are beginning to become available from a variety of organizations. Several of the highlighted systems have or are developing "turn-key" or generalized systems that can be implemented by others. These are available both commercially and as open source software. However, these systems need to be evaluated against a total framework to determine their value for local implementations, particularly those with limited resources.

2. A framework for archiving and preservation

Before addressing the systems that are available for archiving, it is valuable to discuss archiving and preservation requirements within a framework. The framework used in this paper is based on the Open Archival Information System Reference Model [5], which provides high level concepts and a consistent terminology for discussing preservation. The OAIS RM is used by most major preservation projects including those in Australia, the United Kingdom, the Netherlands, and the United States. Many organizations have found this framework to be especially valuable when discussing requirements with system developers or vendors.

2.1. Production and creation of electronic information

Preservation and permanent access begin outside the purview of the archive with the producer or the creator of the electronic resource. Information that is born digital may be lost if the producer is unaware of the importance of preservation, and practices used when electronic information is produced will impact the ease with which the information can be digitally archived and preserved. The archiving and preservation process is more efficient when attention is paid to issues of consistency, format, standardization and metadata description before the material is considered for archiving. Limiting the variability of the incoming material is easier for a small institution or a single company to enforce than for a national archive or library, where a variety of formats must be ingested, managed and preserved.

In the case of more formally published materials, such as electronic journals, efforts are underway to determine standards that will facilitate archiving, long-term preservation and permanent access. Such standardization is considered key to efficient archiving and preservation of electronic journals by third-party archives. The DTD developed by PubMed Central for deposit of biomedical journals is being considered as a generalizable model for all journals. The Archiving and Interchange DTD Suite is based on an analysis of all the major DTDs that were being used for journal literature, regardless of the discipline. The suite is a set of XML building blocks or modules from which any number of DTDs can be created for a variety of purposes including archiving. Using the Suite, NLM created a Journal Archiving and Interchange DTD as the foundation for the PubMed Central archive. In addition, a more restrictive Journal Publishing DTD has been released which can be used by a journal to mark up its content in XML for submission to PubMed Central. Several publishers and projects, such as JSTOR, the Public Library of Science, High Wire Press and CSIRO, are analyzing or planning to use the Journal Publishing DTD [2].

In addition to adhering to specific formats and standards, creators can create metadata. The best practice is for metadata to be created prior to incorporation into the archive, i.e., at the producer stage, where an expert can support the description of the technical content. With recent incorporation of XML and other architectures into software applications, such as MS Word, and PDF, the creation of metadata by creators should become easier and more automatic. In the case of less formally published material such as web sites, the creator may also be involved in providing metadata to support the long-term assessment of the information. The Preservation Office at the National Library of Medicine has implemented a "permanence rating system" based on three factors: integrity, persistent location and constancy of content [3]. These factors are combined into a scheme that can be applied to any electronic resource to manage ongoing preservation activities and alert users about a resource's long-term stability.

2.2. Ingest: Acquisition and collection development

The first function to be performed by the archive itself is acquisition, or ingest. This is the stage at which the created object is "incorporated" physically or virtually into the archive. The acquisition of electronic information for archiving involves the development of collection policies and gathering procedures, and these policies and procedures should be considered in tandem with the development of archiving system requirements.

The collection policy answers questions such as what should be archived, what is the extent of a digital object, how often should the content of an archived site be recaptured, and what are the impacts on version control and identification? The most comprehensive analysis of such guidelines is in the *Digital Preservation Handbook*, which is based on the combined lessons learned of several major projects [1].

A key component of a collection policy is the gathering procedure. There are two general ways in which an archive can acquire material. In the submission approach, the producer of the object sends the object to the archive based on a previously defined relationship and submission guidelines. Attention to standardization and limitations on the number of formats will have a significant impact on the ease with which submissions can be processed. In the second approach, the archive may or may not have a formal relationship with the creator or the producer. In this gathering approach, the archive is proactively acquiring the objects through hand-selection, harvesting programs, or a combination of these techniques. The selection of a harvesting approach involves the determination of parameters for the program. These parameters act as filters to align the results of the harvesting process with the scope of the archive. Parameters may include format types (.doc, .html, .mdb, etc.), date ranges, subject matter, and domain names. In the case of the National Library of Australia, sites are identified, reviewed, hand-selected, and monitored for their persistence before being captured for the archive. In contrast, the Royal Library of Sweden runs a robot to capture sites for its Kulturaw3 project [25] that automatically captures sites from the .se country domain and from foreign sites with material about Sweden, such as travel information or translations of Swedish literature. The NASA Goddard Space Flight Center Library uses a hybrid system that involves the hand-selection of root URLs from which pages are automatically selected based on additional parameters [26].

