Records Transition Survival
- When Used over Several Decades

Feng Luan, Mads Nygård
Department of Computer and Information Science
Norwegian University of Science and Technology
Trondheim, Norway
E-mail: {luan, mads}@idi.ntnu.no

Abstract

More and more digital information is created from personal computers and portable devices, but it is difficult to preserve over a long period as it depends on fast evolving computer technology. Without careful protection, we can neither access digital information in old digital storage media, nor interpret the meaning of digital information from old formats after several decades. Thus, a black hole may form in our society concerning information stored in computers. In this paper, we mainly discuss and compare several preservation strategies and preservation systems. At the end, we introduce three challenges for preservation systems.

Relevant topic: Digital preservation and access management, Digital libraries development, architecture, and management

Keywords: state of the art paper, long-term preservation, preservation strategies, digital repositories, DL policy, migration, emulation,

1. Introduction

Digital information is a type of information represented in terms of computer technologies. It can be texts, images, sound, and video. In a study (Gantz, et al. 2008) released by IDC, they estimated that the amount of the total digital world was 281 exabyte[1] in 2007 and the annual growth in the digital world is approximately 60%. The amount of digital information should equal nearly 1800 exabyte in 2011.

Digital information differs from traditional information that is written on paper or is engraved on stones. Digital information is constituted by three levels: *the physical level*, *the logic level* and *the conceptual level*. The physical level is the lowest level. It is constituted by digital storage media and corresponding media drivers. Any digital storage medium utilizes some physical characteristics, like optic or magnetic, to store digital information. Media drivers translate those physical characteristics into bits of 0s or 1s. The bits are organized together by a set of syntactic and semantic rules, which are defined by a format. Each format has a logic model. How to translate bits into a logic model is the main task in the logic level. As a rule, the specification of a format is the key for the translation. Only when we know the specification of the format can we develop an interpreter for the bits. The conceptual level is the top level. The content of digital information derives from a set of digital objects. Those digital objects can adapt a same format or adapt different formats. Software programs will finally render those digital objects as a unit in a human readable manner. For instance, a web page includes texts and pictures. If end users like to view the web page, they can use a web browser application that has integrated text and image format interpreters.

From the three levels concerning digital information, we can see that digital information heavily depends on computer technologies. Losing any related components within the levels will make digital information unreadable. Common threats for digital information are hardware/software failure, hardware/software obsolescence, loss of

[1] 1 Exabyte = 1024^3 Gigabyte = 1024^6 Byte

format specification, malicious modification, context loss and organization failure. Thus, preserving digital information is not easier than preserving traditional information.

In this paper, we are going to investigate the preservation strategies used at varied preservation project. We expect the readers of this paper can get those benefits:

- Know about varied preservation strategies.
- Easily choose the appropriate preservation strategies based on individual situation according to the matrix designed by us.
- Know about popular preservation systems around the world.
- Know about possible challenges for preservation strategies.

The rest of this paper is organized as follows. We first describe the state of the art of preservation strategies in Section 2. Afterwards, we design a matrix to compare the current preservation strategies in Section 3. In Section 4, we introduce seven famous preservation systems. Finally, we discuss research challenges for the preservation systems.

2. Preservation strategies

In this section, we will introduce the state of the art in preservation strategies. (The Digital Preservation TestBed 2001; Lee, et al. 2002; Thibodeau 2002) summarized several preservation strategies. In practice, these preservation strategies are not exclusive. The curators of a preservation system should use several of them.

2.1 Computer museum

Computer museum is a naive strategy. In this strategy, the old computer system that is used to create digital information should be retained. Through the preserved computer system, future users can run any previous software programs. They can furthermore access and view the preserved digital objects by their original software programs. This strategy is naive in terms of the reason as follows. 1) Every component in a computer system, like CPU and hard-disk, has a limited lifetime. Curators of a preservation system must replace broken or outdated components. 2) Every previous software application must be preserved. 3) Most hardware and software are proprietary. Thus, it is impossible for curators of a preservation system to produce the old hardware and software by themselves. 4) The venders of hardware and software do not support all of their previous products. It is expensive when the curators ask the venders to support all of the previous productions.

