Modular emulation as a long-term preservation strategy for digital objects

Jeffrey van der Hoeven
Digital Preservation Division
Koninklijke Bibliotheek
www.kb.nl
Overview

• The e-Depot

• Digital Preservation

• Emulation

• Conclusions
Mission of the Koninklijke Bibliotheek:

Ensuring permanent availability of information and knowledge

For digital material, it means:

- Long-term preservation and accessibility of electronic publications
- Safeguarding authenticity & integrity
- Continuous effort in Research & Development
The e-Depot

- Electronic version traditional depository
- Developed in collaboration with IBM
- Technical heart: DIAS (OAIS-based)
- Ingest of online journal articles, e-books, and multi-media applications
- Operational since March 17, 2003
- Over 3 million electronic publications ingested
Overview

• The e-Depot

• Digital Preservation

• Emulation

• Conclusions
How does digital feel?

**Functionality**
- Computer file(s)
- Software
- Hardware

**Representation process**
- Offers:
  - structure
  - content
  - context
  - appearance
  - behaviour

**Interactivity**
- User
What do we need?

• Difficulties:
  • Carriers of digitally stored information are vulnerable
  • Software and hardware are required to read the publication
  • Technology is changing rapidly

• To ensure digital preservation, we need:
  - Safe storage
  - Preservation metadata
  - Strategies for access
A fatal exception 0E has occurred at 0167:BFF9DFFF. The current application will be terminated.

* Press any key to terminate the current application.
* Press CTRL+ALT+DEL again to restart your computer. You will lose any unsaved information in all applications.

Press any key to continue
Preservation strategies

Migration
= change the object -> adapt it to a new environment

Risky due to:
- Possible loss of information
- Error propagation
- lack of support for functional behaviour
Preservation strategies

Universal Virtual Computer (UVC)
  = time & platform independent reconstruction of original object

Works well for digital objects...

...but does not offer original functionality.
Preservation strategies

Emulation

= change the environment -> keep the original

- Original object
- Original environment
- Functionality preserved
- No periodic conversions
- No error propagation
Preservation strategies

Emulation -> digital content trapped?
No, just leave object outside disk image.

Electronic publication
Original software
Emulator
Current / future platform
Preservation strategies

Emulation ≠ virtualization
Virtualization is platform dependent (like VMware, MS VirtualPC)

Virtualization

- Platform X
  - Virtual Machine
    - Platform X

Emulation

- Platform X
  - Emulator
    - Platform X
    - Platform Y
    - Platform Z
Overview

- The e-Depot
- Digital Preservation
- Emulation
- Conclusions
Why interesting?

• Retains the original object -> KB focus!

• Solution for complex digital objects:
  • Portable Document Format (PDF)
  • Interactive multi-media applications
  • Scientific data and applications (DARE project)
  • Websites (upcoming web archiving project)

• Indirectly important for migration
The Reference Workstation (RWS) approach

... buys us time, but eventually...

- Multi-media application
- RWS Software
- Emulator
- Future platform

Snapshot of harddisk content

Disk image

e-Depot
Emulation

Difficulties:

- **Costly**: depends on number & type of objects
- **Complex**: accurate, fast, reliable, flexible

... but a lot of work has been done already:

SIMH, Bochs, QEMU, PearPC and more
Emulation project

- Goal: working emulator as substitute for RWS
- Scope: PDF files, databases, multi-media applications
- Together with Nationaal Archief of the Netherlands
- Preliminary research phase finished
- Project started April 2005
Research results

3 theoretical emulation concepts:
1 - Stacked (or layered) emulation

Risk: highly dependent, many emulators needed.
Research results

2 – Migration of emulator (CAMiLEON)

Risk: incompatible, each emulator must be migrated
Research results

3 - Emulation using a Virtual Machine (VM)

Windows 2000
Emulator 1
Virtual Machine
Platform X

Linux
Emulator 2
Virtual Machine
Platform Y

Slightly changes over time, but reliable.
Research results

New approach: modular emulation

Modular emulator

Windows 2000

Module Library

Processor
- x86
- PPC

Memory
- 64MB
- 128MB
- 256MB

Storage
- HD
- CD
- DVD

Emulator specification

CPU
RAM
Graphics
CD-ROM
Research results

Extension: universal virtual platform
Research results

Universal virtual platform <-> controller
Welcome back to earth!

2005

• Testing emulators vs RWS
• DP Testbed

2006

• Start development
• Modular emulator
• Substitute RWS

2007
What more?

- Preservation of environment (OS + apps)
- Documentation (user manuals, technical reports)
- Preservation metadata (management)
Future talk…

2007 2008 2009

tool service network
Overview

• The e-Depot

• Digital Preservation

• Emulation

• Conclusions
Conclusions

Emulation =

...the only way to retain functionality

...proven technology

...new for Digital Preservation

Important: do not re-invent the wheel!
Modular emulation – a viable preservation strategy

KB: www.kb.nl
Contact: Jeffrey.vanderhoeven@kb.nl

Thank you for your attention