



The use of emulation tools as part of a strategy for long-term preservation of digital records



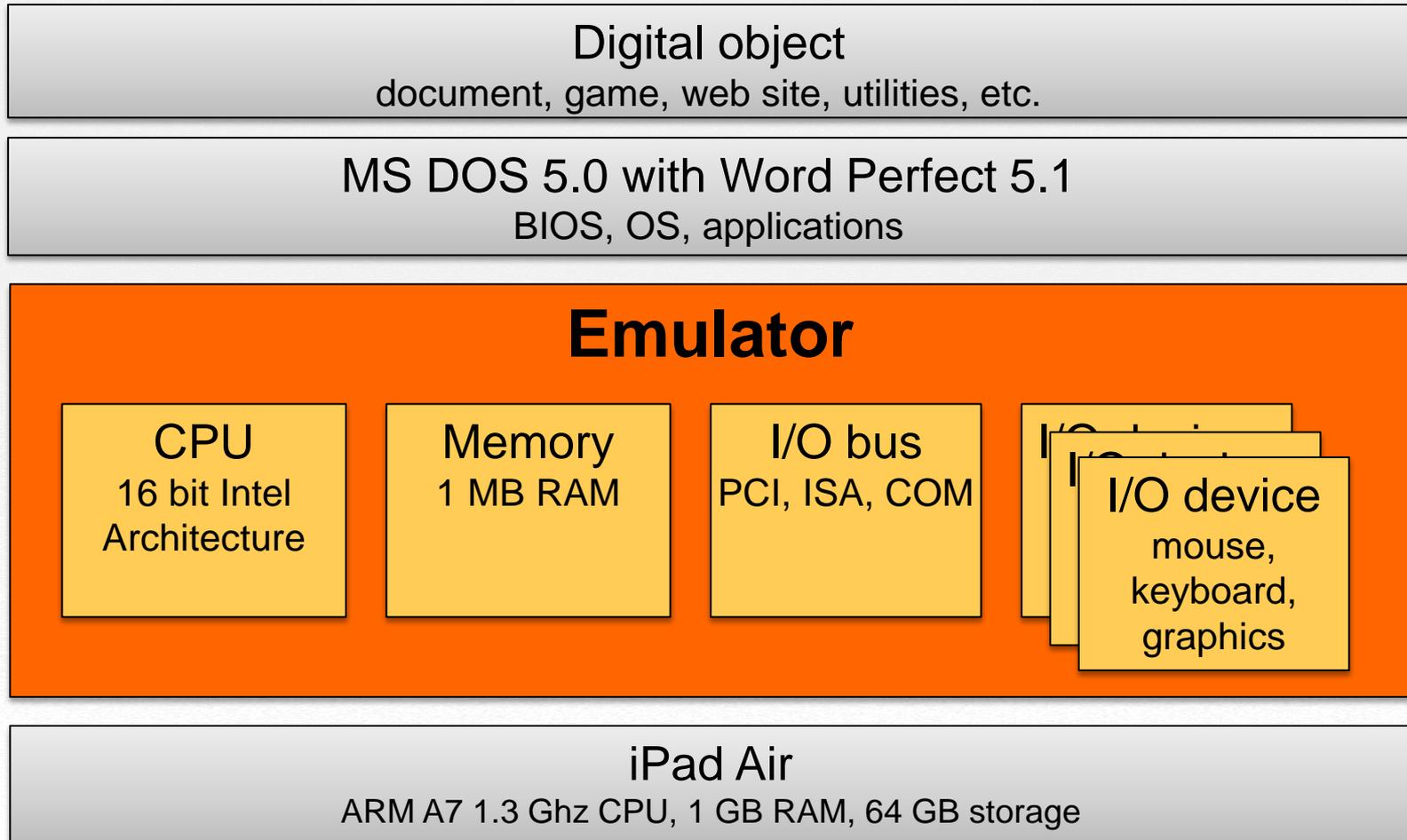
Jeffrey van der Hoeven
3rd LIBER workshop, Vienna, May 20th, 2014



What is emulation?

In a nutshell...

Emulation



Emulation

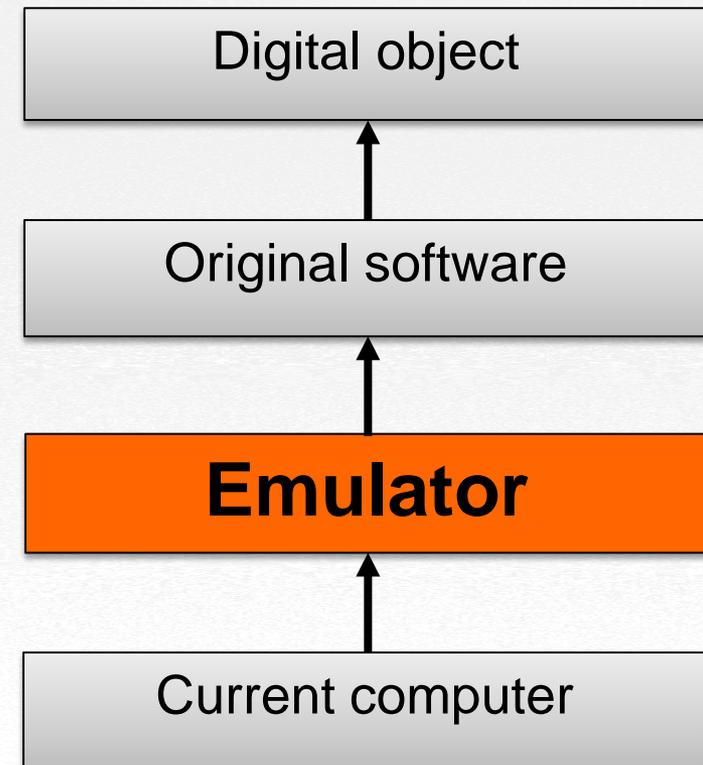
= mimics the original hardware and/or software of a computing system.

Pros:

- Authentic 'look and feel'
- Applicable to any type of object

Cons:

- Requires accurate emulator
- Requires original software
- Requires understanding of the way old computer system and software works
- Copyrights and patents on original software and hardware



Emulation in practice

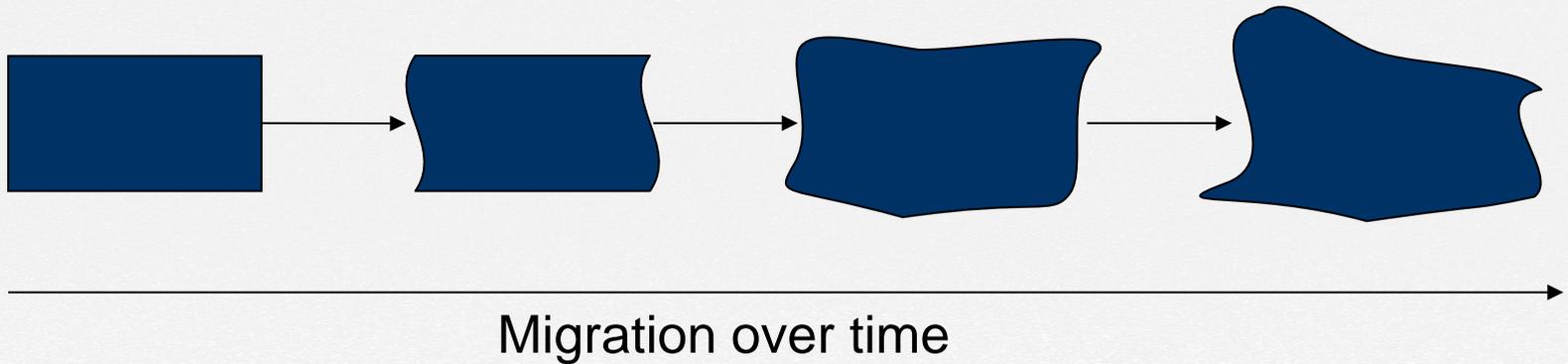
- The concept of emulation dates back to 1941 (to emulate Enigma codes from the Germans by the British)
- Early 90s emulation in context of digital preservation first appeared
- Virtualization = emulation on the same computer architecture
 - Nowadays widely used at server-side computing
- Examples of emulation:
 - Video game console emulators, such as NES, arcade machines
 - Non-HP printers emulating well-known HP-printers
 - DOSbox emulating the MS DOS interface
 - Smartphone/TV/tablet emulators are used to develop and test apps
 - XBOX 360 emulating XBOX (1st gen)

