NRC Publications Archive
Archives des publications du CNRC

Long term digital preservation: an end user's perspective
Doyle, Julie; Viktor, Herna L.; Paquet, Eric

NRC·CNRC

*Long Term Digital Preservation - An End User's Perspective ***

Doyle, J., Viktor, H.L., Paquet, E.
October 2007

Canada

Long Term Digital Preservation - An End User's Perspective

Julie Doyle, Herna L. Viktor
School of Information Technology and Engineering
University of Ottawa
800 King Edward Ave., Ottawa
ON K1N6N5, Canada
Email: jdoyle, hlviktor@site.uottawa.ca

Eric Paquet
Visual Information Technology
National Research Council
M50 Montreal Rd., Ottawa
K1A0R6, Canada
Email: eric.paquet@nrc-cnrc.gc.ca

Abstract— The issue of long term preservation of digital data has, in recent years, become a critical issue with many diverse groups and organisations recognising the need to preserve digital documents before they fall victim to digital obsolescence. However, there is a distinct lack of discussion regarding the needs of future end users of such preserved documents. This is surprising, given that it is the end user who ultimately determines what information should be archived and in what way. Furthermore, it is essential that a preserved document retains its authenticity and usability through time and is easily interpreted by future end users. To this extent, we present in this paper details of our case study: an emulation framework to preserve 2D and 3D anthropometric data. We describe the testing of this emulation environment, by anthropometric experts, against the original environment and compare the two based on a given set of usability and authenticity criteria.

I. INTRODUCTION

There are a number of challenges associated with the long term preservation of digital data. Of most interest to us for this paper are the requirements of future end users of a preserved digital document. It is crucial when implementing an archival system for the long term preservation of digital data, to consider the end user's needs with respect to the preserved digital document. Such considerations aid in determining exactly what information (e.g. metadata, software) should be preserved along with the digital document and in what way. While we cannot predict everything an end user may want to do with a preserved digital document in the future, we can assume that they will expect, at least, to have the ability to view or interact with the data in the same way as today's users. As such, it is critical that preserved documents can be rendered authentically on future computers. Moreover, the digital document should be interpretable and understandable to future end users, as well as remaining usable.

The question we need to address therefore is "What can we do to ensure that a preserved digital document remains authentic and usable in the distant future?" For a definition of authenticity, we consider Rothenberg's authenticity principle which states that for a digital document to be interpreted authentically in the future, it should "exhibit as much as possible of its original behaviour, functionality, look and feel, as well as its content" [1]. This is particularly true

for interactive documents which are inherently dependent on behaviour. The ISO standard definition of usability states that usability is "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [2]. Within the context of long term data preservation, usability then refers to the extent to which future end users can view and interact with the preserved document in the same was as users in the past could, i.e. authentically.

From this, we can compose a list of a typical future end user requirements in terms of usability:

1) The end user should be able to access the preserved digital document.
2) The content should be executable, i.e. the host machine should be able to render the document in its original environment.
3) The end user should be able to interpret and understand the content of the digital document.
4) All the above should be possible without the end user experiencing any errors or complexity [3].
5) The preserved document should be authentic in the sense that it is the same document that was preserved and both its content and behaviour have remained unchanged.
6) Metadata should accompany the digital document instructing the future end user on how to execute the document, as well as explaining the document content, its intended behaviour and a description of the software required to run it.

Given this set of requirements, it is necessary to design and implement a preservation environment that will ensure each is met. This becomes more difficult when the preserved document is interactive in nature and this interactivity must also be preserved.

While a number of different approaches have been proposed as solutions to the problem of long term digital data preservation such as encapsulation [4], reliance on standards, and migration [5], emulation emerges as the most realistic approach if a document's program behaviour as well as its original content is to be preserved for long term future access [1]. Within the context of long term data preservation

emulation is the process of recreating on current hardware (the host environment) the technical environment required to view and use digital documents from earlier times [6]. This is possible through the use of an emulator that runs on an emulation virtual machine. The emulator mimics the digital document's original hardware environment, including CPU, memory and peripherals. Along with a saved bitstream of the software application required to display and interact with the document, and the saved document itself, the emulator can run on a virtual machine, hence recreating the document's original environment.

