ESARBICA NEWSLETTER

In this issue:

Modular emulation – a long term preservation strategy undertaken jointly by the National Library and the National Archives of the Netherlands ***(****By Remco Verdegem****) pg 6-19***

*"*While both the KB and *Nationaal Archief* have in the past struggled to fulfil their mandates on long-term preservation…it has become increasingly clear there needs to be a systematic assessment of long term preservation strategies. In April 2005 a joint project was started in order to emulate the components of the hardware architecture as individual emulators and interconnecting them in order to create a full emulation process ***."***

ALSO Inside:

ESARBICA NEWSLETTER

2006 Issue No: 15

July

For more information, contact the editors Mr. S. Katuu skatuu@yahoo.com or Dr. P Ngulube ngulubep@nu.ac.za

Events

July 2006

10th – 14th **Intermediate archives and records management** Training course to be convened by the Centre of Applied Communication at UNISA. Deadline for registration **30th June 2006.** Contact Mr. Mathews Kokong Tel: (+2711) 471 3896 Fax: (+2711) 471 3906 Email: mkokong@unisa.ac.za Website: http://www.unisa.ac.za/Default.asp?Cmd=ViewContent&ContentID=17563

10th – 14th **Libraries as a bridge to information and knowledge society** Conference convened by the Standing Conference of Eastern, Central and Southern African Library and Information Professionals (SCECSAL XVII) in Dar es Salaam, Tanzania. Contact Dr Alli Mcharazo, P O Box 33433, Dar es Salaam, Tanzania Tel: +255 744 296134 Emails: tla_tanzania@yahoo.com or amcharazo@muchs.ac.tz Website: http://www.tlatz.org/scecsal2006/

12th – 14th **Information Lifecycle Management** Conference to be convened by Longsight Communication**.** Contact Saimon Chirume, Tel: (012) 320 6770, Fax: (012) 326 5080, Cell: 072 661 0714 E-mail: saimon@longsight.co.za Website: http://longsight.co.za/

25th – 28th **SAPCOM conference on preservation** Convened in Kimberly by SAPCOM (South African Preservation and Conservation Group). Contact Lesley Hart, University of Cape Town Libraries, Private Bag X3, 7701 Rondebosch, South Africa Tel: +27216504424 Fax: +27216861505 Email: lesley@uctlib.uct.ac.za

31st – 2nd Aug **Comprehensive document and records management** Conference convened by Marcus Evans in Johannesburg, South Africa. Contact Hennie Potgieter Tel: (27)11516 1073Fax (27)115161004 Website: http://marcusevans.datapro.co.za/pdfs/SA525.pdf

Aug 2006

2nd – 4th **Managing paper based records** Conference convened by Professional Corporate Training in Johannesburg. Contact David Tel: 011 781 6922/7, Fax 011 7816933, Email: pctint@telkomsa.net

21st – 1st September **Records and information management** Course convened by the Institute of Development Management in Lesotho. Contact: Matthias Chida Email: mchida@idmbls.com Website: http://www.idmbls.com/crs/crsrec.html

28th – 15th September **Advanced international records management training program** (phase 1) Conducted by the Swedish National Archives. Deadline for application was 1st March 2006 Contact: Riksarkivet C/o Hanna Eriksson Box 12541

SE-102 29 Stockholm, Sweden Fax +4687376474 Phone +4687376350 Website: http://www.ra.se/ra/recordsmanagement/index_eng.asp

Sept 2006

11th – 15th **Advanced archives and records management** Training course to be convened by the Centre of Applied Communication at UNISA. Deadline for registration **1st September 2006.** Contact Mr. Mathews Kokong Tel: (+2711) 471 3896 Fax: (+2711) 471 3906 Email: mkokong@unisa.ac.za Website: http://www.unisa.ac.za/Default.asp?Cmd=ViewContent&ContentID=17563

25th – 29th **Libraries: partners in learning, nation building and development** Conference convened by Library and Information Association of South Africa (LIASA) Ninth Annual Conference. Contact: Ms Naomi Haasbroek Tel: +(27)0218431259 Fax: +(27)218433525 Email: naomi@tlabs.ac.za Website: http://www.liasa.org.za/conferences/conference2006/cfp.php