2.3. Data management: Metadata for preservation

Archiving and preservation require special metadata elements to track the lineage of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to reproduce it on future technologies. In 2001–2002, the Preservation Metadata Working Group developed a draft set of over 20 elements and numerous sub-elements for metadata preservation in the framework of the OAIS Reference Model [19]. In order to gain consensus on this set and to provide operational and implementation guidance, a follow on group, PREMIS, the PREservation Metadata: Implementation Strategies working group was formed [18]. The draft element set for preservation metadata and the results of the implementation survey were published in 2004–2005. The plan is to provide the preservation metadata set for testing and prototype implementations before moving the results into a standards process.

2.4. Archival storage: Formats for preservation

A major issue for the archiving community is which format(s) should be used for archival storage. Should the electronic resource be transformed into a format more conducive to archiving? Should the organization create a dark archive of archival copies in one format and a light archive for dissemination, which might be in a different format? Should the goal be complete replication of the electronic resource or should preservation provide a copy that is "just good enough"?

Of course, the answers to these questions differ by resource type, and there is little standardization at this point. Most electronic journals, reference books, or reports use TIFF image files, PDF, or HTML. TIFF is the most prevalent for those organizations that are involved with conversion of paper issues of journals. For purely electronic documents, Adobe's PDF (Portable Document Format) is the most prevalent format. PDF provides a replica of the Postscript format of the document, but relies upon proprietary encoding technologies. While PDF is increasingly accepted, concerns remain for long-term preservation and it may not be accepted as a legal depository format, because it is a proprietary format [14]. Therefore,

Adobe, the Association for Information and Image Management (AIIM) and several other organizations have developed a draft standard for archival PDF, called PDF-A. This provides a file specification for a minimal set of PDF features and functions that will continue to be migrated from one version of PDF to another. The draft is currently in the ISO process.

2.5. Preservation planning: Migration and emulation

Preservation planning is the bridge between the decisions made about archival storage of the bits and bytes and issues of future access and user needs. There is no common agreement on the definition of long-term preservation, but some have defined it as being long enough to be concerned about changes in technology and changes in the user community. This may be as short as 2–10 years.

Two strategies for preservation are migration and emulation. Migration means copying the object to be archived and moving it to newer hardware and software as the technology changes. Migration is, of course, a more viable option if the organization is dealing with well-established commercial software such as Oracle or Microsoft Word. However, even in these cases migration is not guaranteed to work for all data types, and it becomes particularly unreliable if the information product has used sophisticated software features. Unfortunately, this level of standardization and ease of migration is not as readily available among technologies used in fields of study where specialized systems and instruments are used.

Emulation, a strategy that replicates the behavior of old hardware and software on new hardware and software, is being considered as an alternative to migration. There are several types of emulation. Encapsulation would store information about the behavior of the hardware/software with the object. For example, a MS Word 2000 document would be labeled as such and then metadata information would be stored with the object to indicate how to reconstruct the document at the engineering – bits and bytes – level. An alternative to encapsulating the behavior with every instance is to create an emulation registry that uniquely identifies the hardware and software environments and provides information on how to recreate the environment. Each instance would point to the registry [23,24]. Taking emulation a step further is the idea of creating a virtual machine – a new machine that based on the information in the registry could replicate the behavior of the hardware/software of the past [15].

While the best practice for the foreseeable future continues to be migration, machine emulation has been tested with some success by the CAMiLEON Project, a joint project between the University of Michigan and the University of Leeds. However, Granger concludes that a variety of preservation strategies and technologies should be available. Some simple objects may benefit from migration, while others that are more complex may require emulation ([9]; see also [11]).

2.6. Access: Current and future

The life cycle functions discussed so far are performed for the purpose of ensuring continuous access to the material in the archive. Successful practices must consider changes to access mechanisms, as well as rights management and security requirements over the long term.