2.2 Emulation

Emulation is to imitate all functionalities of computer hardware by a software application called emulator, which can simulate the functions of different hardware components. In order to access and interpret preserved digital information, we must know the specification of the original hardware such that we can implement an emulator. Moreover, we should preserve a variety of previous software programs ranging from operating system to application software.

(Granger 2000) argued that emulation is a short term preservation strategy since the target of the preservation is the whole computer system rather than digital information in emulation. Digital information management is also hard for emulation since the preserved information is dispersed over various emulators. In addition, providing long-term accessibility for all digital information requires us to develop emulators for every previous computer system.

It is, however, a generic approach. If the users have an emulator and related preserved programs, they should easily access and view the digital information created in that particular computer system that the emulator is used for. No one needs to develop a program for either the format interpretation or the formation conversion. Therefore, it is helpful for complex digital objects and software applications, both of which are too complex and hard to develop again. For instance, database applications and game applications.

2.3 Encapsulation

Encapsulation serves to preserve digital information, which is in its original format, and the specification of the original format. We must rewrite the source codes for the old format based on the preserved specification. Therefore, a new interpreter for the old format can run at the corresponding computers to access the old digital information. There are two significant defects. 1) Many formats are proprietary, so curators do not have their specifications. 2) Curators, who use this strategy, have much job to developing all of the previous formats.

2.4 Universal Virtual Computer (UVC)

The concept of Universal Virtual Computer (UVC) was proposed by Lorie (Lorie 2001; Lorie 2002) in 2001. UVC is similar to JAVA Virtual Machine. All digital information is saved with the instructions of a UVC. The content can be rendered by a UVC interpreter. In addition, UVC can also execute old software applications. For each generation of computer technology, we have to develop a new UVC interpreter. This concept was proved by a joint project called DNEP (van Diessen and Steenbakkers 2002), which was carried out between KB and IBM in 2003.

2.5 Migration

(Waters and Garrett, 1996) defined migration is '*... a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequence generation*' (Waters and Garrett 1996).

(Wheatley 2001) broke migration into six specific cases: 1) *Minimum Preservation*: preserve the bits of the original digital objects; 2) *Minimum Migration*: convert the original digital objects to very general formats requiring little technical work; 3) *Preservation Migration*: preserve digital objects in the same formats as minimum migration and some annotations, e.g. screenshots, comments to the screenshots, and video clips; 4) *Recreation*: replace the original object with a new object; 5) *Human Conversion Migration*: recreate all objects derived from that original object by manual operations; 6) *Automatic Conversion Migration*: convert digital objects to ones suitable for the current environment.

(Consultative committee for Space Data Systems 2002) divides migration into four categories: *refreshment*, *replication*, *repackaging*, and *transformation*. In refreshment, we copy original digital information from one digital storage medium to the same type of medium without changing the original bits, whilst replication is to simultaneously preserve the copies in distributed places. Both of them can effectively avoid accidental bit loss in digital storage media. On the other hand, repackaging will modify the structure of the collection of digital objects, and transformation will reconstitute a digital object based on the original.

Sometimes, transformation is called format conversion or format migration. It aims to preserve only the content of digital information. Therefore, it must change the original logic model and original bits. The disadvantages of transformation are: 1) data/functionality in the preserved digital information might be lost since formats are incompatible or the converting application has logic faults; 2) we do not know when we should do migration; 3) to choose the right format is time-consuming; 4) authenticity is hard to archive.

(Rosenthal, et al. 2005) further analyzed transformation (they called it format migration) in terms of when to do format migration:

- **Migration on Ingest.** Migration on ingest is called normalization. It transforms formats when they are inserted into the preservation system. It postpones the actual migration and reduces the number of formats in the preservation system.
- **Batch Migration.** Batch migration periodically transforms archived digital information to another format before the old format becomes obsolete.
- **Migration on Access.** Migration on access was proposed by (Mellor, et al. 2002). It preserves the original bit streams in the original format and does the migration only when users want to access preserved digital information. For migration on access, bit streams and appearance are preserved in the original form. Hence, it is easy to prove the authenticity of a digital object by the original bits.