Oct 2006

16th –10th Nov **Reducing risks to collections** ICCROM course convened in Ottawa, Canada and organized in collaboration with the Canadian Conservation Institute. The application deadline was 28th February 2006 at http://www.iccrom.org/eng/training/forms.htm

16th –10th Nov **Sharing conservation decisions** ICCROM course convened in Rome, Italy and organized in collaboration with the Institut National du Patrimoine (France), the Instituto Centrale di Restauro and the Opificio delle Pietre Dure (Italy). The application deadline was 28th February 2006 at http://www.iccrom.org/eng/training/forms.htm

23rd – 27th **10th AHILA (Association for Health Information and Libraries in Africa) conference.** Conference to be convened in Mombasa, Kenya. Contact: Nancy Kamau (Kenya Medical Research Institute) P.O.BOX 54840, Nairobi Tel.: (254)-713678 Fax: (254)-720030 E-mail: kemrilib@healthnet.or.ke Website http://www.ahila.org/events.php

Nov 2006

7th – 10th **Culture, memory and trauma – third annual national oral history conference** Convened by Oral History Association of South Africa in Richards Bay, South Africa. Deadline for proposal submission **15th July 2006**. Send proposals to**:** Third National Oral History Conference c/o Prof Philippe Denis, Sinomlando Centre

for Oral History and Memory Work in Africa University of KwaZulu-Natal Private Bag X01 Scottsville 3209 (South Africa) Phone: 033 260 55 44 E-mail: denis@ukzn.ac.za Website: http://www.national.archives.gov.za/oral_assoc.htm

13th – 17th **Intermediate archives and records management** Training course to be convened by the Centre of Applied Communication at UNISA. Deadline for registration **3rd November 2006.** Contact Mr. Mathews Kokong Tel: (+2711) 471 3896 Fax: (+2711) 471 3906 Email: mkokong@unisa.ac.za Website: http://www.unisa.ac.za/Default.asp?Cmd=ViewContent&ContentID=17563

23rd – 24th **East African Government Technology conference** To be convened in Dar es Salaam, Tanzania. Website: http://new.aitecafrica.com/node/133

26th – 29th **11th ISfTeH International Conference : e-Health for all. Developed and Developing countries. Common issues – Universal solutions?** To be convened by the South African Department of Health and to be held in Cape Town, South Africa. Website: http://www.mrc.ac.za/conference/satelemedicine/index.htm

Feb 2007

Feb-March **ICCROM Course on Conservation of Built Heritage** Application deadline **31st July 2006** to ensure inclusion in our selection process. Contact ICCROM - Sites Unit 13, via di San Michele I-00153 ROME RM, ITALY Tel (+39) 06 585531 Fax (+39) 06 58553349 E-mail: builtheritage07@iccrom.org Website: www.iccrom.org

March 2007

Advanced international records management training program (phase 2) Conducted by the Swedish National Archives. Deadline for application was 1st March 2006 Contact: Riksarkivet C/o Hanna Eriksson Box 12541 SE-102 29 Stockholm, Sweden Fax +4687376474 Phone +4687376350 Website: http://www.ra.se/ra/recordsmanagement/index_eng.asp

July 2007

XIV Bi-Annual ESARBICA Conference to be convened in Tanzania, hosted by the National Archives of Tanzania.

October 2007

Advanced international records management training program (phase 3) Conducted by the Swedish National Archives. Deadline for application was 1st March 2006 Contact: Riksarkivet C/o Hanna Eriksson Box 12541 SE-102 29 Stockholm,

Sweden Fax +4687376474 Phone +4687376350 Website:
http://www.ra.se/ra/recordsmanagement/index_eng.asp

July 2008

XVIth International Congress on Archives to be convened in Kuala Lumpur, hosted by the International Council on Archives Website:
http://www.ica.org/calendrier.php?pcalendrierid=237&plangue=eng

Modular emulation – a long-term preservation strategy undertaken jointly by the National Library and the National Archives of the Netherlands

By Remco Verdegem

Background information

The National Library of the Netherlands

As National Library of the Netherlands, the *Koninklijke Bibliotheek* (KB) has the responsibility to serve as the Dutch deposit library. The aim of the deposit library is to collect, catalogue and preserve all publications appearing in the Netherlands. This includes publications on paper as well as in electronic form. In order to meet the challenge of electronic storage, the KB, together with IBM, has developed a large-scale storage system that meets the specific requirements of long-term preservation: the e-Depot. In addition to long-term storage, permanent access strategies have to be worked out. The KB has established a research and development program to develop innovative tools and procedures to ensure future accessibility of stored digital publications.