The way in which access is viewed depends on the purpose of the archive, the audiences it will serve and the anticipated needs of those audiences over the long term. For example, national and institutional archives must be concerned with the ability to provide long-term access to the electronic information in a way that virtually replicates the look and behavior of the object today. This is a requirement because of the legal functions served by these archives of record.

Other organizations are interested in how they might actually improve access to current information in the future. A major reason for storing the information related to the U.S. National Library of Medicine's Profiles of Science materials in TIFF and other standardized forms, such as tagged ASCII, is so that the information can be re-purposed or enhanced. Even in its development stage, the project was able to improve the quality of the video clips by converting them to High Definition Video. The belief is that there will always be newer and better technologies, and a goal of the archive is to be able to take advantage of these advances in the future.

One of the most difficult access issues for digital archiving involves rights management. What rights does the archive have? What rights do various user groups have? What rights has the owner retained? Of specific importance for the development of system requirements are the questions: How will the access mechanism interact with the archive's metadata to ensure that these rights are managed properly and how will access rights be updated as the copyright status or security level of the material changes? Numerous groups, including the IEEE, ContentGuard, and MPEG, are developing digital rights management standards including expression languages to support interoperability in e-commerce transactions.

3. Systems development

Given the framework described above, what systems are available and what functions do they perform? This section identifies several highly visible systems. It is not intended to be comprehensive but to give a sense of the state of the art and practice with regard to the integration of requirements into operational systems.

3.1. Digital Information Archive System

The Digital Information Archive System (DIAS) is a commercially available system, originally developed to handle the electronic deposit of electronic documents and multimedia files for the Koninklijke Bibliotheek (KB) (the National Library of the Netherlands) [28]. It is based on the results of the various NEDLIB Projects led by the KB over the last several years. In the initial implementation of the DIAS system, IBM addressed the initial ingest, transformation, storage and metadata creation. The DIAS system was implemented as KB's Deposit of Netherlands Electronic Publications (DNEP) system in December 2002, making it the first system of its kind. KB's initial implementation was for e-journal publishers to deposit e-journals, but it is being extended to include other e-materials such as e-books. KB has entered into agreements with several publishers including Elsevier [21] and Kluwer to archive electronic journals.

In 2003 the KB started a joint project with IBM to develop the preservation subsystem of DIAS, called Preservation Manager [20]. The work began with a series of studies around key preservation issues such as authenticity, media migration management, archiving of web publications, and a proof of concept of the Universal Virtual Computer. This subsystem will consist of a preservation manager, a preservation processor and tool(s) for permanent access. The Preservation Manager will manage and control the long-term durability of the digital objects using technical metadata. This is considered to be an essential part of the DIAS solution, since technical metadata will allow a future hardware environment to take the software bit stream and the content bit stream and provide access to the content. The problem that remains to be addressed is the obsolescence of the hardware of the rendering environment. Two major approaches are emulation and the use of a basic virtual computer. The aim is to have the turnkey system able to be generalized to other libraries and archives. Therefore, the system must be independent of the

either of these preservation strategies. DIAS' research is being extended through a new project called KOPAL, which began in October 2004 with the German national library, Die Deutsche Bibliothek [13].

3.2. OCLC Digital Archive

As an outgrowth of the preservation services that OCLC has provided to its member libraries for many years, OCLC has developed the OCLC Digital Archive [17]. It provides long-term access, storage and preservation for digital materials or "objects". The system is based on the OAIS. Records can also be ingested in batch. Currently the OCLC Digital Archive can ingest text and still images in formats such as PDF, HTML, TEXT, JPEG, BMP, GIF and TIFF. The goal is to accept more input formats in the future. This system is connected to OCLC's Connexion cataloging system, and the cataloger begins by creating a WorldCat record for the object, followed by a record that includes the preservation metadata based on an earlier version of the PREMIS metadata effort. These two records are linked. In principle, the OCLC Digital Archive follows the Metadata Encoding and Transmission Standard (METS) structure, providing descriptive, administrative, technical and structural metadata.