In later years, scientists and curators have tried to enhance their preservation cooperation by distributed technology. Distributed technology can bring many advantages. 1) We save multiple copies in different location and administrate them in various organizations. LOCKSS (LOCKSS 2009) and DSpace (DSpace 2009) tried to build a federation of digital repositories. 2) Distributed technology makes the whole migration system open and adaptable. Entities within a migration system can be developed by different techniques and each changing entity will not affect any other entity. 3) Distributed technology makes a migration system able to cooperate with web services. For example, we can invoke an external service to do format conversion; we can dynamically extract the format's metadata and identify and validate particular formats by web services, e.g., JHOVE (JHOVE 2009) and DROID (DROID 2009). We can get impartial and definitive information about the formats from format registries, e.g., Global Digital Format Register (Global Digital Format Register 2009), PRONOM (PRONOM 2009). In addition, PANIC (Hunter and Choudhury 2006) and CRiB (Ferreira, et al. 2007) are two migration projects, which use web service techniques to find format transformation services and choose the most appropriate service to do migration.

3. Comparing the preservation strategies

In this section, we are going to compare the above preservation strategies, i.e., computer museum, emulation, encapsulation, UVC and migration, based on a matrix. For migration, we adapted Rosenthal's classification, i.e., migration on ingest, batch migration and migration on access. Migration on ingest is removed from the candidates since it does not overcome the risks but reduce the risk degree. The matrix contains six criteria derived from the hierarchy of digital information. Table 1 lists the comparison results. All criteria are discussed below Table 1.

Table 1. Matrix for comparing preservation strategies

	Computer Museum	Emulation	Encapsulation	UVC	Batch Migration	Migration on Access
Change in bits	no change	no change	no change	change once	change many times	change once
Hardware specification	no	yes	no	no	no	no
Hardware components	yes	no	no	no	no	no
Format specification	n/a	n/a	the original format's specifications	the mediatory format's specifications	the latest format's specifications	the original format's specifications
Format converter	n/a	n/a	n/a	from the original format to a mediatory format	from the format currently used to a new format	from the original format to a new format
Format interpreter	the original interpreter	the original interpreter	the new interpreters	the new interpreter	the current interpreter	the current interpreter

- '**Change in bits**' means how many times the bits of the digital object have been changed over its lifetime. For a preservation system, any preserved digital objects usually do not permit to modify. However, it is allowed to do format conversion. Since the models of the formats are varied, the bits of the formats are different also. Here, change refers the format's conversion rather than the content's modification. If the format has changed many times, it is possible to lose some parts for a digital object, such as its structure and functions. Possible answers for this criterion is 'no change', 'change once' or 'change many times'.
- '**Hardware specification**' means whether we should preserve hardware specifications. Sometimes, curators need hardware specification to develop emulation programs to simulate the functionality of the old hardware. For some strategies, this is a critical criterion for preservation systems. The answers should be either 'no' or 'yes'.
- '**Hardware components**' means whether we should preserve hardware components. For some strategies, curators do not need to develop emulation programs, but directly preserve the old hardware components. Through the old hardware components, the digital object can be viewed. The possible answers should be 'no' or 'yes'.
- '**Format specification**' means what format specifications we should preserve. Any format specification is a key to translate the bits of the digital object since the specification defines the format's model and syntax. For a preservation system, formats possibly change over time. The answers should be 'the original format's specifications', 'the latest format's specifications', 'the mediatory format's specifications' or 'n/a'.
- '**Format converter**' means what format converter we need. Format converters can transform the original format to a mediatory format (e.g. the UVC format), or from the original format to a new format already used in the market, or from the format that is currently used to a new format used in market. Hence, the possible answer is one of the conversions above.

- '**Format interpreter**' means what kind of interpreter programs we need to translate formats. For any format, users must rely on an interpreter program to dynamically translate the bits to a human readable manner. According to different types of the formats' specifications, an interpreter program should be 'the original interpreter' for the original format, 'the current interpreter' for the format is in the market, or 'the new interpreter' developed for the old format.

In terms of the criteria listed in Table 1, we define two characteristics, i.e., the possibility for losing data and the difficulty for implementing of preservation strategies, to sort the above strategies.

Figure 1 illustrates this sorting. Computer museum is the easiest strategies and it has the lowest possibility to loss data. We have to remove it from the candidates since it is a naive approach and it is just suitable for a very short-term preservation system.