This paper discusses the initial design, implementation and testing of a preservation framework for 2D and 3D anthropometric digital data with the aim of focusing on future end users' needs to aid in this process. The contribution of the paper relates to the demonstration, through our experimental process, that future users of a digital document can access, view, interact with and interpret a digital document under emulation in exactly the same way as was possible within the original environment. This is a valuable contribution as a digital record can only be considered authentic and usable if it can be rendered faithfully on future computers and interpreted intelligibly by future end users.

The remainder of this paper is organised as follows. Section 2 outlines other approaches to authentically preserving digital documents for long term future access. Section 3 describes our preservation framework and discusses emulation in more detail. Our experimental design is presented in Section 4, while Section 5 discusses our findings. Finally, Section 6 concludes and outlines some areas for future work.

II. Other Approaches

Despite the importance of end user needs in long term preservation, comparatively little research addresses this issue. Most of the current literature on preservation focuses on archiving and preservation techniques. As such, factors such as ensuring the authenticity and usability of preserved digital documents are often overlooked. An exception to this is Gladney's work on preserving trustworthy digital objects [3], [7]. However, Gladney's viewpoint on authenticity relates to trusting the content of a preserved digital document. As such, he suggests signing a digital object which will inform future users that the content can be trusted. The end user's requirements, with regard to how they can view and use the preserved document, are not considered.

There are many arguments in the literature for the use of emulation as a digital preservation strategy, particularly in cases where it is important to preserve program behaviour as well as the content of a digital document. However, the majority of the work so far is purely theoretical. The greatest proponent of the emulation approach is Jeff Rothenberg who defined the concept of the Emulation Virtual Machine (EVM) [8]. An EVM is an additional layer that sits between the host hardware platform and the actual emulator. The main advantage of the EVM is its ability to run on various host platforms. Other components of this approach include an emulator specification document that describes the document's original hardware platform, and an emulator specification interpreter which should produce a working emulator of the original hardware based on this specification. However, a drawback of this approach is that Rothenberg suggests that the emulation interpreter be developed in the future, by some future user, rather than at present. Such practice, however, cannot guarantee that either the original document will be accessible and interpretable in the future or that the program behaviour will be faithfully preserved.

Raymond Lorie takes a somewhat different approach to solving the digital longevity problem [9], [10]. With regard to program archiving, Lorie suggests archiving the actual program required to display/interact with the data and writing a Universal Virtual Computer (UVC) program that emulates every instruction of the current hardware machine. In this way, a future UVC interpreter will be able to interpret the original machine's instructions. This approach assumes that for each existing or new machine at present, manufacturers will provide an emulator of that machine written in UVC code. This is a big assumption in itself, as again there is no guarantee that manufacturers will agree to this. Furthermore, Lorie does not specify the details of the UVC code.

Ongoing research at the National Library of the Netherlands describes the concept of modular emulation in an attempt to preserve digital documents [11]. This model combines Lorie's idea of the Universal Virtual Computer with Rothenberg's ideas on an emulator specification. The authors agree that emulation is the best strategy for preservation of digital documents dependent on functionality. They propose a modular structure for emulation, ensuring both flexibility and the easy assimilation of future hardware components. However, this approach has yet to be implemented and tested.

III. Our Preservation Framework

The focus of this paper is to illustrate that a digital document, preserved through emulation, can remain both authentic and usable by future end users over long periods of time. As such, we provide here an overview of our preservation framework, including a description of the type of data we are preserving in addition to a discussion on emulation within our framework.

