TOWARDS A UNIVERSAL VIRTUAL INTERACTOR (UVI) FOR DIGITAL OBJECTS

Yale University Library
Euan Cochrane (P)
Seth Anderson
Ethan Gates

Open SLX GmbH
Klaus Rechert (P)

Educopia Institute
Jessica Meyerson
"I know you were for me"
So they found a small cafe
In a cobbled street
He told her that his name was Fred
She said, "mine's Marguerite"
She said, "the memory of our day"
"Is one I'll always keep"
And after taking bread and wine
The sparrows fell asleep
He woke up to the sound of bells
And said, "those bells I know"
"Those ain't the bells of Notre Dame"
"Them there's the Bells of Bow"
It's just another sparrow's dream
Above a London street
But I'll remember all my life
A bird called Marguerite."
Overview

- Demo
- Background/History
- EaaSI
- UVI
they [future generations] should be able to generate an emulator to run the original software that will display my document.

Users need the option to interact with objects in original software

Original content in original software (WordPerfect in Windows 95)

Original content in newer software (LibreOffice Writer in Windows Vista)
Research results are at risk of loss

Original content in original software (WordStar for DOS in Microsoft DOS)

[NB: equation predicting tree growth rates includes exponents documented using upper line of text]

Original content in newer software (LibreOffice Writer in Windows Vista)

[NB: equation layout and meaning changed]
The Universal Virtual Computer (UVC)
Dioscuri is the computer hardware emulator for digital preservation and access. Its design maximizes durability and flexibility. Because Dioscuri is written in Java it is possible to run it on almost any computer platform now and in the future.

Due to its modular structure, each hardware component is represented as a software module. By configuring all modules the way you like it any virtual computer can be created!

Dioscuri takes its name from the Greek myth of the twins Castor and Pollux: one of them is mortal while the other becomes immortal. Symbolically this represents the idea behind emulation and long-term preservation: giving mortal digital objects their immortal equals.

Click emulator -> start to begin the experience!
Dioscuri Modular Emulator Design
Keeping Emulation Environments Portable (KEEP)

1. When a user requests an item from a digital collection and this item requires an old computer environment to render, the Emulation Framework is used.

2. The Emulation Framework automatically selects and runs the best available emulator and configures the software dependencies required to render the object (operating system, applications, etc.).

3. A virtual screen shows the digital item with original software of that time.
Keep limitations

- Each environment had to be configured and documented in each installation
- Few view-paths available out of the box
- Desktop-only software
bwFLA Emulation as a Service (EaaS)
What is Emulation-as-a-Service?
Simplifies access to various emulators
Enables management of persistent, citable emulation environments
Derivatives enable storage savings

- "Base" Environment (e.g. Windows XP)
  - 5 GB (Compressed)

- Derivative x1 (e.g. Windows XP + SPSS 13)
  - 150MB

- Derivative x2 (e.g. Windows XP + STATA 8.2)
  - 250MB

- Derivative x3 (e.g. Windows XP + R 2.0.0)
  - 50MB

- Derivative x1.1 (e.g. Windows XP + SPSS 13 + SPSS code)
  - 1MB

- Derivative x3 (e.g. Windows XP + STATA 8.2 + STATA code)
  - 1MB
Project Goal

Deploy and scale infrastructure and services for software emulation, including distributed management, sharing, documentation/discovery, and access.
EaaSI - Updated UI Wireframe
Distributed Mgmt

- A network of distributed nodes, each contributing to the EaaS/I service and the software development roadmap.
Sharing

- In-network sharing of software images and configured environments.
- Yale University Library is configuring and sharing at least 3000 pre-configured software applications running in configured environments.
Documentation/Discovery

- Incorporating services developed by Wikidata for Digital Preservation
- Comprehensive, open, machine-readable documentation
- Defining profile for description of software and computer environments
Corel Presentations
11 “Open” file operation formats:

Access

- Emulated CD-ROM environment sharing service
- Virtual Reading Rooms Service
- Scientific Software Portal
- API to automatically render objects in original software via emulation
Access

- Emulated CD-ROM environment sharing service
- Virtual Reading Rooms Service
- Scientific Software Portal
- API to automatically render objects in original software via emulation
Universal

- it is intended to be able to be "universal" and (theoretically) work with any files/digital objects.