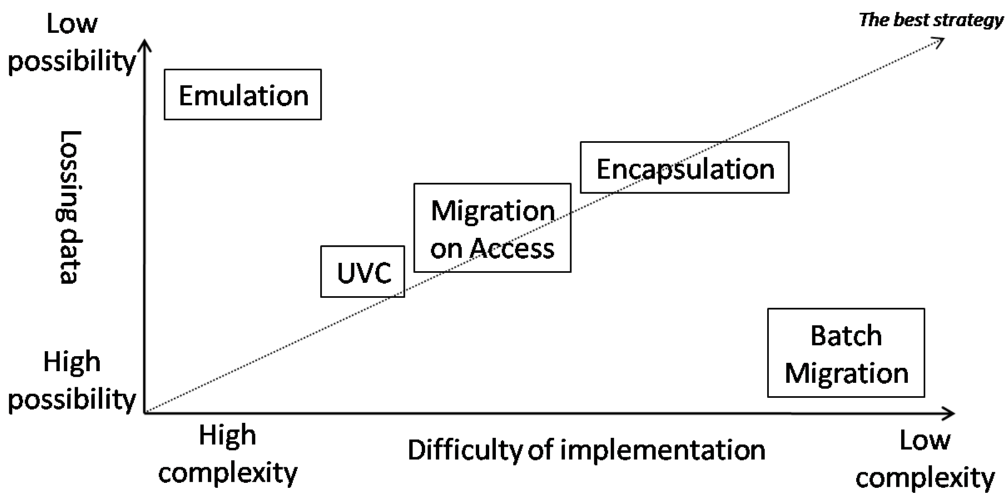

Figure 1. Sorting preservation strategies

The possibility of losing data depends on change in bits, format converter and format interpreter. Those three criteria include two kinds of processes: conversion and interpretation. Conversion is carried out between two different formats. Sometimes, one format cannot completely transform to another format. Some parts will be lost. Interpretation just involves one format. It still has possibility to lose data if the format's specification is not well documented. Moreover, it is too complex to understand the old format specification after several decades. Therefore, change in bits is ranked at the first place among those three factors since the digital object probably involves many times conversion. Format converter is at the second place since it does conversion, whereas format interpreter is the last since it does interpretation. In addition, conversion and interpretation for recent formats will have less probability since the adjacent generations have smaller generation gaps. Additionally, it is easier to implement more functions with the specialists' help.

Difficulty of implementation of preservation strategies includes hardware specification, format specification, format converter, and format interpreter. The difficulty is determined by implementation targets and time interval between technologies. According to our practical experience, implementing hardware simulation programs is harder than developing format converters or format interpreters. Moreover, developing format conversion is more difficult than developing format interpreter. For format conversion, it requires the developers to understand both the old format and the target format. For format interpreter, it just needs the developers to understand the old format. We also argue that understanding recent technology is easier than understanding archaic technology since we might lose related documentations over time. Time interval is the time difference between the created date and the attended date. For recent formats, developers can seek the help from external experts, even the designer, whilst for remote formats few experts can understand them. The developers have to try to implement by themselves.

4. Relevant preservation systems

There are a lot of scientists who design preservation systems. Some preservation systems are small, while some preservation systems are huge and they are used at knowledge centers and culture protection centers. In this section, we are going to briefly introduce seven popular preservation systems.

4.1 LOCKSS (Lots of Copies Keep Stuff Safe)

LOCKSS is an open-source system, under auspices of Stanford University. The first version was released in 2000. After LOCKSS was tested at 50 libraries worldwide, the system was finally released into production in 2004.

The preservation strategy deployed by LOCKSS is migration on access. LOCKSS can embed any plug-in format conversion program at the back end. When a format becomes obsolete or unsupported, this format will be converted to any recognized formats by the plug-in program as users try to view it (Rosenthal, et al. 2005). In addition, LOCKSS is a distributed preservation system, which can manage multiple copies at varied remote data repositories. Through this way, it can effectively avoid bit-rot and malicious modification for a file by recovering this file from another site.

4.2 Eprints

Eprints is a set of open-source software programs for building open access services. Eprints was developed at the University of Southampton in UK in 2000. The most current version is Eprints 3. Comparing to previous versions, Eprints 3 improves in several aspects, i.e., architecture, automatic extraction of metadata, access control, flexible workflows, format support, thumbnails, etc.