The system also includes an Administration Module, which allows the user to modify existing records. The Administrator can also set privileges for a variety of functions so that various pieces of the metadata creation, ingest and dissemination processes can be assigned to different people with proper security. The Administration Module also allows the administrator to create collections and user groups for specific end-user access to the metadata and the content. Virus and fixity checks are run and results are reported through both the Administration and the cataloging (Connexion) modules.

3.3. PANDORA Digital Archiving System (PANDAS)

The PANDAS (PANDORA Digital Archiving System) has been operational since June 2001 [16]. The second version was installed in August 2002. Prior to the development of its own system, NLA tried to buy an archiving management system. From the response to the Request for Information, it became apparent that there was no affordable system on the market that met the requirements and so NLA decided to build the system in-house. PANDAS enabled PANDORA (http://pandora.nla.gov/au/index.html) to increase the efficiency of capturing and maintaining the archived Australian online publications and therefore, PANDORA's productivity. It also provides PANDORA's partners, primarily the state libraries, with more effective Web-based software for contributing to PANDORA.

The NLA has received a number of requests for access to the PANDAS software, since the current software options to support the creation and management of digital archives are limited. UKOLN recommended use of PANDAS for pilot web archiving projects it proposed for both Wellcome Trust and JISC [6]. In response, PANDORA will soon make available an evaluation module, which will allow interested parties to have trial access to PANDAS.

3.4. Lots of Copies Keep Stuff Safe (LOCKSS)

LOCKSS (Lots of Copies Keep Stuff Safe) is an automated, decentralized preservation system developed by Stanford University to protect libraries against loss of access to digital materials (http://lockss.stanford.edu/index.html). LOCKSS development is supported by the National Science Foundation, Sun Microsystems, and the Mellon Foundation. LOCKSS software, which is free and open-source, is designed to run as an "Internet appliance" on inexpensive hardware and to require minimal

technical administration. LOCKSS has been operational at Stanford for five years and the production version of the software was released in April 2004.

LOCKSS creates low-cost, persistent digital "caches" of authoritative versions of http-delivered e-journal content at institutions that subscribe to that content. LOCKSS uses the caching technology of the web to collect pages of journals as they are published, but unlike normal caches, the cached pages are never flushed. The LOCKSS server runs an enhanced web cache that collects new issues of the e-journal and continually compares its contents with other caches via a peer-to-peer polling system. If damage or corruption is detected in an institution's cache it can be repaired from the publisher or from another cache. LOCKSS safeguards the institution's access to the content while enforcing the publisher's access control systems and, the LOCKSS model generally does not harm the publisher's business model since it is based on the original subscription to the e-journal.

LOCKSS is moving toward becoming a self-sustaining alliance. "The LOCKSS Alliance will provide a small core of central support for technology, collections, and community services. In addition to a range of specific services, the Alliance will transfer knowledge, skills and responsibility for the LOCKSS Program from Stanford University" [10].

3.5. DSpace Institutional Repository System

The DSpace Institutional Digital Repository System began as a joint project of the MIT Libraries and Hewlett-Packard Co. [29]. The architecture for the system is based on a number of preceding projects including those at Cornell, CERN, OCLC, LC and OAIS. DSpace 1.1 was released in November 2003 via an open source license (available from SourceForge). DSpace 1.2.2 was released in May 2005 (www.dspace.org).

As an institutional repository system, the current emphasis of DSpace is on ingesting material. It is based on the submission approach to gathering in which people within the institution who have been given the proper authorization can submit both their manuscripts and metadata. The submission is reviewed before full incorporation into the archive. The latest version of the software gives the flexibility of customizing the submission forms.

Dublin Core is the basis for DSpace's descriptive metadata. This can be modified, but requires local programming capabilities. The support for preservation metadata is limited at this point, focusing on administrative control elements. DSpace does support the incorporation of detailed metadata elements for specific versions of software applications; this level of detail is critical to the successful migration and rendering of objects in the future.

The MIT Libraries' implementation of DSpace defines various levels of support for different input formats. For example, "Supported" means that the format is recognized and the institution is confident that it can make the format useable in the future through whatever technique is desirable (emulation, migration, etc.). Note that there is no attempt to dictate the preservation method. "Known" means that the format is recognized and the institution will preserve the bitstream as-is, without a complete guarantee that it will be able to render the object completely in the long-term future.