A. CAESAR^{TM} 3D Data and its Long-Term Preservation

The CAESAR^{TM} database is a repository of anthropometric data. Anthropometry is the study of human body measurements (e.g. weight, height, proportions) and its biochemical characteristics (e.g. stature, size of body parts etc) [12]. There are many application areas of anthropometry including design of restricted spaces such as airline and vehicle seats, design of spaces to provide easy access to elderly and disabled people, and retail tailoring. The latter is the focus of our current research. It represents an important application area of anthropometry, to ensure that clothing fits a person well, in addition to being aesthetically pleasing and comfortable.

Our database resides in an IBM DB2 object-relational database - our archive. Specifically, our archive houses 3D human body scans along with descriptive relational data and 2D views of the scans. Preserving this information, particularly the 3D data, is very important for a number of reasons. Many 3D formats are proprietary and so their lifetime is limited to the lifetime of the underlying corporation, which may only span a few years. The same is true of the software required to view and interact with such 3D formats in addition to raw data formats. Neglecting to preserve such information may result in a 'Digital Rosetta Stone', whereby application software becomes obsolete, resulting in data that can no longer be accessed and interpreted by future end users [13].

Furthermore, the field of 3D anthropometry is still in its infancy. This implies that we cannot predict what future users will want to do with 3D anthropometric data. We therefore need to ensure that all the data are available, accessible and usable by future users. This means preserving not only the final 3D data, but also the software necessary for viewing and interacting with them, as well as the raw data produced from body scans. Preserving this raw data will allow for post-processing of it in the future with more sophisticated algorithms and techniques. For example, there are a number of flaws related to current scanning technology which result in holes in body scans, discrepancies in colours, etc. The technology to fix these holes is much better now than it was even two years ago. So we might want to compare what we can see now with what we could see two years ago; and we might want to do the same thing five or ten years from now. Thus, preserving 3D data over long periods of time can also result in the improvement of the 3D data. To ensure that this preserved information remains accessible and interpretable in the future by end users, emulation is used as a preservation strategy.

B. Emulation as a Preservation Strategy

Our primary motivation for focusing on emulation as a preservation strategy is to allow us to preserve the software environment within which a particular digital document is displayed and manipulated. This essentially allows us to preserve program behaviour in addition to the document's content. A significant benefit of this preservation strategy for future end users is that they can access a preserved document and use its original software environment to view and interact with the document in the same way as original users could.

Preserving a digital document requires encapsulating (within the preservation archive) the following information:

1) The digital document to be preserved (saved as a bitstream).
2) The document's original software environment (saved as some executable bitstream).
3) An emulator of the document's original computing hardware platform. All attributes of this hardware platform required to recreate the behaviour of the digital document in its original software environment need to be specified in the emulator.

4) Metadata, explaining to future users how to 'open' and use the preserved environment. Metadata should also include information on the document's origin, details of its software environment and any other information deemed relevant to aid future users in accessing and using the preserved document. Such information should be saved in human-readable format.

When a future user wishes to access a preserved digital document, a request is sent to the archive and upon approval the encapsulated information is returned to the user. To ensure the preserved environment remains viable through time it is necessary to refresh the storage media as required and to test the emulator in each new computing environment that emerges.

C. Our Emulation Environment

As discussed in Section 2, the majority of current research into emulation as a preservation strategy focuses on developing techniques to specify a computer's hardware architecture along with a program which can be run on this specification in the future to 'interpret' it and create the emulator. It is our opinion, however, that the emulator should be developed at present to avoid the need to rely on an interpreter and also to allow testing at present. While we considered writing an emulator program for this current research, we realised that this would be a time-consuming task which would delay the initial testing of our framework. We therefore decided to integrate an existing commercial emulator into our preservation framework, with the aim of further investigating the development of our own emulator program for future iterations of testing.