However, Eprints 3 is not a long-term preservation system. It does not specify any preservation strategies. They focus on the functionalities to import and export digital objects. For example, there are a number of format options for exporting searching results; relevant metadata are automatically extracted as the digital object has been deposited.

4.3 DSpace

DSpace is another open-source software package, developed by the Massachusetts Institute of Technology Libraries and Hewlett-Packard. The current version is 1.5.2. The objective of DSpace is to provide a repository for the research data sets and educational material. DSpace is more flexible than LOCKSS or Eprints. LOCKSS and Eprints are mainly used for digital journals and digitized documents, whilst DSpace supports more formats, such as 3D digital objects, research data sets and films.

DSpace has specified two kinds of strategies: bit preservation and functional preservation (Smith, et al. 2003). Bit preservation ensures the integrity of the bits for a digital object, while functional preservation ensures that the formats for a DSpace system are always interpretable with existing software programs. Moreover, all formats are classified into ‘supported’, ‘known’ and ‘unsupported’. ‘Supported’ means the specification of a format is public and well-documented. ‘Known’ means a format is popular but its specification is unknown since this format is proprietary. ‘Unsupported’ means a format is either unknown to the librarians or extremely rare. Only the supported formats in DSpace will be periodically transformed to the new formats. The known and unsupported formats will not be converted unless they become supported.

4.4 e-Depot

E-Depot is a long-term preservation system used at the National Library of Netherlands. The core of this system called DIAS, i.e., Digital Information Archiving System (DIAS 2003), was developed by IBM.

In e-Depot, the National Library of Netherlands adapted the UVC strategy (van der Hoeven, et al. 2005). In their experiments, they proved that UVC is a possible approach for long-term preservation systems. They believe that, if more and more institutions develop format decoders for UVC and share them, UVC will be a good approach for long-term preservation systems. They concluded the biggest challenge for UVC is performance. However, the UVC’s designer, Lorie, argued it is difficult to improve (Lorie 2001).

4.5 FEDORA (Flexible Extensible Digital Object Repository Architecture)

FEDORA is an open source digital preservation infrastructure. FEDORA began in 1997 and was developed by Cornell University and University of Virginia. The latest version is 3.2.1. In comparison with other preservation systems, FEDORA provides not only the basic functions for preservation systems but also a model based on the Resource Description Framework (RDF) to provide relation between digital objects.

FEDORA aims at providing an infrastructure for various digital storages. They do not propose any preservation strategies. However, they have a function ‘content repurposing’. This function is opaque to users. It dynamically transforms (i.e., migrates) a digital object at the back end by a content transform service as users try to view a digital object. This function can be extended as a preservation strategy, which is similar to migration on access, when the format in FEDORA becomes obsolete.

4.6 DRIVER (Digital Repository Infrastructure Vision for European Research)

DRIVER is funded under the EU Sixth Framework Programme, Research Networking Testbeds. DRIVER’s main objective is to establish a confederation infrastructure of European digital repositories, offering sophisticated functionality services to both researchers and the general public.

As DRIVER aims at providing an infrastructure to aggregate heterogeneous content sources into uniform shared information spaces, they do not argue preservation issues. Like Eprints and FEDORA, they focus on exporting and importing data sources. (Feijen, et al. 2007)

4.7 iRODS (Integrated Rule Oriented Data System)

iRODS is an open source data grid software system developed by the Data Intensive Cyber Environments (DICE) research group and collaborators. It is the successor to the Storage Resource Broker (SRB). Currently, the most recent version is 2.1.

iRODS does not specify any preservation strategies. It mainly aims at providing a function of distributed storage management, which allows storage media are independent from hardware and software. Actually, users of iRODS can define their preservation strategies according to individual situations. A project was carried out by the Centre for e-Research (CeRch)[2] at King's College London (KCL), and continuing investigations begun by the former Arts and Humanities Data Service (AHDS)[3]. They adapted migration on ingest and batch migration for their iRODS system (Hedges, et al. 2009; Hedges, et al. 2007).