Ref:

<http://kaluszka.com/vt/emulation/history.html>

Alternative: file format migration

= Convert one file format into another.



Pros:

- Easily applicable to text and pictures
- Converted files can be accessed using current software

Cons:

- Chance of losing important information each conversion
- Not applicable to all kinds of data (e.g. software)



Why emulation?

A viable strategy for permanent access

KB's vision:

“We offer everyone everywhere access to everything published in and about the Netherlands.”

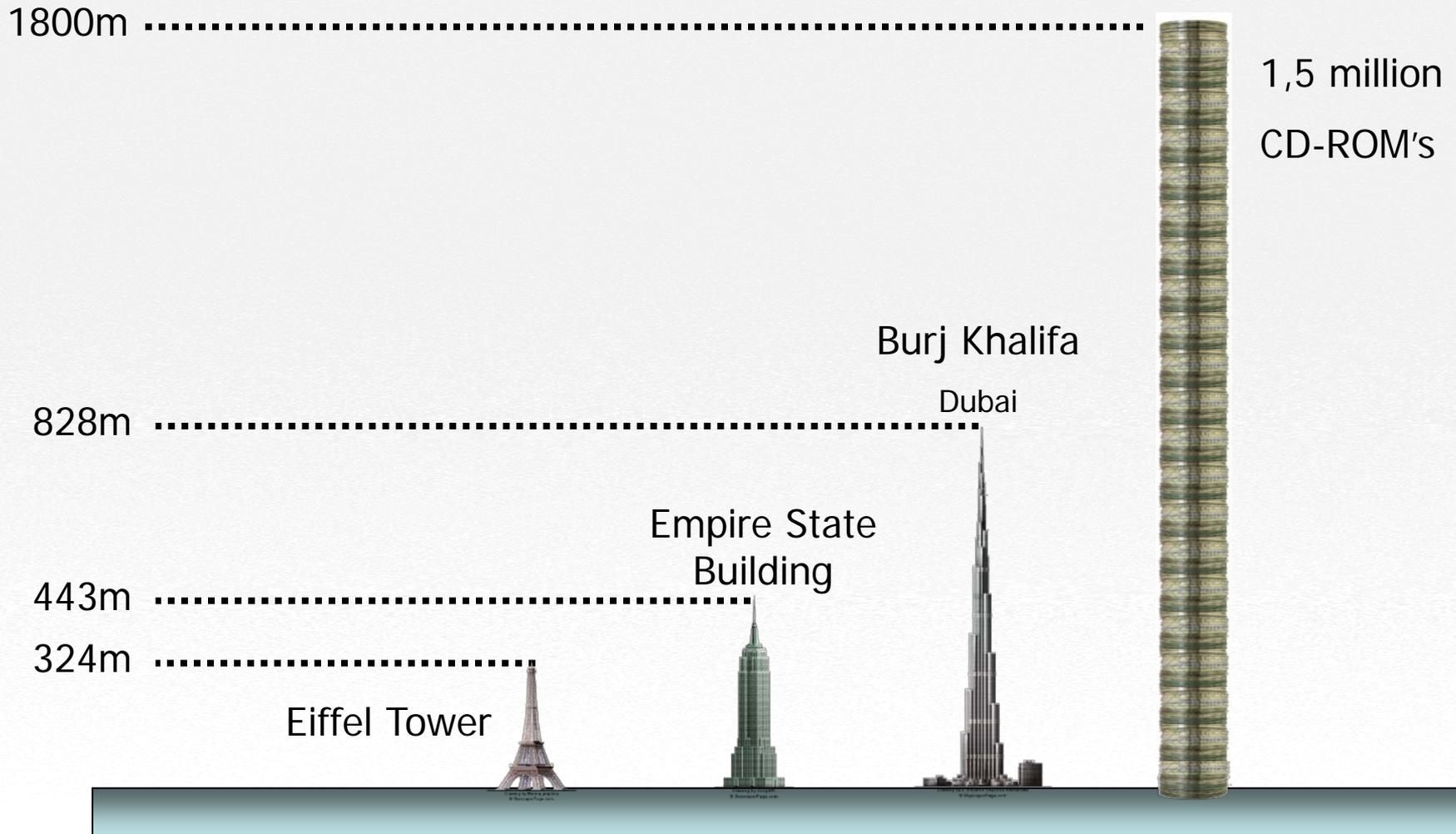
KB's policy plan 2009-2013

What we preserve:

- 6M books, magazines, newspapers
- 1000+ Medieval manuscripts
- 18M Scientific journals
- 450K Parliamentary notes
- 1000+ Websites