Commercial emulators are widely used and accepted within the field of computer science. While such emulators have not necessarily been implemented for long term data preservation purposes, some are ideally suited for such purposes as they allow one type of computer or operating system to be run on another (e.g. Apple on Intel, Windows on Linux). VMware Server is one such emulator. [1] It is available as freeware and is the chosen virtual machine for our emulation environment as it is both robust and easy to use. There are a number of advantages to using such an existing commercial emulator. Firstly, it has been developed and tested by hardware experts. This is important because it guarantees that all hardware components have been correctly emulated. Secondly, it enables us to efficiently test an emulation environment and show that emulation, using existing emulators, is a viable strategy for long term data preservation. This in turn allows us to concentrate our research efforts on ensuring the preserved document can be viewed and used authentically within our emulation environment.

Our emulation environment consists of our host (future) computer platform along with our Virtual Machine (VM),

---

[1]The concepts of a Virtual Machine and Emulator are very closely related and the two terms are often used interchangeably. The difference lies in the fact that a Virtual Machine does not correspond to any one physical machine but is rather a generalisation of a hardware architecture implemented through software.

VMware, which sits atop the host platform. The VM recreates the hardware architecture of our original platform, which runs the original application software necessary for rendering and interacting with the digital document. Thus, running the VM on the host computer virtually recreates the digital document's original environment. Our experiment, the design and setup of which is described next, aims to determine how authentically usable the preserved document is compared to its original counterpart.

IV. EXPERIMENTAL DESIGN

A very important question often ignored by many researchers in the field of long term data preservation is "What can we do to ensure that a preserved digital document remains authentically usable in the future for its intended end users?"

While there may be no definitive answer to this question, there are a number of steps we can take in an attempt towards achieving this authentic usability to the best of our current ability. These include: (1) Creating a preservation environment and determining what needs to be preserved and how; (2) Testing this environment on future computers with intended end users and (3) Soliciting information from intended future end users regarding their opinions on the authenticity and usability of the preserved document compared to its original counterpart. As such, we describe in this section details of the initial phase of an iterative experimental process that will be examined over a period of time, namely the effects that preserving digital documents have on future end users. This preliminary experimental procedure consists of a usability test conducted by a number of anthropometric experts, to (1) determine whether our emulation environment has successfully preserved a digital document's authentic look and feel and behaviour and (2) provide opinions on how to increase the usability of the preserved environment for future end users.

As the CAESAR™dataset consists of both 2D and 3D image data, it was decided to test two separate environments for viewing and manipulating each respective type of image. These two environments will be discussed below. As this was a preliminary phase to future, large-scale experiments, we had 3 expert subjects participate. Each subject was very familiar with the CAESAR™dataset and had worked with this data on numerous occasions prior to the experiment. It was important for this experimental phase to have anthropometric experts participate for a number of reasons. Firstly, our 3D viewer has been developed for domain experts, so it was essential that such experts test it. Furthermore, experts are highly qualified and thus capable of easily observing any differences between the original and preserved environments. Therefore, they could comment faithfully on the capability of emulation in retaining the usability and authenticity of the preserved digital documents (both 2D and 3D). Moreover, they could provide significant feedback regarding what metadata should be preserved within the preservation environment to further increase usability for future end users. As such, our expert users would play a critical role in 'fine-tuning' our preservation
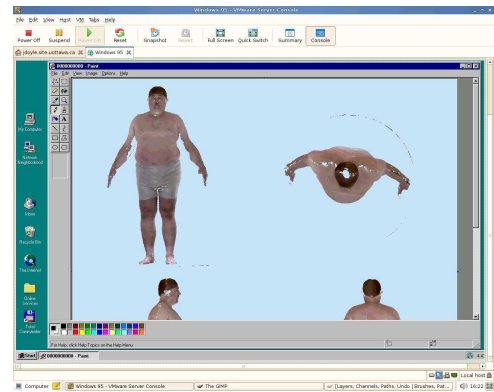

Fig. 1. 2D Emulation Environment

environment prior to future iterations of our experimental process.