In addition to these components of DSpace that are specifically preservation-oriented, the DSpace suite includes search and browse capabilities and support for the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This allows DSpace sites to harvest metadata from other sources and to offer services based on the metadata that is harvested. The OAI-PMH will be critical to the plans for developing a federation of DSpace archives.

3.6. Fedora™ – Flexible Extensible Digital Object Repository Architecture

The University of Virginia Library has teamed with Cornell University's Digital Library Group to develop Fedora, an open-source digital repository architecture on which a variety of digital library implementations can be based [27]. Fedora 1.0 was released as open source software (Mozilla Public License) in May 2003. Release 2.0 was made available in January 2005 [30]. The first phase production repository was launched in 2004, but the full functionality described in the original design proposal will not be completed until 2005.

Because Fedora is an architecture around which local repository services can be built, it requires more local programming support to implement. However, this architecture also provides more flexibility for local implementers. The most recent release provides support for XML and web services, which allow easier integration and interoperability with other applications [8].

The largest implementation of Fedora is at the University of Virginia Library's Central Digital Repository. A 2001 Mellon Foundation grant resulted in joint development of a production-quality system by Cornell and the University of Virginia. The system currently includes XML objects, text (full text and page images of e-books) and images in multiple resolution [22]. A number of other institutions and organizations are using or evaluating Fedora, including The British Library, the National Library of Portugal and the Thailand Office of Defense Resources.

3.7. Portable PubMed Central

PubMed Central (http://www.pubmedcentral.gov/about/intro.html) is the National Library of Medicine's digital archive for electronic journals in the biomedical and life sciences. In an effort to decentralize the PubMed Central archive and improve the archival capacity of other institutions, NLM developed Portable PubMed Central (pPMC), a suite of software applications to ingest, manage and provide access to electronic journals.

pPMC involved a collaboration with Microsoft, with Microsoft doing the first-level support. The software is about to undergo beta testing of installation at several sites including the Wellcome Trust in the UK and the Italian National Research Council (CNR).

Initially pPMC is a mirror for PMC. The content, table of contents browsing, rendering and retrieval is performed at the mirror site. However, the searching will initially take place at the central PMC site. In the next stage, the pPMC will be enhanced to allow for independent input and accessioning of content by the pPMC site. NLM hopes that capacity will be developed by other institutions and eventually there will be a federation of pPMC archives.

4. Trends and issues

The trend in archiving and preservation has moved from theoretical discussions to pragmatic projects and operational systems. There are more initiatives focused on the realistic details of metadata, selection criteria, technologies and systems for archiving. There are several systems, commercial, not-for-profit, and open source, that are being developed with the intent of providing long-term archiving, preservation and permanent access.

However, to-date all the systems have their limitations. No system addresses all of the functions of the OAIS RM. IBM's DIAS system is the most advanced of those systems that are available commercially or via open-source for local implementation. The preservation component is being added and tested

through additional implementations of DIAS at national libraries. The implementations of DIAS have focused primarily on e-journals and e-books, though DIAS can reportedly accommodate a wider variety of digital objects. The OCLC Digital Archive addresses most of the OAIS RM functions, but it exists in a third-party environment, with close ties to the Dublin Core metadata scheme and a limited set of formats. The institutional repository environments of D-Space and Fedora are able to handle a variety of file formats, but they have not yet tackled many of the preservation issues. While LOCKSS is moving to become a more comprehensive environment, it has historically focused on addressing the specific problem of loss and corruption of digital content. Portable PubMed Central is still in beta test and will primarily address e-journals, though it will be based on a software suite that has been in production for several years.

While some projects are developing operational systems, others continue to work in the background to achieve consensus on standards among/between projects. Key to this more pragmatic approach is the availability of draft standards such as PDF-A and the PREMIS metadata for digital preservation, which will allow all the systems to move ahead with community consensus in key areas required for interoperability. The National Science Foundation (NSF) and the Library of Congress have awarded research grants in problematic areas such as extremely large data sets and long-term access to complex multimedia objects. The International Internet Preservation Consortium's Deep Web Working Group is investigating the capture of the dynamic web. A recent study shows that an increasing number of countries, including Canada, Denmark, New Zealand, Norway, South Africa and the United Kingdom have addressed the thorny issue of intellectual property rights by enacting legislation or putting a legislative process in place that covers some form of digital publications [10].