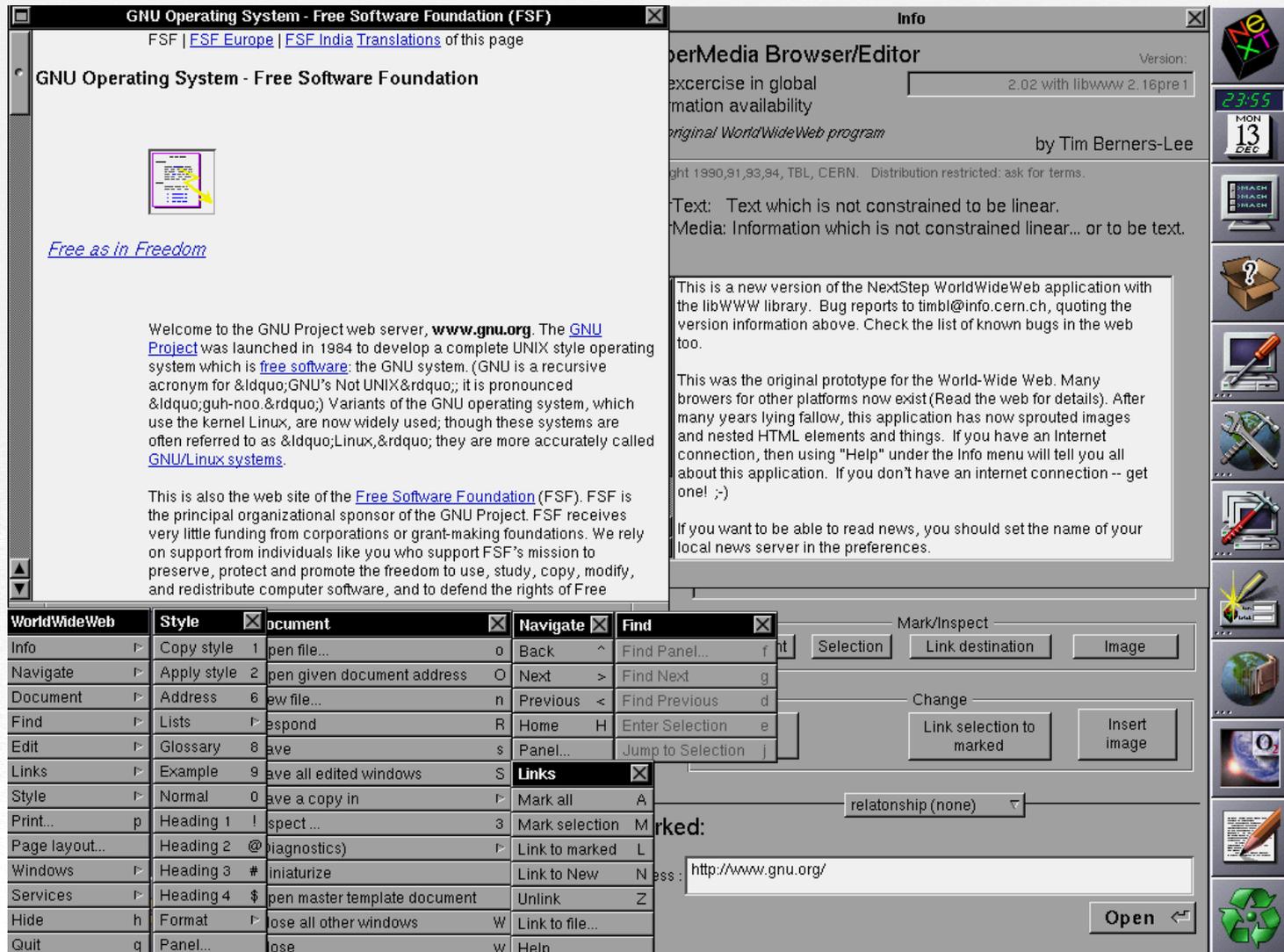
KB data collection: 1 PB in 2018



Collection characteristics

- Increase of volume
- Increase of complexity of unstructured data
- Increase of software dependency
 - for analyses
 - for presentation

WorldWideWeb browser - 1991

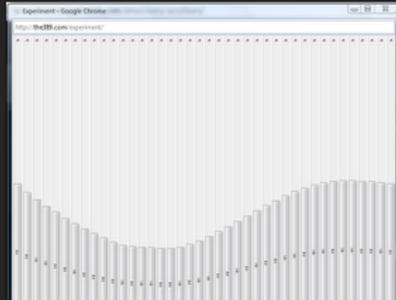


Google Chrome - 2014

Featured Experiments 1-3 of 9 [Prev](#) | [Next](#)



[Canopy](#)
Ryan Alexander
★★★★★



[Wavy Scrollbars](#)
Andrey
★★★★★



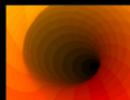
[Ball Pool](#)
Mr.doob
★★★★★

All Experiments

Most Recent 1-15 of 48 [Prev](#) | [Next](#)



[twitterbrowse](#)
tom
★★★★★



[Orange Tunnel](#)
txd
★★★★★



[3D JavaScript with Sandy-HX](#)
Matthew Casperson
★★★★★



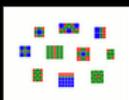
[JavaScript Platformer Demo](#)
Matthew Casperson
★★★★★



[Depth of Field](#)
Mr.doob
★★★★★



[Vectomatic](#)
Lukas Laag
★★★★★



[Roto Game](#)
Felix E. Klee
★★★★★



[JS Fireworks](#)
Kenneth Kuflik
★★★★★



[Google Sphere](#)
Mr.doob
★★★★★



[World of Solitaire](#)
Robert Schultz
★★★★★



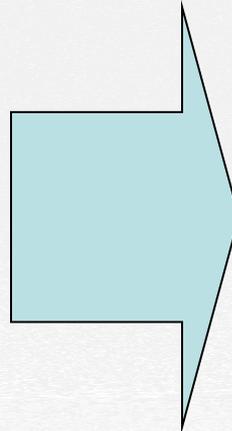
[Canvas Sketch](#)
Henrique Moreira
★★★★★



[HasCanvas](#)
Robert O'Rourke
★★★★★

Functional changes...

- Static web sites (HTML)
- News groups
- Offline working (history & caching)
- Bookmarking
- Little plugins



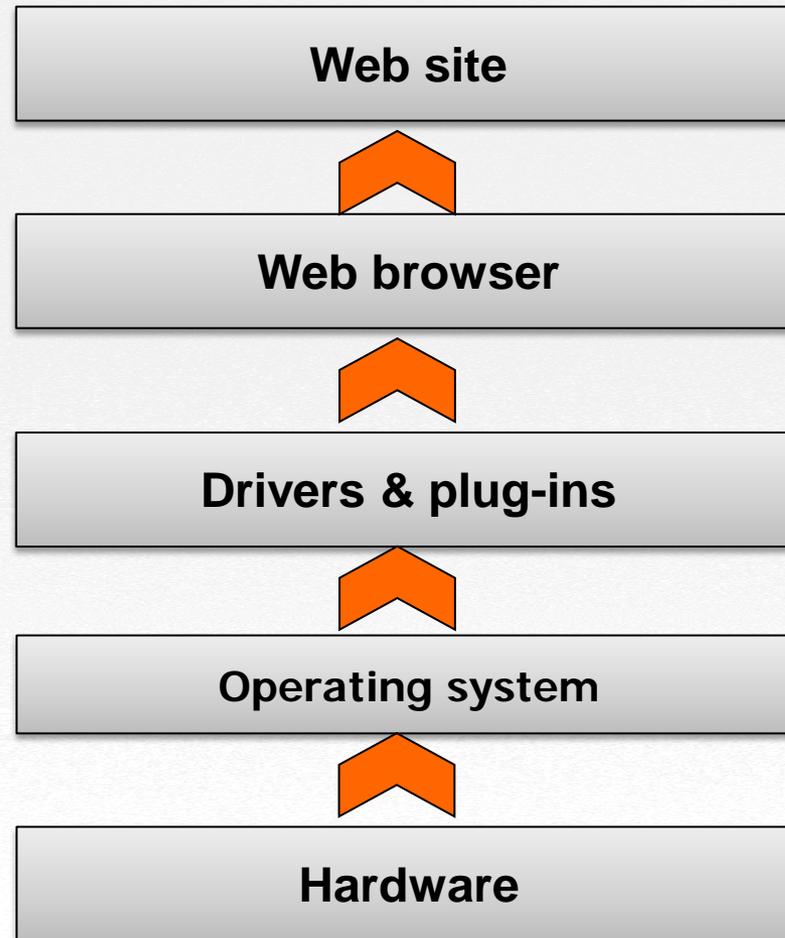
Dynamic web sites (HTML, XHTML, CSS, XML, XSLT, SSL, Javascript, AJAX, .NET)

Personalised content (RSS, RDF, ATOM)