Virtual

- A homage to the Universal Virtual Computer (UVC) concept developed by IBM and the Koninklijke Bibliotheek, KB

Interactor

- Rendering and viewing are primarily passive activities but digital object experiences are not passive, they’re interactive
UVI Overview

- Upload file
- Identify age and format family
- Extract any useful metadata
- Match to environments and rank “best match” or a hex editor
- Autostart rendering object
Identification

- Identify software to render file
  - Siegfried / Droid

- Use file creation/modification date to determine software version

- Better tool support is required to match software (and version)

- Metadata, describing software capabilities
Corel Presentations

11 “Open” file operation formats:

<table>
<thead>
<tr>
<th>Default Software Environment Name</th>
<th>Executive Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adobe Dynamic Media &amp; Web Solutions</strong></td>
<td>C:\Program Files\Adobe Acrobat 4.0\Reader\English (\backslash) 1.0.0.6 English (us)</td>
</tr>
<tr>
<td><strong>Adobe Illustrator 8.0</strong></td>
<td>C:\Program Files\Adobe Illustrator 8.0</td>
</tr>
<tr>
<td><strong>Adobe Photoshop 6.0</strong></td>
<td>C:\Program Files\Adobe Photoshop 6.0</td>
</tr>
<tr>
<td><strong>CorelDRAW 3.0</strong></td>
<td>C:\Program Files\CorelDRAW 3.0</td>
</tr>
<tr>
<td><strong>Microsoft Office 2000</strong></td>
<td>C:\Program Files\Microsoft Office 2000</td>
</tr>
<tr>
<td><strong>Microsoft Internet Explorer 4.0</strong></td>
<td>C:\Program Files\Microsoft Internet Explorer 4.0</td>
</tr>
<tr>
<td><strong>Microsoft Outlook 98</strong></td>
<td>C:\Program Files\Microsoft Outlook 98</td>
</tr>
<tr>
<td><strong>Microsoft Word 97</strong></td>
<td>C:\Program Files\Microsoft Word 97</td>
</tr>
<tr>
<td><strong>Microsoft Excel 97</strong></td>
<td>C:\Program Files\Microsoft Excel 97</td>
</tr>
<tr>
<td><strong>Microsoft PowerPoint 97</strong></td>
<td>C:\Program Files\Microsoft PowerPoint 97</td>
</tr>
<tr>
<td><strong>Microsoft Access 97</strong></td>
<td>C:\Program Files\Microsoft Access 97</td>
</tr>
<tr>
<td><strong>Microsoft FrontPage 98</strong></td>
<td>C:\Program Files\Microsoft FrontPage 98</td>
</tr>
<tr>
<td><strong>Microsoft Publisher 98</strong></td>
<td>C:\Program Files\Microsoft Publisher 98</td>
</tr>
<tr>
<td><strong>Microsoft Works 98</strong></td>
<td>C:\Program Files\Microsoft Works 98</td>
</tr>
</tbody>
</table>

*Note: The above table provides a sample of software environments and their default locations.*
More data enables more accurate matching
Object Preparation

- Choose appropriate media type based on
  - Environment and OS metadata
  - Supported media types
  - Supported filesystems
  - Read-only rendering or writeable
Auto-opening after system boot

- Scale
  - Automate rendering as much as possible
  - Keep environments generic (unmodified)
    - Multipurpose environments
    - Reduce maintenance
  - Automation included with user object
Future developments

- Improve automation
- Automate and document SW-related tasks
  - Preserve usage knowledge
  - Automate common tasks
- Work with complex (multi – FFMT data sets)
  - Complex SW setups (RDM)
  - Processing Pipelines
Our Team

- Euan Cochrane Principal Investigator
- Seth Anderson Program Manager
- Ethan Gates Software Preservation Analyst
- Klaus Rechert & Oleg Stobbe (OpenSLX) Technical Architecture and Development
- PortalMedia UX/UI Development
- Jessica Meyerson (Educopia/SPN) Communications/Outreach
- Kat Thornton (Data Current/WikiDP) Semantic Architect
Fair use applies to institutions making software available on a cooperative basis to broaden research opportunities, including off-premises access using technology such as Emulation as a Service...

– ARL Code of Best Practices in Fair Use for Software Preservation
A Very Special Thanks to our Funders...
Thank you

https://www.softwarepreservationnetwork.org/eaasi/