Despite these positive trends, key non-technical issues remain. The cost of archiving and the lack of established business models that will sustain long-term preservation may prove to be significant stumbling blocks in the advancement of the cause of preservation. However, even these issues are being addressed in a pragmatic fashion. OCLC, Stanford University Libraries/HighWire Press, JSTOR, and major publishers such as Elsevier are actively dealing with questions of cost and how and who will pay for the archiving. The development of value-added services that can help to subsidize basic archiving and preservation activities is being considered.

A review of the highly visible operational and near-operational systems shows the beginning of a body of best practice for digital archiving that will be embedded in operational systems. The early adopters in the development of systems for archiving and preserving digital content are providing lessons that can be adopted by others in the stakeholder communities. Most encouraging is the speed with which these systems are being enhanced with new releases and functionality that provides increased flexibility for local implementations. While the more difficult issues related to preservation continue to be worked on in the research labs and budget meetings, systems are increasingly available to support the archiving of information before it is lost completely. Through the collaborative efforts of the various stakeholder groups – creators, librarians, archivists, funding sources, and publishers – a new tradition of stewardship will be developed to ensure the preservation and continued access to our intellectual heritage.

References

[1] N. Beagrie and M. Jones, *Preservation Management of Digital Materials: A Handbook*, The British Library, London, 2001.

[2] J. Beck, PubMed Central and the NLM DTDs. Presented at the ASIS&T DASER Summit, Cambridge, MA, November 12–23, 2003 (Online.) Available: http://www.asis.org/Chapters/neasis/daser/Jeff_Beck_presentation.ppt (13 May 2005).