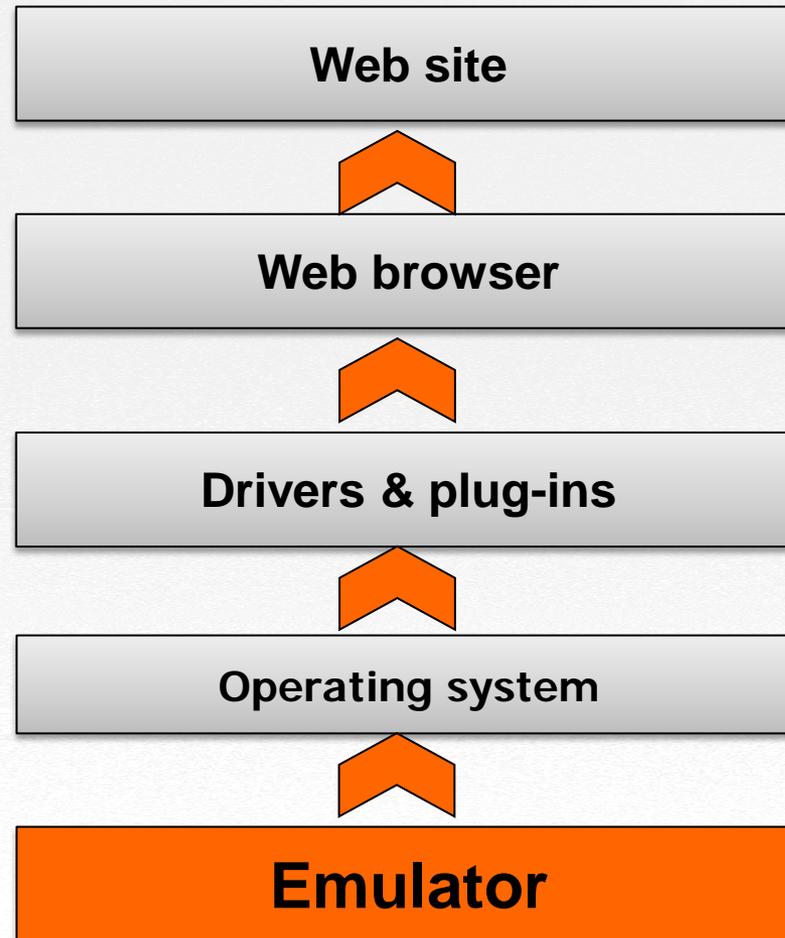
Many plugins (Flash, PDF, ActiveX, Silverlight, 3Dmodelling, video/audio codecs, etc.)

SaaS (Software as a Service)