A. 2D Environment

One of the primary requirements of preserving digital data is the ability to access the data on future computers. However, as no 'future computer' currently exists, we looked to the past and decided to work with an older computing environment, namely one running Windows 98. Preserving an older computing environment and testing it on current computer hardware would provide our 'future' context. As we wanted our anthropometric experts to compare the preserved environment (under emulation) with the original environment, we set both up side by side. Our 2D original environment consisted of the following: (1) An old IBM PC purchased in 2000 running Windows 98; (2) Microsoft Paint run under Windows 98; (3) A 2D bitmap image from our CAESAR™database displayed within Paint. Our 2D emulation environment consisted of the following components (Figure 1): (1) An IBM ThinkCentre PC purchased in 2007 running Suse Linux (version 10.1); (2) VMware Virtual Machine running Windows 98; (3) Microsoft Paint run under Windows 98 on VMware; (4) A 2D bitmap image from our CAESAR™database displayed within Paint.

We chose Paint as our current 2D image viewer as it is a basic application that suits our purposes of looking to the past. Decades from now, image viewers and editors will exist that are far more sophisticated than they are today. Hence, future users will perceive an application environment that has been preserved for many years as very basic. We are interested in how future users will deal with working within such a basic environment, considering how accustomed they will be to more advanced technology. Our expert users simulate these future users, as they too must acclimate to a more basic environment than they are used to.

Each expert subject was given a task scenario, consisting of a number of sub-tasks, which they were to complete in both the original and test environments. The tasks were related to the application area of virtual tailoring. The subject was presented with the digital document - a 2D image of a virtual mannequin which was created from a 3D body scan of a human. Such scans frequently produce imperfections in the
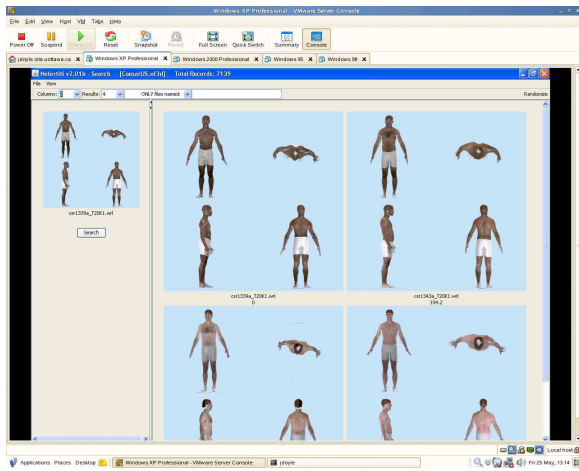
Fig. 2.   3D Emulation Environment

resulting images, such as holes and tears on the body, due to the flaws of current scanning technologies. The subjects' tasks therefore consisted of filling and fixing holes on the body, ensuring the colours of these 'fills' matched (as closely as possible) skin/hair colour, fixing any seams on the body, exchanging body parts of mannequins and simulating clothing design (e.g. sketching a dress on the body). Each subject then completed a questionnaire comparing the two environments.

B. 3D Environment

A similar environment was set up to test the emulation of Cleopatra, a 3D content-based indexing and retrieval system for anthropometric databases. Cleopatra describes human bodies according to shape, data and measurements. The result is a unique data mining system for anthropometric databases. Cleopatra retrieves similar individuals and clusters the population according to various criteria like three-dimensional shape, anthropometric measurements and statistical data. Cleopatra has been developed for Windows XP and it is not backward compatible with Windows 98. Therefore, this environment required the emulation of Windows XP. Specifically, our 3D original environment consisted of the following: (1) An IBM ThinkCentre PC purchased in 2007 running Windows XP; (2) Cleopatra run under Windows XP; (3) 3D images from our CAESAR^TM database displayed within Cleopatra. Our 3D emulation environment consisted of the following (Figure 2): (1) An IBM ThinkCentre PC purchased in 2007 running Suse Linux (version 10.1); (2) VMware Virtual Machine running Windows XP; (3) Cleopatra run under Windows XP on VMware; (4) 3D images from our CAESAR^TM database displayed within Cleopatra. Again, each environment was set up side by side for easy comparison by our anthropometric experts.