The majority of electronic publications in the e-Depot are deposited and stored in Portable Document Format (PDF). For a long time, PDF documents were fixed-format text documents. However, newer versions of PDF can contain all kinds of embedded formats, in-text cross references, forms and even pieces of scripting code that make PDF more and more a dynamic format [Ockerbloom]. In addition to PDF, other formats are accepted for storage in the e-Depot as well as Microsoft Office-documents and image formats.

Since the e-Depot became operational in 2003, another type of digital object has been stored as well: interactive multi-media applications, which are supplied on CD- or DVD-ROM by publishers. These applications are installed on a standardized computer platform, which is called a Reference Work Station. After installation, a snapshot is taken from the hard disk's content by creating a disk-image of the

installed application including the operating system. This image is stored in the e-Depot and can be reinstalled on the RWS at any time.

To retrieve the stored digital objects in the future, the e-Depot has to be extended with new functionality. Using the RWS for the rendering of interactive multi-media applications is just one strategy which will buy us some time to work on another solution. PDF documents and other file formats that are stored are still accessible, but we have to work on rendering strategies to enable future access. The KB aims at keeping publications accessible in the format in which they were delivered to the KB by publishers. Migration may alter the appearance of a publication, which also may change its meaning. Migration does not offer a solution for interactive and complex digital objects either. Therefore, the KB will invest in the development of an emulator for preservation purposes.

The *Nationaal Archief* of the Netherlands

The *Nationaal Archief* of the Netherlands has officially existed in its current form since 4 June 2002. But the national archives service in the Netherlands is now 200 years old. The *Nationaal Archief* is the leading centre for the study of Dutch history and culture at the national level. As the 'national memory', the *Nationaal Archief* manages not only historic government records but also the archives of private individuals who play or have played an important part in the life of the nation. It is the task of the *Nationaal Archief* to gather these archives, to maintain them in good, ordered and accessible condition, and to present them to a broad public.

The *Nationaal Archief* has formulated a number of ambitious activities for the next few years. These are set out in the policy statement *'Geschiedenis binnen handbereik'*. At the heart of our ambitions are efforts to make our collection more visible and accessible, and to provide a broad and varied public with a greater insight into the history of the Netherlands. The most important social issues are promoting a transparent and accountable government and maintaining cultural heritage. As the *Nationaal Archief* gives access to 'authentic' sources, the *Nationaal Archief* is able to address these social issues. The challenges of the digital developments require a new strategy for the future management and use of digital collections. Sustained

accessibility, continuity and linkage of digital sources are important conditions for 'future-proof' public services. Digital cultural heritage and electronic records lend themselves admirably to the purpose of applications which serve these public services. Condition is that the digital records are accessible in a sustainable way.

The *Nationaal Archief* of the Netherlands has the legal task to secure the sustained accessibility of the records of the government. This applies to paper records as well as to electronic records. For paper records good facilities are in place. In order to meet this challenge for electronic records, which the Dutch government is producing increasingly, the *Nationaal Archief* is implementing a digital depot for the storage and continued accessibility of digital records.

Modular emulation as a joint project

While both the KB and *Nationaal Archief* have in the past struggled to fulfil their mandates on long-term preservation, for the KB managing publications in the e-Depot and for the *Nationaal Archief* managing electronic records generated by the Dutch government, it has become increasingly clear there needs to be a systematic assessment of long term preservation strategies. In April 2005 a joint project was started in order to emulate the components of the hardware architecture as individual emulators and interconnecting them in order to create a full emulation process.

In order to understand this ongoing process, a few simple questions could be asked:

a) Why do we need preservation strategies?
b) What are the different preservation strategies?
c) What is emulation?
d) What are the different levels of emulation?
e) What are the different approaches of emulation?

a) Why do we need preservation strategies?