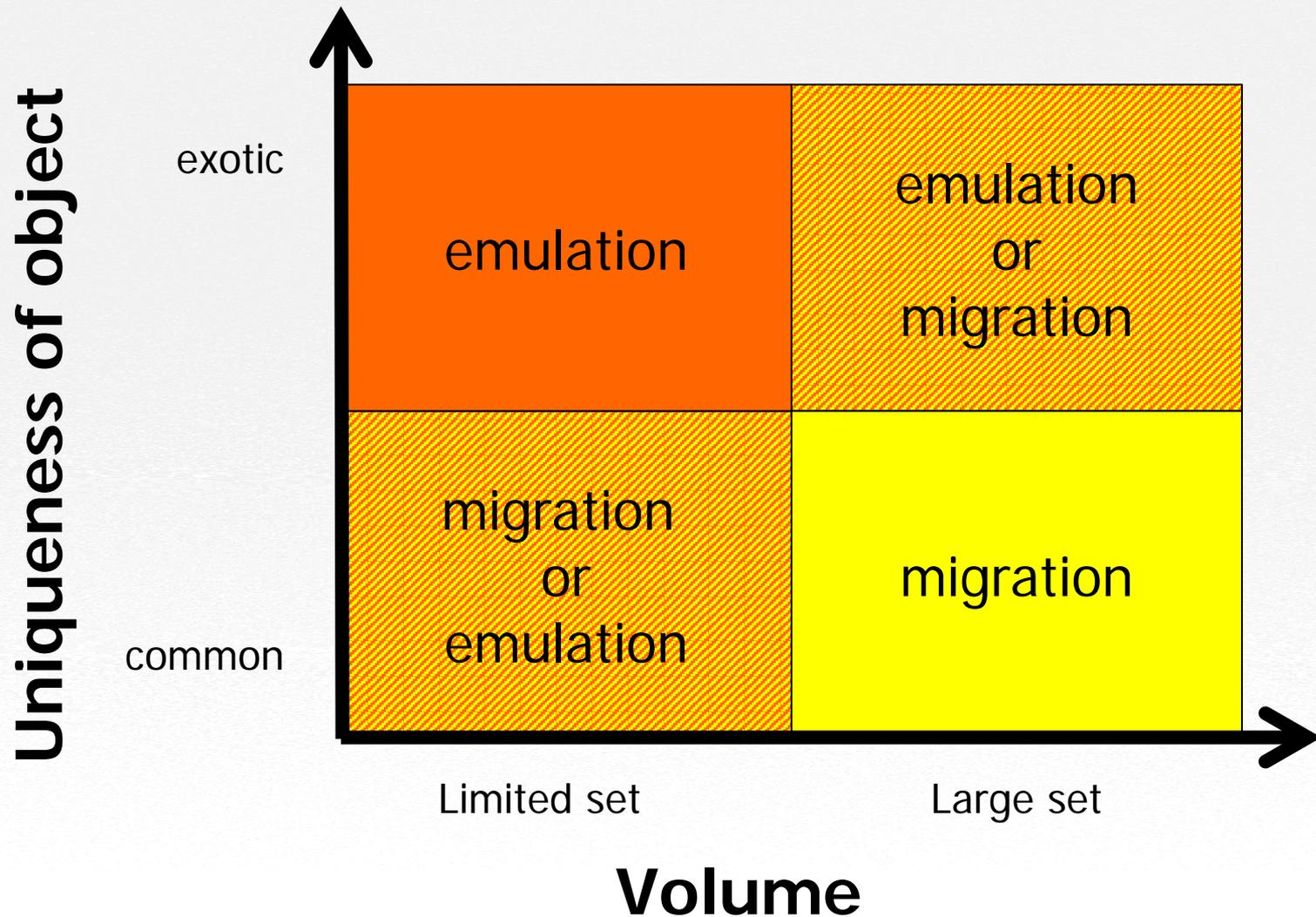
Software & hardware stack



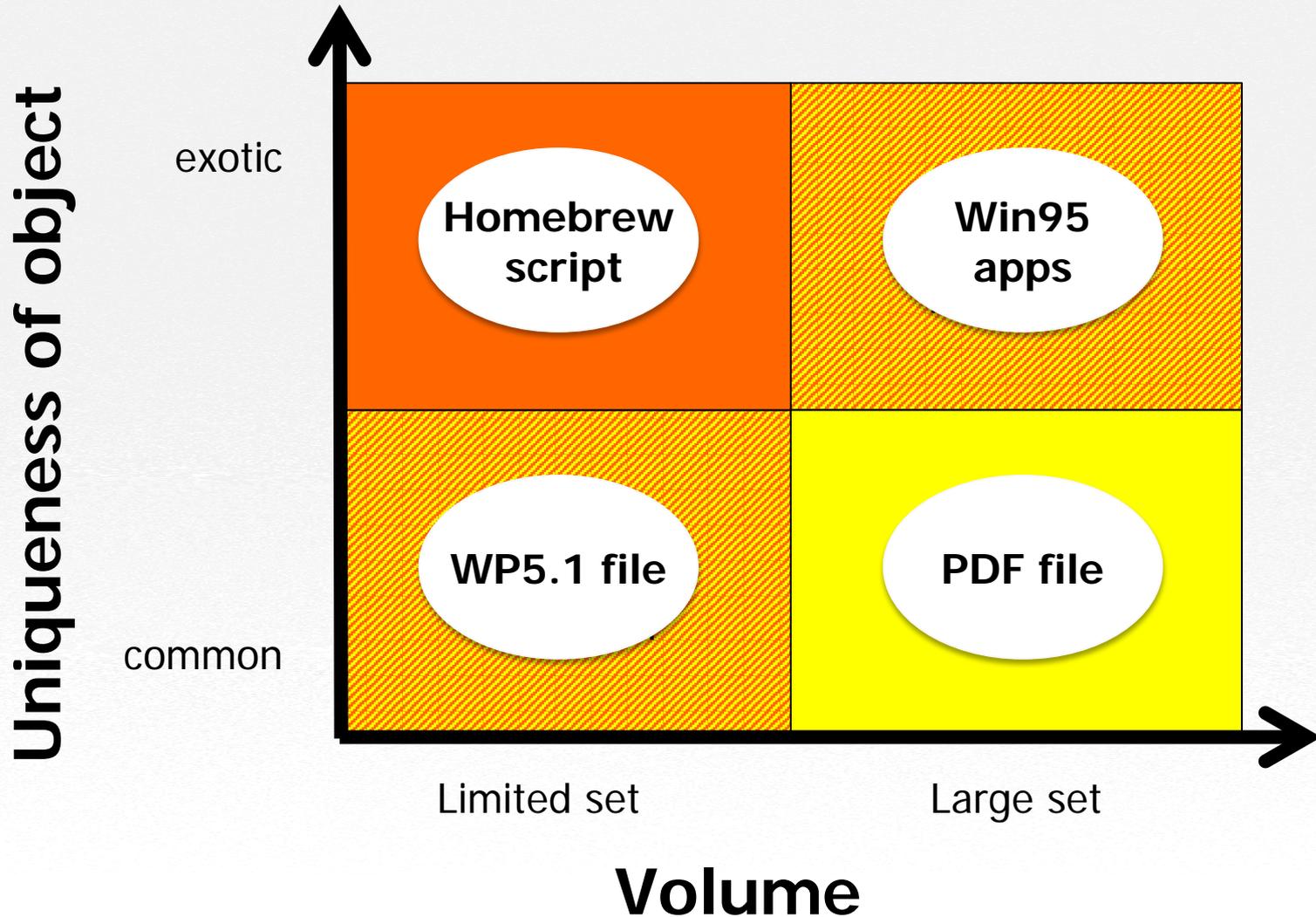
Software & hardware stack



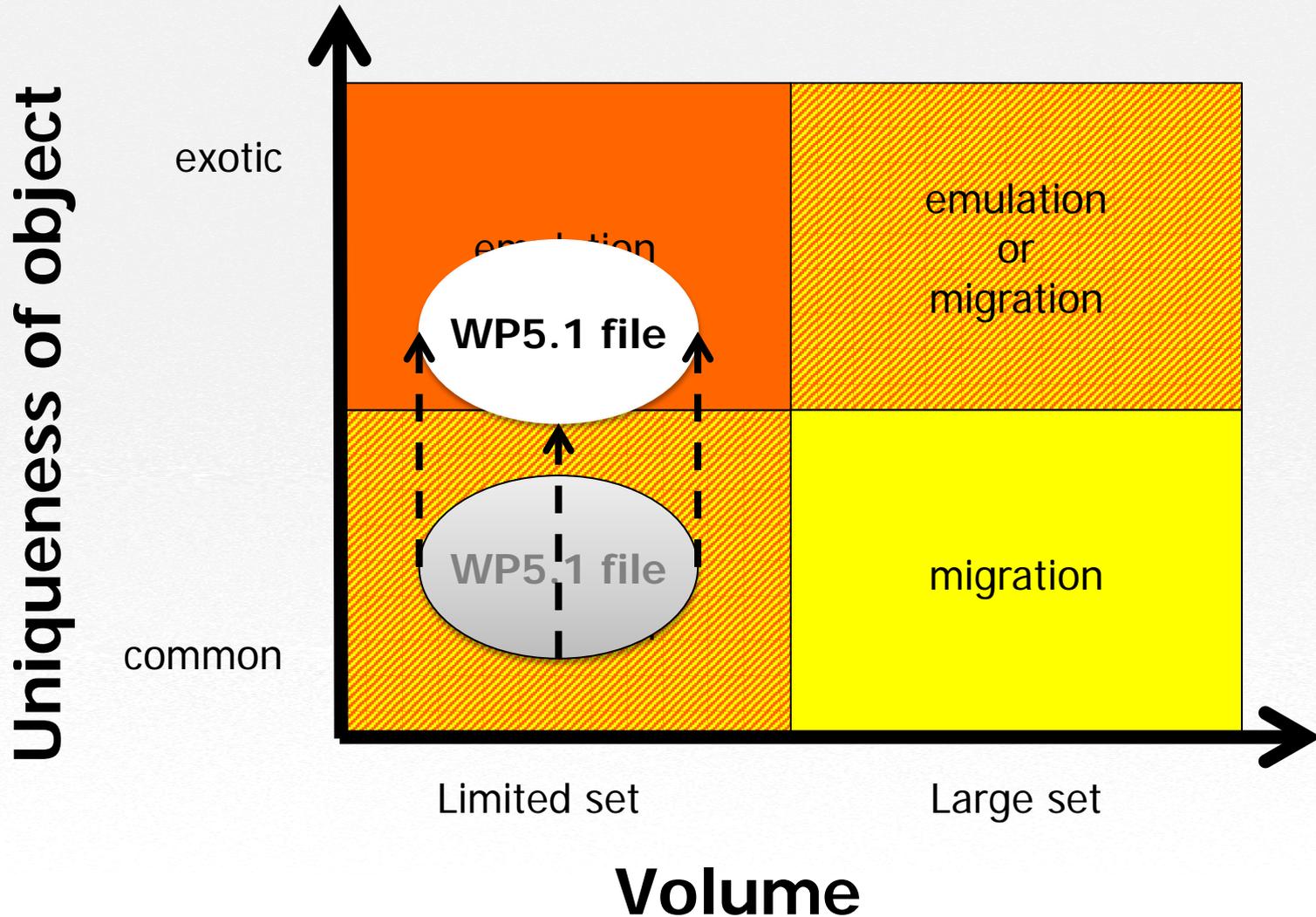
Migration vs emulation



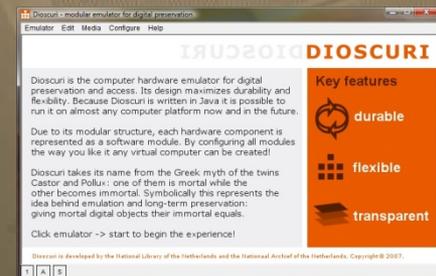
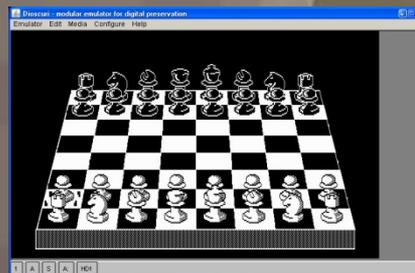
Migration vs emulation (2)



Migration vs emulation (3): over time...