A new task scenario was provided for the 3D environment. These tasks consisted of loading a number of databases, searching for similar 3D bodies within the database as well as from outside the database, reinitiating the search from one of the results, saving the results and exporting a database.

In addition, our experts were asked to view four views of a body at high resolution and to obtain metadata about a body. A similar questionnaire was then completed by our subjects, comparing the the usability and authenticity of the preserved environment against that of the original.

C. Recorded Information

The main objective of this experiment was to determine whether a digital document is as usable and authentic when preserved under emulation as its original counterpart. We were therefore interested in recording a number of things in both our 2D and 3D environments:

1) Was the emulation environment as usable as the original? I.e. was it as easy to complete the required tasks in the emulation environment as it was in the original environment?
2) Was the document under emulation 'authentic'? I.e. was it the same as the document in the original environment?
3) Was there any loss of functionality in the emulated environment?
4) Did the emulated environment have the same look and feel as the original?
5) Was the content of the document under emulation the same as the original? I.e. had it remained unchanged?
6) Was there any notable difference in response time when loading the two environments, saving within the environments or performing individual tasks within the two environments?
7) What extra metadata should be stored along with the emulation environment to ensure the preserved document remains usable by future users in terms of how they can view and interact with the preserved document.

This information was recorded by means of a questionnaire completed by each subject at the end of each testing session (2D and 3D). We hypothesized that our subjects would observe no notable differences in the usability, content, behaviour and look and feel of the preserved document. We felt there would be discrepancies in response times of the two environments but were unsure whether the emulated environment would be faster or slower, given the 'extra layer' of the virtual machine.

V. DISCUSSION

Our primary objectives for this preliminary experiment were to illustrate that, even to the trained expert eye, 2D and 3D anthropometric data remains as authentic and usable under emulation as it was in its original application environment. We were also interested in soliciting expert opinions on what metadata should be saved with the emulated environment to ensure a high level of usability for future end users. As we hypothesized, our expert subjects observed no differences in the look and feel and general usability of the original and emulation environments in both the 2D and 3D testing contexts. Hence, they agreed that emulation is a viable preservation strategy for maintaining the original look and feel, behaviour and usability of a digital document. It was noted, however, that the emulated environment was slower when loading the

initial application (both within Paint and Cleopatra) and when saving the manipulated images, although the difference was put at no more than 10%.

Within the 2D testing environment, all expert subjects commented on the lack of advanced functionality for interacting with the images and stated that it is easy to forget the functionalities of an older program when you have become accustomed to more sophisticated application programs. As such, the subjects recommended adding detailed metadata explaining the functionalities and behaviour of the preserved application, specifically its limitations. A further source of surprise for each subject were the 16-bit images and colours. Even though each subject was aware they were working with an older computing environment, each commented that they expected to see 32 bit (or more) colour images. This is due to the fact that our appreciation of colours in the past is not the same as it is today. As such, some subtle details may have been overlooked in older application environments in the past. It is therefore necessary to explain "old" image quality to future end users to provide them with a clearer picture of the preserved data they are working with.

As mentioned above, the Cleopatra viewer was slower to load within the 3D emulation environment than it was in the original 3D environment. However, this difference in load speed can be attributed to the fact that Cleopatra was run directly from a usb 2.0 disk drive, whereas VMware Server currently only supports usb 1.0. Once the database had fully loaded, however, response time for searching, changing the number of results and other tasks was perceived as no slower than within the original environment. The main issue concerning the 3D emulation environment was that each time a subject tried to import and view a new 3D body scan into the database, Cleopatra shut down. This is due to the incompatibility of the graphics card, the emulator and the Cleopatra software. The lack of transparent 3D support is a problem with emulators in general and will remain a focus for us in future research.