Digital objects are fragile. The debate as to the best means of preserving digital objects over the long term has been underway for many years and will no doubt continue for years to come. Various theoretical solutions have been proposed, and research is carried out worldwide to identify ways in which digital objects can be

authentically maintained while remaining accessible and usable over the long term.

The problem is basically due to the high pace of technological change resulting in obsolescence. There is therefore is no guarantee that current computer environments and digital objects can be read in future

In 1999 Jeff Rothenberg posited “old bit streams never die they just become unreadable”. There is therefore a need to preserve more than just the bit stream. One needs to be able to interpret the bit stream taking into account the object’s characteristics. Since digital objects depend on their environment, the key question therefore is how to carry authentic records through time in a usable and understandable way.

b) What are the different preservation strategies?

The main preservation strategies that have been posited so far include

- Migration
- Use of XML
- Encapsulation
- Virtual machine software, and
- Emulation.

c) What is emulation?

The theory behind emulation is that the only way to ensure the authenticity and integrity of the digital object over the long term is to continue to provide access to it in its original environment. Emulation does not focus on the digital object itself, but on the original environment in which the object is created and rendered. It aims at recreating an environment in which the digital object can be rendered in its authentic form.

d) What are the different levels of emulation?

Emulation can be done at three different levels:

- application software level
- system software (operating system) level and
- hardware level.

Emulating both application and system software requires knowledge of their design and implementation. These products are complex and very often proprietary, which makes it difficult to emulate. Another issue with application-level emulation is that each application requires a specific emulator.

This means the most efficient level of emulation is at the hardware level which is done by mimicking the hardware architecture through software. This is called *software-emulation-of-hardware* or *full emulation*. In this way, computer hardware like a processor is emulated by a software surrogate. In fact, the emulator is placed on top of the host operating system (OS), instead of running directly on hardware. This allows us to take advantage of the services and flexibility offered by the host OS, like the use of its Application Programming Interface (API). An emulator is therefore a software program that runs on a host platform (hardware and OS) and recreates the hardware of the target platform under emulation, as depicted in figure 1.

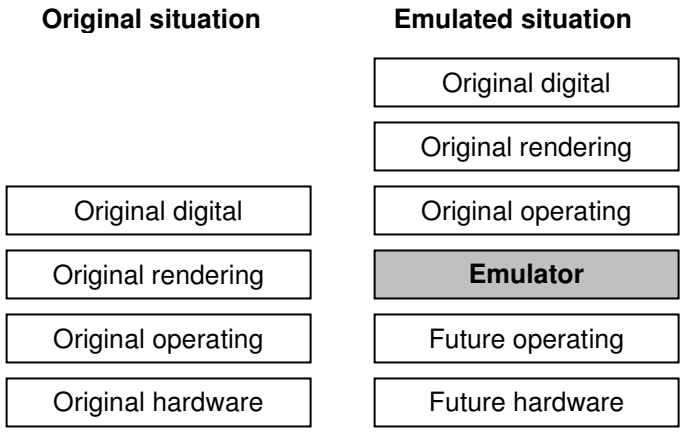

Figure 1: Software emulation of hardware

Although full emulation can also be quite complicated, it has a more straightforward behaviour than emulating higher levels. For example, higher level features such as graphical user interfaces and running multiple applications side by side (known as multi-threading) are difficult to emulate accurately. Emulation of a hardware platform does not incorporate these aspects, but requires the reproduction of the functional

behavior of the original platform in such a way that the original software is not able to distinguish the difference between emulation and reality. Because hardware specifications are well defined and most often available, this behavior is easier to reproduce than that of an OS or software application. Moreover, this approach retains the original OS, applications, drivers and configuration, which secures better authenticity of the original environment.

Since an emulator for digital preservation needs to meet specific needs, there are different approaches to using emulators. According to experts

> the central issue (affecting which approach is taken) is how to ensure that an emulator developed today, can still run in the distant future. Other important aspects concern the accuracy of an emulated platform, such as timing and performance: to be sure an emulator can render digital objects in an authentic way, it has to be tested while the original platform still works.[1]

e) What are the different approaches to emulation?