Current state of affairs



KB's take on emulation

- Since 1999 : emulation tests (with Jeff Rothenberg)
- 2002 – 2004 : Universal Virtual Computer (with IBM)
- 2005 – 2007: Dioscuri project (with Nationaal Archief)
- 2008 – 2010: emulation in PLANETS project
- 2009 – 2012: KEEP project on emulation

Emulation Expert Meeting - 2006

- Milestone for Digital Preservation:
 - Brought together emulation experts in 1 room
 - Acknowledged the need for emulation strategies for long-term preservation
 - Set the scene by defining a technical roadmap for the years to come...

More about the EEM 2006, available at:

http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatie-eemstatement-en.html



KB Where are we in 2014...

Nr	Roadmap step	Achievement
1	Create and demonstrate example emulators.	Done! Dioscuri and the like
2	Develop fidelity criteria.	Partly done in PLANETS
3	Develop validation test suites.	Done! PLANETS, KEEP & SCAPE
4	R&D device-independent input/output mechanisms.	Partly done in KEEP
5	Develop methods for capturing and preserving contextual information.	Partly done by Freiburg, PLANETS project
6	Develop methods for describing, managing, and automatically interpreting information about the versions and configurations of software and hardware needed to render digital artifacts under emulation.	Done! In KEEP & University of Freiburg. Should be enhanced.
7	Define and develop a long-lived emulation environment (EVM)	Partly done in KEEP project & IBM's UVC.
8	Develop network-based services for providing remote access.	Done! EF & GRATE

GRATE / bwFLA

Emulation as a Service

Developed by the University of Freiburg (2005-2014)

What is the GRATE / bwFLA solution?

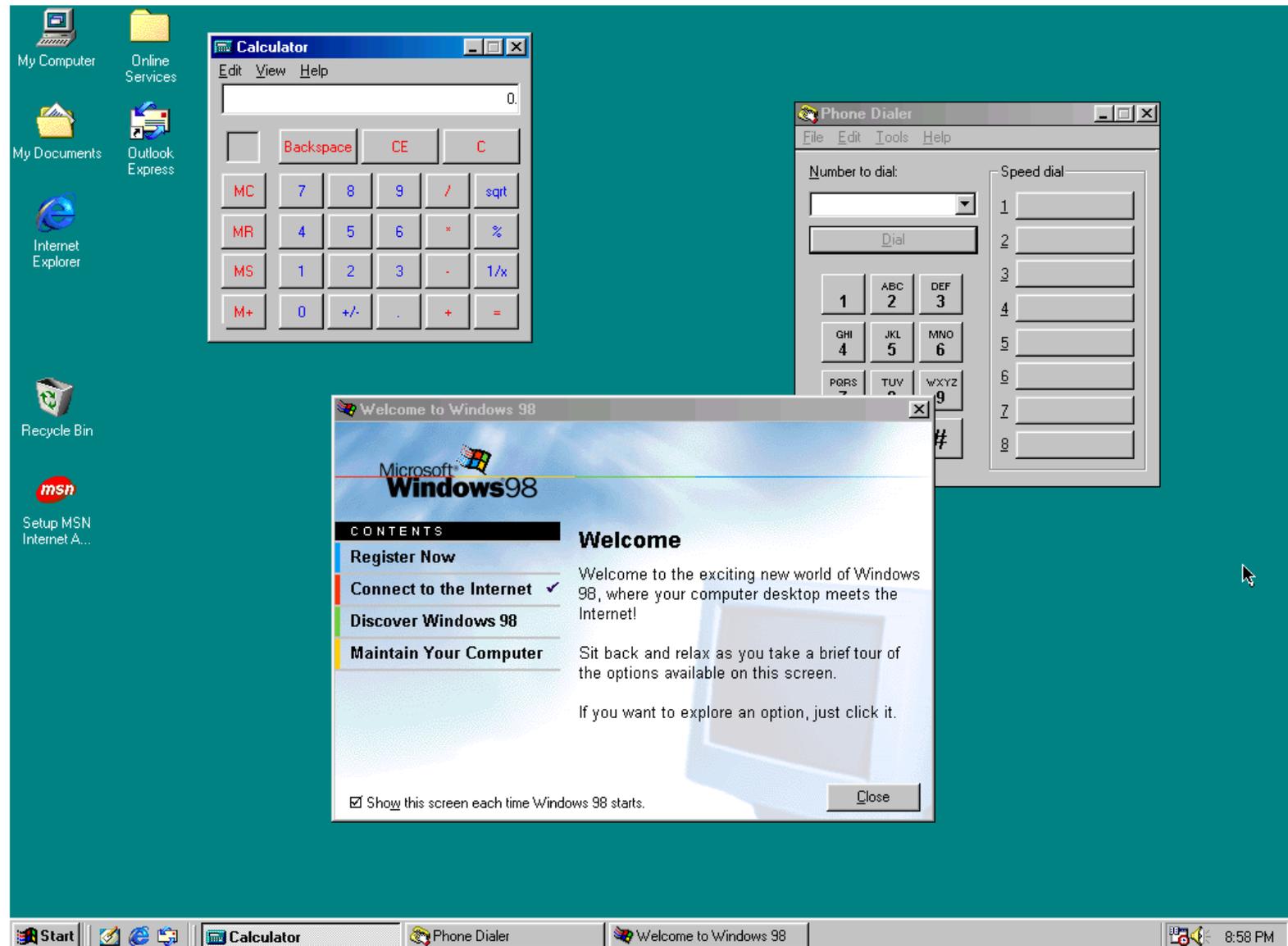
Key features:

- Emulation as a Service (no software installation)
- Computing takes place remotely
- Manages emulation tools
- Automated setup of emulation processes
- Can be integrated with existing archiving solutions



Share

Cite



bwFLA
Legacy Environments at Your Fingertips

Home Emulation as a Service Use Cases / Demos Publications twitter blog

bwFLA — Emulation as a Service

Functional Long-Term Archiving

Digital data is highly volatile, the processes and systems that make use of it even more so. Initially creating data and establishing workflows is very costly, in an ever-changing technological environment the upkeep presents a long-term challenge. To fulfill their designed purpose over a long period of time or for making research results, documents and decisions of today verifiable in the future, efficient and authentic access is required.

In most cases the best way to render a certain digital object is using its creating applications, since these cover most of the object's significant properties, hence providing an authentic and possibly an interactive user experience. Therefore, emulation is a key strategy to provide a digital object's native environment and thus maintain its original **characteristics, look & feel and utility.**

Emulation as a Service

Emulation as a strategy for digital preservation is about to become an accepted technology for memory institutions as a method for coping a large variety of complex digital objects. Hence, the demand for ready-made and especially easy-to-use emulation services will grow. In order to provide user-friendly emulation services a scalable, distributed system model is required to be run on heterogeneous Grid or Cluster infrastructure.

unmodified environment, managed by user

base images, managed by EaaS

Artifacts

Applications

Customization

Operating Systems

Hardware-Emulation

Emulation Component

Ethernet

VNC

RDP

HTML5

WS-API

Emulation Framework

Developed by the KEEP project (2009-2012)

What is the Emulation Framework?