VI. CONCLUSION AND FUTURE WORK

In recent years, the issue of long term preservation of digital data has reached new heights of awareness and has been afforded more discussion in the literature. However, the requirements of future end users of preserved documents have been largely overlooked, despite the fact that it is ultimately the end user who determines exactly what should be saved within the preservation environment and in what way. Moreover, it is critical to ensure that a preserved document retains its authenticity and its usability through time so as to allow future end users to access, view and interact with the document in the same way as users in the past could. The aim of this paper was therefore to address this issue, that is, to focus on end user needs when creating a preservation environment and to have typical future end users test this environment to determine the usability and authenticity of a preserved document when compared to its original equivalent.

As such, this paper presented details of our emulation preservation environment, focusing on a preliminary testing of this framework by anthropometric experts. The aim of this experiment was to determine whether our emulation environment was a viable framework for preserving the content, look and feel and behaviour of 2D and 3D digital documents as well as ensuring that future end users could easily interpret the preservation environment. It was the opinion of our expert subjects that the preservation environment was indeed as authentic and usable, given the above criteria, as the original environment. In addition, each subject provided us with highly beneficial ideas on what metadata we should include within our framework to allow future end users to easily understand and interpret the preserved digital document and its environment.

We are currently planning a larger scale experiment which will consist of approximately 40-60 subjects. Such an experiment will incorporate our results from this current evaluation and focus more on determining if the metadata saved with the preserved 2D and 3D digital images is sufficient for end users to understand, interpret and use all aspects of the preservation environment. We are also interested in investigating how future users might interact with preserved documents considering interaction techniques will most likely have completely changed decades from now.

REFERENCES

[1] J. Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Council on Library and Information Resources, Washington, USA, 1998.

[2] International Organisation for Standardization: *ISO 9241-11- Guidance on Usability*. In http://www.iso.org, 1998.

[3] H. M. Gladney, *Principles for Digital Preservation: Focusing on End User' Needs rather than those of Archiving Institutions*. In Communications of the ACM, Vol. 49., No. 2, 2006.

[4] A. Waugh, R. Wilkonson, B. Hills and J. Dell'oro, *Preserving Digital Information Forever*. In 5th ACM Conference on Digital Libraries, pp. 175-184, San Antonio, Texas, 2000.

[5] P. Wheatley, *Migration - a CAMiLEON Discussion Paper*. In Ariagne, No. 49. Available at http://www.ariadne.ac.uk/issue29/camileon/.

[6] D. Holdsworth and P.R. Wheatley, *Emulation, Preservation and Abstraction*. In RLG Diginews, Vol. 5, No. 4, 2001. Available at http://www.rlg.org/preserv/diginews/diginews5-4.html.

[7] H.M. Gladney, *Trustworthy 100-Year Digital Objects: Evidence After Every Witness is Dead*. In ACM Transactions on Information Systems, Vol. 22, No. 3, pp. 406-436, 2004.

[8] J. Rothenberg, *An Experiment in Using Emulation to Preserve Digital Publications*. Koninklijke Bibliotheek, The Hague, The Netherlands, 2000.

[9] R.A. Lorie, *Long Term Preservation of Digital Information*. In 1st ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 346-352, Roanoake, Virginia, 2001.

[10] R.A. Lorie, *A Methodology and System for Preserving Digital Data*. In 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 312-319, Portland, Oregon, 2002.

[11] J. van der Hoeven and H. van Wijngaarden, *Modular Emulation as a Long-Term Preservation Strategy for Digital Objects*. In 5th International Web Archiving Workshop (IWAW '05), Vienna, Austria, 2005.

[12] H.L. Viktor, E. Paquet and H.Guo, *Measuring to Fit: Virtual Tailoring through Cluster Analysis and Classification*. In ECML/PKDD '06, Springer-Verlag, 2006.

[13] A.R. Heminger and S.B. Robinson, *The Digital Rosetta Stone: A Model for Maintaining Long-Term Access to Static Digital Documents*. In Communications of the Association for Information Systems, Vol. 3, No. 2, 2000.