1) stacked emulation

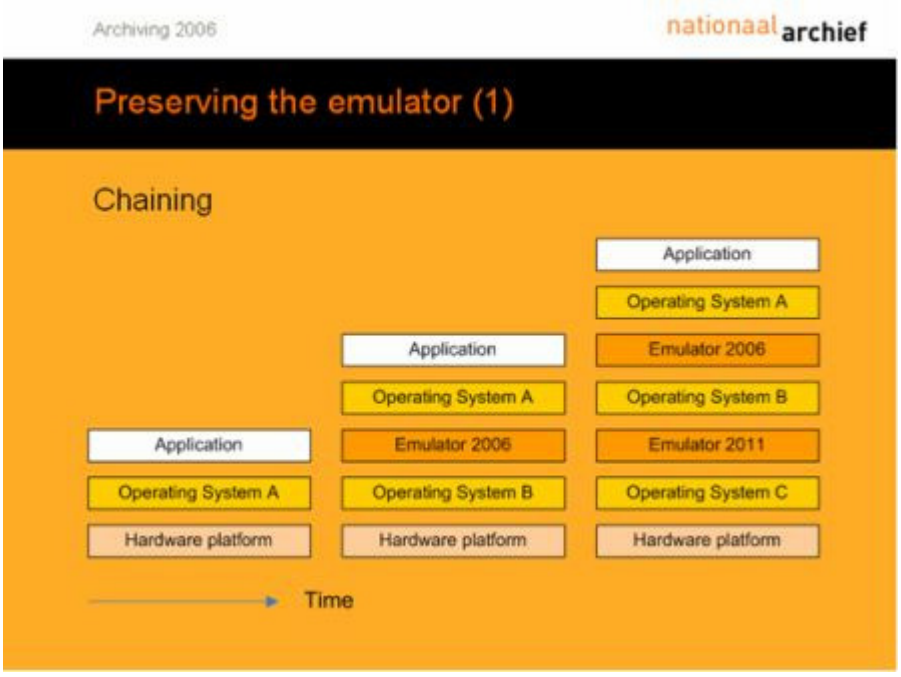

Figure 2: stacked emulation over time

[1] Jeffrey van der Hoeven, Hilde van Wijngarden (2005) "Modular emulation as a long term preservation strategy for digital objects" Paper presented at the 5th International Web Archiving workshop in Vienna, on 23rd September http://www.iwaw.net/05/papers/iwaw05-hoeven.pdf pg 7

With chaining or platform-dependent emulation as shown in figure 2 above, a specific hardware platform is emulated directly on top of a current platform and OS. According to Hoeven and Wijngarden, "the advantage of applying this kind of emulation is the efficiency that can be gained by specifically developing the emulator for one particular host platform. This generally enhances performance and functional behaviour of the emulated platform."[2]

However, using this approach the main challenge is the lack of compatibility with other platforms since the emulator is bound to one particular configuration of the platform. Therefore in order to keep these emulators running on subsequent platforms over time, they have to be stacked or layered (figure 2). With this layering, the risks of errors in the layers of emulators increases exponentially.