5. Future work

Digital preservation systems are a set of complex systems containing multiple functions, e.g., ingest function, archive function, dissemination function, search function, and user interface function. From the management view, we identified three main challenges:

5.1 Very large amounts of data

As we mentioned in the introduction section, the survey of IDC (Gantz, et al. 2008) indicates that the future digital world is huge and possibly the volume of digital objects exceed the capacity of storage media. They found that disks, tapes and optical occupied approximate 99% of the storage market. Additionally, the volume of information created has exceeded the capability of storage media in 2008. Moreover, the difference of the information created and storage capability will reach almost two times in 2011.

Besides the storage capacity challenge, storage performance is another challenge. According to the situation of the National Library of Norway in 2007, they preserved 957 terabytes of data. And the annual growth for data volume at the National library of Norway is also near 60% per year. For their last replacement of storage media, they spent nearly three months to transfer all digital objects to the new storage system. We have tested the read/write performance of normal storage media, e.g., a hard disk for desktops, an external disk and a memory stick. We found that the hard disk has the highest read and write speed, i.e., both speeds around 50 MB/s, while the memory stick has the lowest read and write speed about 10MB/s and 1.5MB/s, respectively. Thus, for those 957 terabytes, we need

2 www.kcl.ac.uk/iss/cerch/

3 www.ahds.ac.uk

232 days if we use the hard disk, whilst we need 7743 days if we use the memory stick. Therefore, as the amount of digital information rises exponentially, there is a real concern that the time required for the next system replacement will exceed by far the support time, i.e. the previous copying process of all data is still not finished before the next has to be initiated.

5.2 Lack of comprehensive preservation strategies

The preservation system is to provide services for users. In generally, the users have four requirements for preserved digital objects: *Accessibility*, *Viewability*, *Understandability* and *Trustability*. Accessibility requires that the users should be able to retrieve the preserved digital objects from storage media; viewability requires that the users should be able to read the preserved digital objects with the help of some software; understandability requires that the users should be able to fully understand what the content means; and trustability requires that the users should trust the digital objects are authentic.

Among all of the mentioned preservation strategies, emulation is the most suitable and easiest strategies that can satisfy the four requirements. However, this strategy is probably difficult to implement for the curators of a preservation system. For other preservation strategies, especially batch migration and migration on access, most effort is put at storage and format issues. Just accessibility and viewability are satisfied. We need more complete strategies for other requirements. As a system for preserving critical digitized knowledge and information, we believe that any preservation system must have a clear and comprehensive preservation strategy.

5.3 What metadata should be preserved?

Metadata are data about data. They provide supplementary information about an object. Currently, there are numerous metadata standards, such as Dublin Core (DCMI Usage Board 2008), MOREQ 2 (MOREQ2 2008), ISO 23081 (ISO-23081-2 2009), PREMIS (PREMIS Editorial Committee 2008), and the OAIS metadata model (The OCLC/RLG Working Group on Preservation Metadata 2002). However, there are few metadata standards for preservation strategies. We believe that the metadata for preservation strategies should cover those aspects: 1) what metadata is necessary for preservation strategies? 2) what metadata can improve authenticity of the preserved digital objects? 3) what metadata should be recorded after curators carry out a preservation strategy? 4) how to visualize the history for the preserved digital objects?

6. Conclusion

Preservation systems are complex information management systems in terms of computer technologies. Ingest, management, search and dissemination are the important functions for the preservation systems. In this paper, firstly we introduced and discussed several preservation strategies related to the management issue as a digital object is going to be preserved over several decades. Afterwards, we presented seven preservation systems popular worldwide. Finally, we identified three research challenges for preservation strategies. We expect that the readers of this paper will have obtained a general overview for preservation strategies in varied preservation systems.

7. References

Consultative committee for Space Data Systems
2002 Reference Model for an Open Archival Information System (OAIS). <http://public.ccsds.org/publications/archive/650x0b1.pdf>, accessed on Oct 22, 2008.

DCMI Usage Board
2008 Dublin Core Metadata Element Set. http://dublincore.org/documents/dcmi-terms/, accessed on Aug 6, 2009.

DIAS
2003 http://www-935.ibm.com/services/nl/dias/, accessed on Aug 6, 2009.

DROID
2009 http://sourceforge.net/apps/mediawiki/droid/index.php?title=Main_Page, accessed on Aug 6, 2009.