7 emulators

Dioscuri, Qemu, VICE, UAE,
BeebEm, JavaCPC, Thomson

25 file formats

e.g. PDF, TXT, XML, JPG, TIFF, PNG, BMP, Quark, ARJ, EXE,
disk/tape images



6 platforms

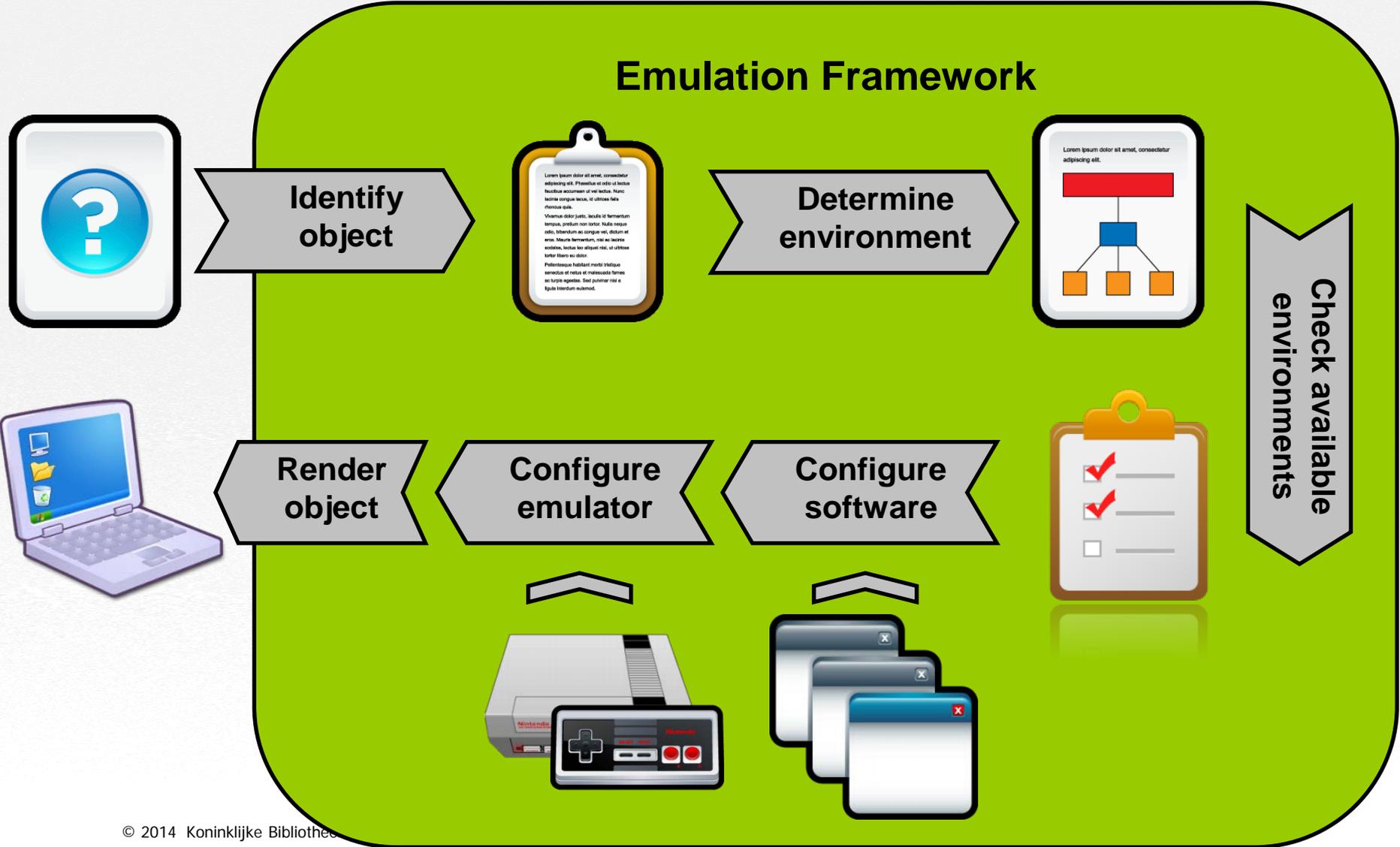
x86, C64, Amiga, BBC Micro,
Amstrad, Thomson T07

What is the Emulation Framework?

Key features:

- Manages emulators
- Decision support for selecting the best emulator
- Partly automated setup of emulation processes
- User guidance when operating old computer environments
- Can be integrated with existing archiving solutions
- Includes a software and emulator archive web service
- Open source (Apache 2.0 license)