2) migrated emulation

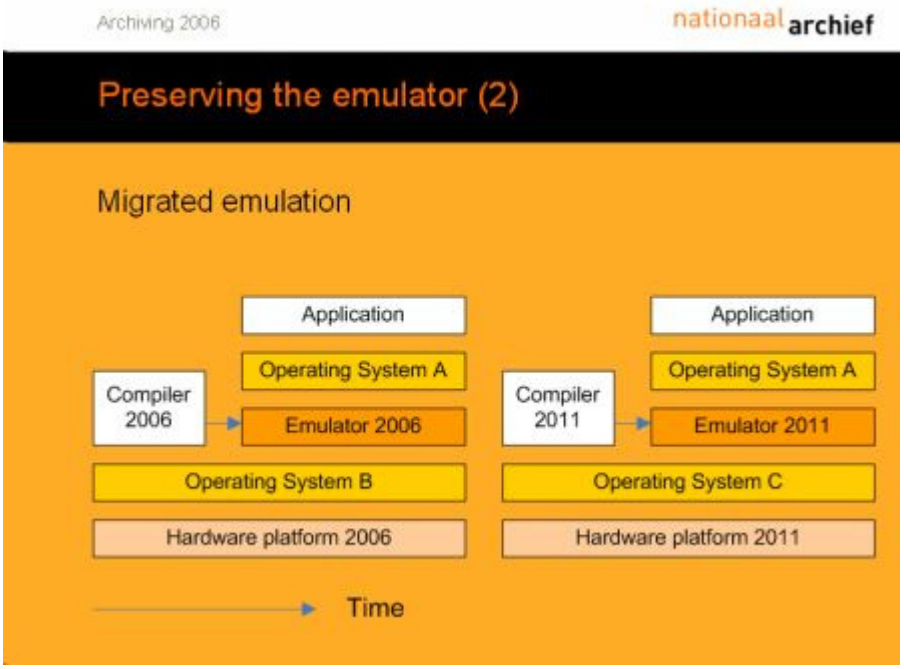

Figure 3: migrated emulation over time

With migrated emulation the emulator is adapted to a different environment. First, an emulator is created for one particular host environment and then if this environment

[2] Ibid

changes, for example, the OS becomes obsolete, the emulator is translated to run on a new host environment.

According to Hoeven and Wijngarden "this translation process requires a compiler that is able to convert the source code written in language X into a binary executable that will work on the new platform."[3] As shown in figure 3, the original emulator is periodically migrated to subsequent environments. However, considering that compiler languages change over time, compatibility is an issue, which could easily lead to malfunctions.

3) emulation virtual machine

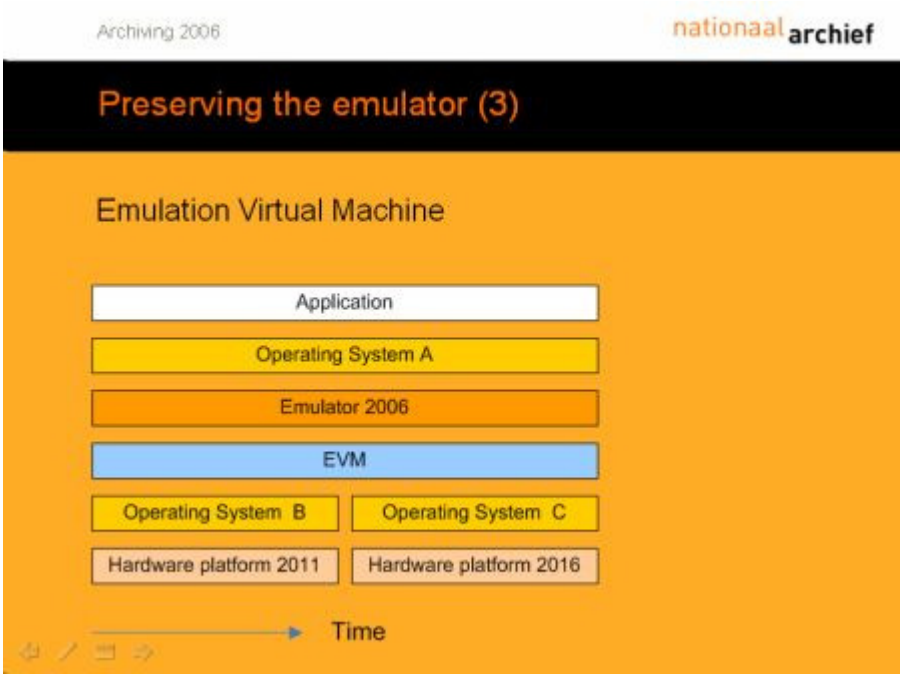

Figure 4: emulation virtual machine over time

Since ideally an emulator should be independent both in time and over platform (figure 4), Jeff Rothenberg has argued that in order to reach this independency one

[3] Ibid pg 8

has to introduce an additional layer between the host platform and emulator, called an Emulation Virtual Machine (EVM).[4]

Hoeven and Wijngarden argue that together with an emulator specification and an emulation interpreter, emulators can be created to run on the EVM.[5] Rothenberg adds that the EVM is stable over time and able to run on various host platforms since it is no longer bound to any particular platform and, it is possible to run multiple emulators on the same EVM.

However, since the EVM has to be maintained over time, it can end up being a very complex process.

An outline of the joint modular emulation project

The *Nationaal Archief* of the Netherlands and the National Library of the Netherlands have started a project to develop a preservation strategy that is based on emulation. The objective of this project is to develop an open source emulator that is modular and portable. Figure 5 shows the conceptual model of modular emulation in the context of digital preservation.