[3] M. Byrnes, Assigning permanence levels to NLM's electronic publications. Presented at Information Infrastructures for Digital Preservation: A One Day Workshop, Dec. 6, 2000, York, England. (Online.) Available: http://www.rlg.org/en/page.php?Page_ID=244 (13 May 2005).
[4] CAMiLEON: Creating creative archiving at Michigan & Leeds: Emulating the old on the new (2001). (Online.) Available: http://www.si.umich.edu/CAMILEON/ (13 May 2005).
[5] CCSDS (Consultative Committee for Space Data Systems), Reference model for an Open Archival Information System (OAIS). Blue Book. Issue 1. ISO 14721:2002, 2002. (Online.) Available: http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf (13 May 2005).
[6] M. Day, Collecting and preserving the World Wide Web: A feasibility study undertaken for the JISC and Wellcome Trust (2003). (Online.) Available: http://library.wellcome.ac.uk/assets/WTL039229.pdf (13 May 2005).
[7] Digital Preservation Coalition, DPC/PADI What's new in digital preservation (2004). (Online.) Available: http://www.dpconline.org/ graphics/whatsnew/ (21 June 2004).
[8] Fedora 2.0: A powerful open-source solution for digital repositories. Summary report in *D-Lib Magazine* In-Brief. **11**(3) (March 2005). (Online.) Available: http://www.dlib.org/dlib/march05/03inbrief.html (13 May 2005).
[9] S. Granger, Emulation as a digital preservation strategy, *D-Lib Magazine* **6**(10). (2000) (Online.) Available: http://www.dlib.org/dlib/october00/granger/10granger.html (13 May 2005).
[10] G. Hodge and E. Frangakis, Digital preservation and permanent access to scientific information: The state of the practice. A joint report by the International Council for Scientific and Technical Information and CENDI (2004). (Online.) Available with free registration: http://www.icsti.org/icsti_reports.html (21 April 2004).
[11] D. Holdsworth and P. Wheatley, Emulation, preservation and abstraction, *RLG DigiNews* **5**(4) (2001), Feature #2. (Online.) Available: http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2 (13 May 2005).
[12] IBM, Royal Dutch Library preserves culture with Content Manager and DB2 (2003). (Online.) Available: http://www-5.ibm.com/nl/dias/resource/rdl.pdf (13 May 2005).
[13] IBM, IBM helps Die Deutsche Bibliothek to develop digital long-term archives (2004). (Online.) Available: http://www-306.ibm.com/software/success/cssdb.nsf/CS/MCAG-65VEUX?OpenDocument&Site= (13 May 2005).
[14] W. Lefurgy, PDF-A: A new digital preservation format, *Government Records News* (May–June, 2003). (Online.) Available: http://www.archivists.org/saagroups/gov/newsletters/May2003.asp#digital (13 May 2005).
[15] R. Lorie, A project on preservation of digital data, *RLG DigiNews* **5**(3) (2001), Feature # 2. (Online.) Available: http://www.rlg.org/preserv/diginews/diginews5-3.html#1 (13 May 2005).
[16] NLA, PANDAS Manual (2004). (Online.) Available: http://pandora.nla.gov.au/manual/pandas/index.html (13 May 2005).
[17] OCLC, OCLC Digital Archive (2005). (Online.) Available: http://www.oclc.org/digitalarchive/default.htm (13 May 2005).
[18] OCLC, PREMIS (PREservation Metadata: Implementation Strategies) (2003). (Online.) Available: http://www.oclc.org/research/projects/pmwg/ (13 May 2005).
[19] OCLC/RLG Working Group on Preservation Metadata, Preservation metadata and the OAIS information model: A framework to support the preservation of digital objects (2002). (Online.) Available: http://www.oclc.org/research/projects/pmwg/pm_framework.pdf (13 May 2005).
[20] E. Oltmans et al., Preservation functionality in a digital archive, in: *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, Tucson, AZ, 2004, pp. 279–286. (Online.) Available for fee: http://portal.acm.org/citation.cfm?id=996350.996416&col (13 May 2005).
[21] Partners go Dutch to preserve the minutes of science, Research Information (September/October 2004). (Online.) Available: http://www.researchinformation.info/risepoct04archiving.html (13 May 2005).
[22] S. Payette, The Fedora Project. Presented at the DLF Forum (17 November 2003). (Online.) Available: http://www.fedora.info/presentations/DLF-Nov2003.ppt (13 May 2005).
[23] J. Rothenberg, Avoiding technological quicksand: Finding a viable technical foundation for digital preservation. Report to CLIR (January, 1999). (Online.) Available: http://www.clir.org/pubs/reports/rothenberg/contents.html (13 May 2005).
[24] J. Rothenberg, An experiment in using emulation to preserve digital publications. NEDLIB Report Series; 1 (April, 2000). (Online.) Available: http://www.kb.nl/coop/nedlib/results/NEDLIBemulation.pdf (13 May 2005).
[25] Royal Library, National Library of Sweden. Kulturaw3 – Heritage project: Long term preservation of published electronic documents (n.d.). (Online.) Available: http://www.kb.se/ENG/kbstart.htm (13 May 2005).
[26] A. Senserini et al., Archiving and accessing web pages: The Goddard Library web capture project, *D-Lib Magazine* **10**(11) (2004). (Online.) Available: http://www.dlib.org/dlib/november04/hodge/11hodge.html (13 May 2005).
[27] T. Staples, R. Wayland and S. Payette, The Fedora Project: An open-source digital object repository management system, *D-Lib Magazine* **9**(4) (2003). (Online.) Available: http://www.dlib.org/dlib/april03/staples/04staples.html (13 May 2005).
[28] J. Steenbakkers, Treasuring the digital records of science: Archiving e-journals at the Koninklijke Bibliotheek, *RLG DigiNews* **8**(2) (2004). Available: http://www.rlg.org/ en/page.php?Page_ID=17068&Printable=1&Article_ID=990 (13 May 2005).
[29] R. Tansley et al., The DSpace institutional digital repository: Current functionality, in: *Proceedings 2003 Joint Conference on Digital Libraries*, C. Marshall et al., eds, pp. 87–97. (Online.) Available: http://portal.acm.org/citation.cfm?

id=827140.827151&coll=GUIDE&dl=ACM&type=series&idx=827140&part=Proceedings&WantType=Proceedings&title=International%20Conference%20on%20Digital%20Libraries&CFID=11792774&CFTOKEN=13954292 (13 May 2005).

[30] University of Virginia Library, UVA Library Central Digital Repository (2005). (Online.) Available: http://www.lib.virginia.edu/digital/resndev/repository.html (13 May 2005).