Emulation Framework workflow



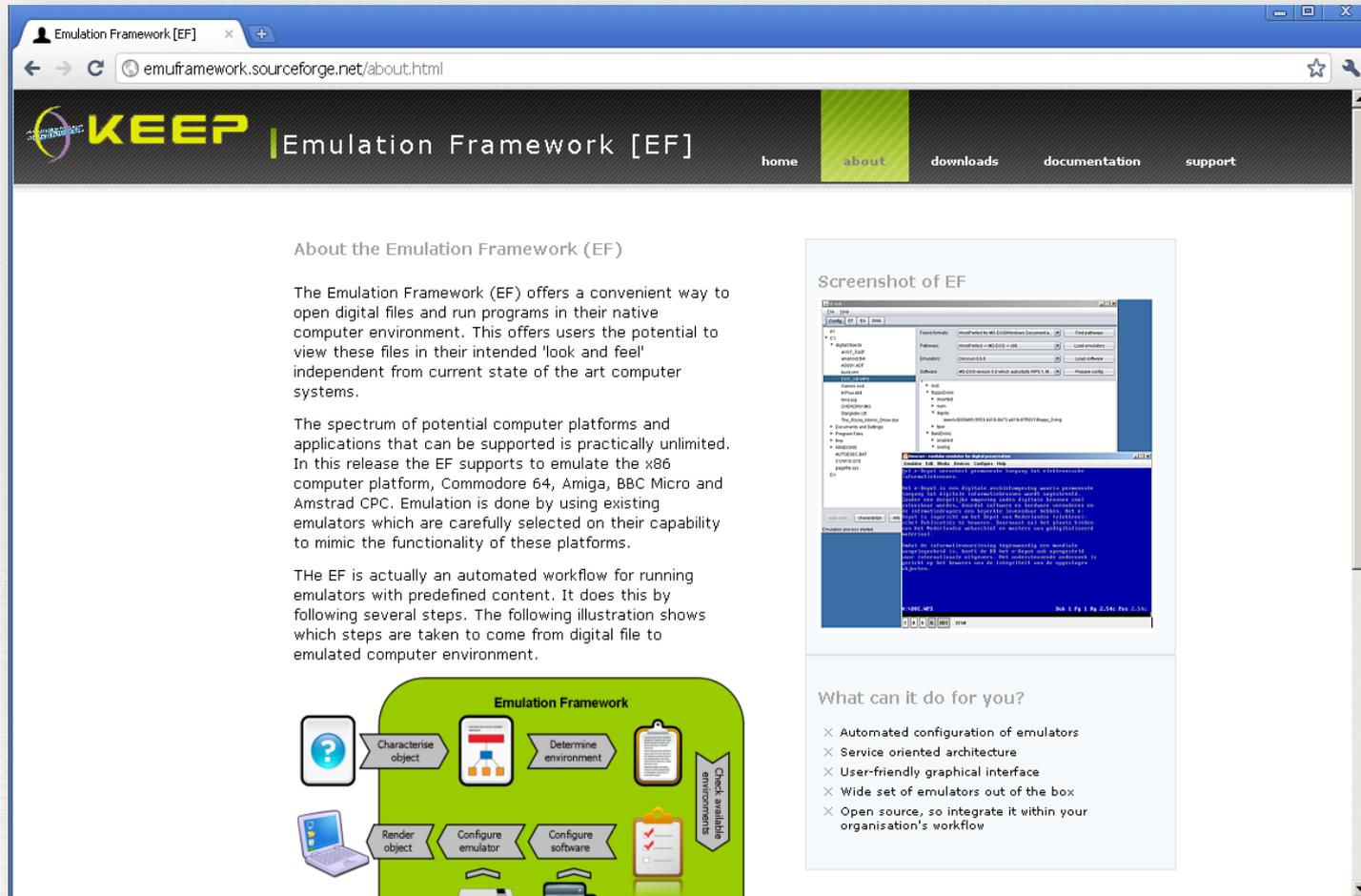
Emulator and Software Archives

**Emulator
Archive service**



**Software
Archive service**





The screenshot shows the website for the Emulation Framework (EF). The browser address bar displays `emuframework.sourceforge.net/about.html`. The website header features the KEEP logo and navigation links for home, about, downloads, documentation, and support. The main content area is titled "About the Emulation Framework (EF)" and contains the following text:

The Emulation Framework (EF) offers a convenient way to open digital files and run programs in their native computer environment. This offers users the potential to view these files in their intended 'look and feel' independent from current state of the art computer systems.

The spectrum of potential computer platforms and applications that can be supported is practically unlimited. In this release the EF supports to emulate the x86 computer platform, Commodore 64, Amiga, BBC Micro and Amstrad CPC. Emulation is done by using existing emulators which are carefully selected on their capability to mimic the functionality of these platforms.

The EF is actually an automated workflow for running emulators with predefined content. It does this by following several steps. The following illustration shows which steps are taken to come from digital file to emulated computer environment.

The illustration shows a workflow diagram for the Emulation Framework. It starts with a question mark icon labeled "Characterise object", followed by "Determine environment", "Render object", "Configure emulator", and "Configure software". A vertical arrow on the right side of the diagram is labeled "Check available environments".

Next to the text is a "Screenshot of EF" showing a graphical user interface for configuring an emulator. It includes fields for "Emulator" (set to "CPC6400"), "Software" (set to "BBC BASIC v1.10"), and "Path". Below the configuration panel is a terminal window displaying the output of the emulation process.

Below the screenshot is a section titled "What can it do for you?" with a list of features:

- × Automated configuration of emulators
- × Service oriented architecture
- × User-friendly graphical interface
- × Wide set of emulators out of the box
- × Open source, so integrate it within your organisation's workflow

The EF is free and open source (Apache 2.0 license).

Why isn't emulation used in preservation?

middelijc wt sijns vaders stat. en hi dede veel
 scepen maken en daer mede voer hi te creten
 en in lythouien En al dat lat torch hi mit sub
 tijlheit aen hē en hier nae wan hi Cecilien En
 corts daer na voer hi door Italien en daer ont
 finc hē die stad van Romem eerlich ende wel.
 Want sij lenden hē heymelic in sijn ghemet:
 daer alle die stat vā Romem te samen an stont
 en die Romeynen brochten alexander en sijn
 die kroone ghemet van sijnen vater
 Alexander onsch dese kroone herde gherne
 en hi onthaelde Elminisse harde wel mit veel
 schoonten talen die hi wou conste: en daer nae
 sijn hē sijn vater van sijnen vater
 sijnen vater Romeynen alexander te
 vater en sijn vater sijnen vater
 voer alexander te Afrikaen waert En sijn
 sonder eenige wederstoet mit al sijn heer alle
 lybien door: en hi quam aldus mit alle sijnen
 volcke in Egipten mit scepen te water ende te
 lande Doe hi in Egipten quam soe ontfinghen
 hē eerlic en sijn hilden hem voor haren heere

Hier na volcht hoe Alexander Cyrien en
 Cyrien belejde en hoe dat hi se te samen ver
 wan:
 Capittel xiiij.

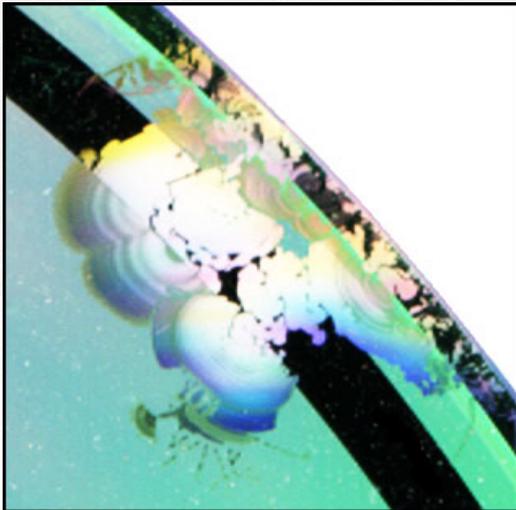


Alexander sach in Egipten staen een beel
 de ghemaect van swarten marbersteene
 Doe vraechde hi wat dat beelde beduyde Dat
 beelde antwoorde hem ende seide dattet was
 Reptanabus die wel eer die vā Perlen stulo
 gen was wt Egipten en liet sijn rijk als hem
 Othus die coninc vā Perlen vdeef Want die
 aventure liep Reptanabus soe teghens c

Various reasons:

- GLAMs aim for low-hanging fruit: mass digitization!
- We just don't know the size of the problem:
 - unknown number of obscure files and applications reside on ancient media carriers
 - the old generation forgot...
 - the new generation doesn't know there was a pre-internet computer period...
 - Technological quicksand after all!
- Copyright & patents infringement

But the longer
we wait...



Does it have a touch screen?



My colors are gone!?

```

H:INTRO PAGE 1 LINE 9 COL 11          INSERT ON
      <<<      M A I N      M E N U      >>>
--Cursor Movement--      | -Delete-      | -Miscellaneous-      | -Other Menus-
^S char left ^D char right | ^G char      | ^I Tab      ^B Reform | (from Main only)
^A word left ^F word right | DEL chr lf  | ^V INSERT ON/OFF      | ^J Help  ^K Block
^E line up   ^X line down  | ^T word rt  | ^L Find/Replce again | ^Q Quick ^P Print
      --Scrolling--      | ^Y line      | RETURN End paragraph | ^O Onscreen
^Z line down ^W line up    |              | ^N Insert a RETURN    |
^C screen up ^R screen down|              | ^U Stop a command     |
!-----!-----!-----!-----!-----!-----!-----!-----!-----R

```

1. Introducing WordStar

WordStar is highly flexible and very visible. Watch the screens as you give commands, and information in various parts of the screen will guide you. You won't see all the information all the time, but it will be there when you need it.

WHERE YOU ARE

The seven WordStar menus are your greatest aids. They are like signposts at the top of your screen, showing you where you are.

```

1HELP 2INDENT 3SET LM 4SET RM 5UNDLIN 6BLDFCE 7BEGBLK 8ENDBLK 9BEGFIL 10ENDFIL

```

Emulation to the rescue!



Analyse your digital collection for rare formats, software & old media.



Define the value of it.



Get valuable data from old media carriers & start collecting software for it.



Invest in an emulation service or tool. Do not re-invent the wheel but use existing solutions.

Windows

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

Thank you!



Contact:

www.kb.nl

Jeffrey DOT vanderhoeven AT kb DOT nl