[4] Jeff Rothenburg (2000) "An experiment in using emulation to preserve digital publications". Koninklijke Bibliotheek, The Hague, The Netherlands. Available at http://nedlib.kb.nl/results/emulationpreservationreport.pdf Accessed June 11, 2006

[5] Op cit pg 8

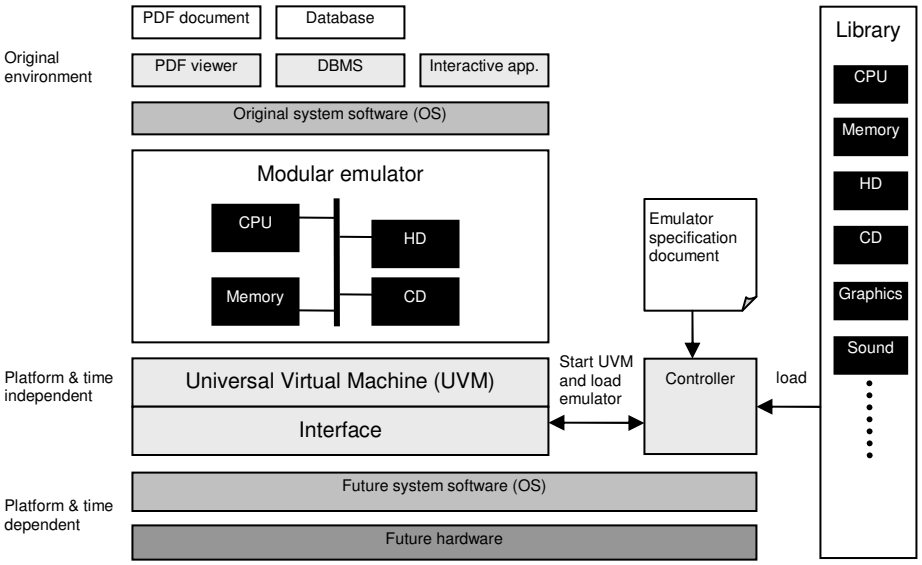

Figure 5: Conceptual model of a modular emulator

The core of this model is defined by the modular emulator that is able to run the original system and application software. The emulator consists of distinct modules, each of them emulating specific hardware functionalities, like a CPU, memory or storage device. Each module can be reused and rearranged to create different emulators. All modules are preserved in a module library. Based on an emulator specification document that defines which modules should be used, the controller loads the required modules and creates a new emulator on the fly. On top of future hardware and software, an EVM-like machine (known as the Universal Virtual Machine or UVM) will create the desired platform independency. In turn, the UVM will execute the modular emulator. The modularity will help to minimize the effort required in emulating a new target machine, as it allows existing modules to be re-used or new modules to be added.

Development of the modular emulator has now started, and the work is scheduled to be completed in April 2007.

The design process for the modular emulator

The design of the emulator is based around a sequence of prototypes, where each subsequent iteration includes more functionality. The final product should provide the

functionality of the modular emulator as described in the User Requirements Document.

This approach was chosen to facilitate design, allow simple implementation to show where bottlenecks lie, and defer research for in-depth knowledge until necessary during component creation.

The table below outlines the sequence in the design phases:

Sequence	Prototype	Details
1.	Registers, memory model, test harness	• 8 general purpose, 6 segment, EFLAGS, EIP • Test segmented memory access
2.	Subset of instruction set/address modes, simple interrupts, clock (?)	• Instruction decoding • Real-mode Effective Address (EA) computation • Simple instructions: Load, Store, Store Immediate, Increment, Conditional-Jump, Loop, subroutine Call & Return, Return from Interrupt (IRET), Interrupt Flag instructions, LEA (Load Effective Address), HALT • Respond to interrupts: Save EFLAGS, EIP, CS registers and optional error code on stack • Log/report attempts to execute instructions or modes that are not implemented yet
3.	Simple Emulation Environment	• Show registers & requested memory locations • Allow typing in register/memory data or loading from files • Allow typing in machine language/ASM or loading from files • Allow running emulator at specified start location • Allow single-stepping • Allow forcing an interrupt request (IRQ)
4.	Minimal machine-language test programs	• Store, Increment, Load • Store Immediate into Segment Register, Store, Increment, Load • Loop • Call in response to interrupt • Return from Call • Use a debugger to try same code on real h/w or validated simulator