DSpace
2009 http://www.dspace.org/, accessed on Aug 6, 2009.

Feijen, Martin, et al.
2007 DRIVER: Building the Network for Accessing Digital Repositories across Europe. Ariadne 53.
Ferreira, Miguel, Ana Baptista, and José Ramalho
2007 An intelligent decision support system for digital preservation. International journal on digital libraries 6(4):295-304.
Gantz, J.F., et al.
2008 The Diverse and Exploding Digital Universe: An updated forecast of worldwide information growth through 2011. http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf, accessed on April 20, 2008.
Global Digital Format Register
http://hul.harvard.edu/gdfr/, accessed on Aug 6, 2009.
Granger, Stewart
2000 Emulation as a Digital Preservation Strategy. D-Lib Magazine 6(10).
Hedges, Mark, Tobias Blanke, and Adil Hasan
2009 Rule-based curation and preservation of data: A data grid approach using iRODS. Future Generation Computer Systems 25(4):446-452.
Hedges, Mark, Adil Hasan, and Tobias Blanke
2007 Management and preservation of research data with iRODS. *In* Proceedings of the ACM first workshop on CyberInfrastructure: information management in eScience. Lisbon, Portugal: ACM.
Hunter, Jane, and Sharmin Choudhury
2006 PANIC: an integrated approach to the preservation of composite digital objects using Semantic Web services. International Journal on Digital Libraries 6(2):174-183.
ISO-23081-2
2009 Information and documentation -- Managing metadata for records. http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?ics1=01&ics2=140&ics3=20&csnumber=50863.
JHOVE
2009 http://hul.harvard.edu/jhove/, accessed on Aug 6, 2009.
Lee, Kyong-Ho, et al.
2002 The State of the Art and Practice in Digital Preservation. Journal of Research of the National Institute of Standards and Technology 107(1):93-106.
LOCKSS
2009 http://www.lockss.org/lockss/home, accessed on Aug 6, 2009.
Lorie, Raymond A.
2001 Long term preservation of digital information. *In* Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. Pp. 346-352. Roanoke, Virginia, United States: ACM.
Lorie, Raymond A.
2002 A methodology and system for preserving digital data. Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, Portland, Oregon, USA, 2002, pp. 312-319. ACM.
Mellor, Phil, Paul Wheatley, and Derek M. Sergeant
2002 Migration on Request, a Practical Technique for Preservation. Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries, 2002, pp. 516-526. Springer-Verlag.
MOREQ2
http://www.moreq2.eu/index.htm, accessed on Aug 6, 2009.
PREMIS Editorial Committee
2008 PREMIS Data Dictionary for Preservation Metadata.
PRONOM
2009 http://www.nationalarchives.gov.uk/pronom/, accessed on Aug 6, 2009.
Rosenthal, David S. H., et al.
2005 Transparent Format Migration of Preserved Web Content. D-Lib Magazine 11(1).
Smith, MacKenzie, et al.
2003 DSpace: An Open Source Dynamic Digital Repository. D-Lib Magazine 9(1).
The Digital Preservation TestBed
2001 Migration: Context and Current Status. http://www.digitaleduurzaamheid.nl/bibliotheek/docs/Migration.pdf, accessed on Oct 22, 2008.

The OCLC/RLG Working Group on Preservation Metadata

2002 Preservation Metadata and the OAIS Information Model. A Metadata Framework to Support the Preservation of Digital Objects. http://www.oclc.org/research/pmwg/pm_framework.pdf, accessed on

Thibodeau, Kenneth

2002 Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. CLIR Reports, Conference Proceedings of The State of Digital Preservation: An International Perspective.

van der Hoeven, J. R., R. J. van Diessen, and K. van der Meer

2005 Development of a Universal Virtual Computer (UVC) for long-term preservation of digital objects. Journal of Information Science 31(3):196-208.

van Diessen, R. J., and J.F. Steenbakkers

2002 The Long-term Preservation Study of the DNEP Project: An Overview of the Results. IBM/KB Long-term Preservation Study Report Series.

Waters, Donald, and John Garrett

1996 Preserving Digital Information. Report of the Task Force on Archiving of Digital Information.

Wheatley, P.

2001 Migration–a CAMiLEON discussion paper. Ariadne 29(2).