Sequence	Prototype	Details
		• Evaluate execution speed
5.	Simplified memory-mapped character display in a Java window	• Hand load characters into memory-mapped area and show that they appear on the display
	Simplified AGP (PCI) for memory-mapped I/O	• Do bus management (if any) for memory-mapped I/O control
	Bit-mapped Java I/O display card emulation, test harness	• Hand load VGA bit-mapped graphics into memory-mapped area, use Java graphics to display the result in a window
	Simplified sound card output, test harness	• Hand load sound-generation data into memory-mapped area and use Java sound to play resulting sound
6.	Minimal graphics/sound machine-language test programs	• Add CPU instructions as needed to run programs that move graphics/sound data to memory mapped areas and cause resulting display/sound output (no I/O port control yet) • Use a debugger to try same code on real h/w or validated simulator • Evaluate execution speed
7.	Port address space, programmed I/O	• Trap I/O ports to call back-end Java device emulators • Keyboard & mouse I/O, test harness • Map I/O port addresses to devices • Implement IN and OUT instructions • Perform keyboard input and mouse sensing • Show mouse-controlled cursor on bit-mapped display • Return mouse-click location at bit-mapped display position • Set interrupt priorities on APIC • Trap and log unrecognized I/O calls
8.	Disk I/O (simple ATA / DMA)	• Cycle clock, instruction cycle emulation • Real-time clock • Set up parameters for DMA transfer via Programmed I/O • Access disk image (stored on real disk) by track/sector • Transfer data between emulated disk and main memory • Signal completion of transfer via IRQ • Implement clock functionality if needed for DMA emulation (e.g., to emulate CPU cycle-stealing)
9.	Expand instruction set, emulation	• Add additional instructions, modes needed for BIOS • Extend emulation environment to read, test files of machine

Sequence	Prototype	Details
	environment	code • Run tests/benchmarks • Evaluate Java speed, graphics, sound, disk access
10.	Implement BIOS	• Including simple I/O, disk-access, boot • Use Bochs / QEMU code if appropriate
11.	Boot emulator	• Make emulation environment load low-memory, BIOS, and initialize VRWS
12.	Boot LILO, DOS and/or minimal Linux in Real Mode	•
13.	Expand emulator as needed for Windows	• Emulate transparent MMU, TLB, etc. to avoid page-faults • Add virtual addressing and other modes • Build test harness and page tables by hand, to test • Run test programs that cause paging
14.	Boot Windows	• Redesign as needed and iterate until Windows runs

For more information

Remco Verdegem, project manager
Nationaal Archief of the Netherlands
remco.verdegem@nationaalarchief.nl

Jeffrey van der Hoeven, co-ordinator Test Team and member of the Development Team
Koninklijke Bibliotheek (National Library of the Netherlands)
jeffrey.vanderhoeven@kb.nl

About the author

After finishing his studies in IT, Remco worked for more than 8 years as an information analyst at a health insurance company. In October 1998, he joined the *Nationaal Archief* of the Netherlands, where he was among others responsible for the functional maintenance of the archival system for paper records. From October 2000 till October 2003 he was working as the project manager for the Digital Preservation Testbed, which was sponsored by the *Nationaal Archief* of the Netherlands and the

Ministry of the Interior and Kingdom Relations. After the project had finished Remco returned to the *Nationaal Archief*, where works as senior advisor Digital Longevity.

Additional information

1. CAMILEON Project – an emulation project conducted by the University of Michigan http://www.si.umich.edu/CAMILEON/
2. Canadian Heritage Information Network's guide to creating and managing digital content http://www.chin.gc.ca/English/Digital_Content/
3. Digital preservation tutorial by Cornell University http://www.library.cornell.edu/iris/tutorial/dpm/
4. International Research on Preservation of Authentic Records in Electronic Systems (InterPARES) project with its headquarters at the University of British Columbia http://www.interpares.org/
5. Wikipedia's entry on Emulation http://en.wikipedia.org/wiki/Emulation
6. Wikipedia's entry on Virtual machine http://en.wikipedia.org/wiki/Virtual_machine