



Emulation as a Service (EaaS) at Yale University Library

August 20, 2014 by [Trevor Owens](#)

The following is a guest post from Euan Cochrane,   Digital Preservation Manager at Yale University Library. This piece continues and extends exploration of the potential of [emulation as a service](#) ([//blogs.loc.gov/digitalpreservation/2012/12/the-foundations-of-emulation-as-a-service-an-interview-with-dirk-von-suchodoletz-part-one/?loclr=blogsig](http://blogs.loc.gov/digitalpreservation/2012/12/the-foundations-of-emulation-as-a-service-an-interview-with-dirk-von-suchodoletz-part-one/?loclr=blogsig)) and [virtualization platforms](#) ([//blogs.loc.gov/digitalpreservation/2011/09/toward-a-library-of-virtual-machines-insights-interview-with-vasanth-bala-and-mahadev-satyanarayanan/?loclr=blogsig](http://blogs.loc.gov/digitalpreservation/2011/09/toward-a-library-of-virtual-machines-insights-interview-with-vasanth-bala-and-mahadev-satyanarayanan/?loclr=blogsig)) .

Increasingly, the intellectual productivity of scholars involves the creation and development of software and software-dependent content. For universities to act as responsible stewards of these materials we need to have a well-formulated approach to how we can make these legacy works of scholarship accessible.

While there have been significant concerns with the practicality of emulation as a mode of access to legacy software, my personal experience (demonstrated via [one of my first websites](#) (<http://web.archive.org/web/19990210183008/http://www.geocities.com/SiliconValley/Horizon/6149/>) about Amiga emulation) has always been contrary to that view. It is with great pleasure that I can now illustrate the practical utility of Emulation as a Service via three recent case studies from my work at Yale University Library. Consideration of interactive artwork from 1997, interactive Hebrew texts from a 2004 CD-ROM and finance data from 1998 illustrate that it's no longer really a question of if emulation is a viable option for access and preservation, but of *how* we can go about scaling up these efforts and removing any remaining obstacles to their successful implementation.

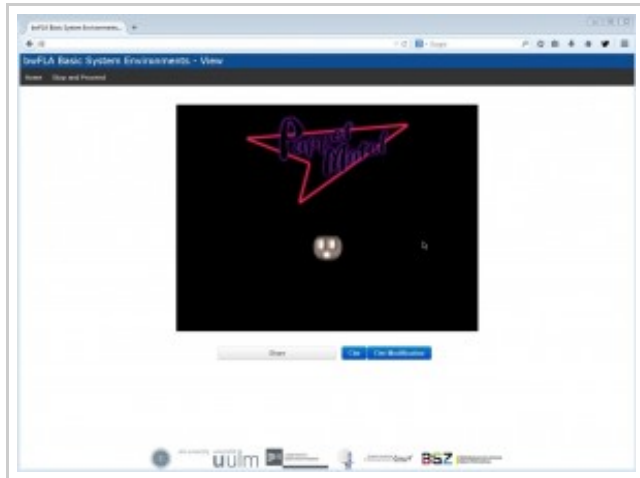
At Yale University Library we are conducting a research pilot of the [bwFLA Emulation as a Service](#) (<http://bw-fla.uni-freiburg.de/>) software framework. This framework greatly simplifies the use of emulators and virtualization tools in a wide range of contexts by abstracting all of the emulator configuration (and its associated issues) away from the end-user. As well as simplifying use of emulators it also simplifies access to emulated environments by providing the ability to access and interact with emulated environments from right within your web browser, something that we could only dream of just [a few years ago](#) (<http://www.openplanetsfoundation.org/blogs/2012-05-14-next-generation-access-emulation-onlive-gaikai-dp>) .

At Yale University Library we are evaluating the software against a number of criteria including:

1. In what use-cases might it be used?
2. How might it fit in with digital content workflows?
3. What challenges does it present?

The EaaS software framework shows great promise as a tool for use in many digital content management workflows such as appraisal/selection, preservation and access, but also presents a few unique and particularly challenging issues that we are working to overcome. The issues are mostly related to copyright and software licensing. At the bottom of this post I will discuss what these issues are and what we are doing to resolve them, but before I do that let me put this in context by discussing some real-life use-cases for EaaS that have occurred here recently.

It has taken a few months (I started in my position at the Library in September 2013) but recently people throughout the Library system have begun to forward queries to me if they involve anything digital preservation-related. Over the past month or so we have had three requests for access to digital content from the general collections that couldn't be interacted with using contemporary software. These requests are all great candidates for resolving using EaaS but, unfortunately (as you will see) we couldn't do that.



Interactive Artwork, Circa 1997: Use Case One

An Arts PhD student wanted to access an interactive CD-ROM-based artwork (Laurie Anderson's "[Puppet Motel](#)")

([//blogs.loc.gov/digitalpreservation/files/2014/07/interactive-artwork.jpg](http://blogs.loc.gov/digitalpreservation/files/2014/07/interactive-artwork.jpg))

Screenshot of Puppet Motel running in the emulation service using the Basilisk II emulator.

(<https://web.archive.org/web/20121207084410/http://www.wordyard.com/dmz/digicult/anderson-motel-5-10-95.html>) from the general collections. The artwork can only be interacted with on old versions of the Apple Mac "classic" operating system.

Fortunately the Digital Humanities Librarian ([Peter Leonard](http://resources.library.yale.edu/StaffDirectory/detail.aspx?q=718) (<http://resources.library.yale.edu/StaffDirectory/detail.aspx?q=718>)) has a collection of old technology and was willing to bring a laptop into the library from his personal collection for the PhD student to use to access it on. This was not an ideal or sustainable solution (what would have happened if Peter's collection wasn't available? What happens when that hardware degrades past usability?).

Since responding to this request we have managed to get the Puppet Motel running in the emulation service using the [Basilisk II emulator](http://basilisk.cebix.net/) (<http://basilisk.cebix.net/>) (for research purposes).

This would be a great candidate for accessing via the emulation service. The sound and interaction aspects all work well and it is otherwise very challenging for researchers to access the content.

Hebrew Texts, Circa 2004: Use Case Two

One of the Judaica librarians needed to access data for a patron and the data was in a Windows XP CD-ROM (Trope Trainer) from the general collections. The software on the CD would not run on the current Windows 7 operating system that is installed on the desktop PCs here in the library.

The solution we came up with was to create a Windows XP virtual machine for the librarian to have on her desktop. This is a good solution for her as it enables her to print the sections she wants to print and export pdfs for printing elsewhere as needed.

We have since ingested this content into the emulation service for testing purposes. In the EaaS it can run on either the virtualization software from Oracle: [VirtualBox](https://www.virtualbox.org/wiki/Downloads) (<https://www.virtualbox.org/wiki/Downloads>) (which doesn't provide full-emulation

(http://www.computerworld.com/s/article/print/338993/Emulation_or_Virtualization_) or QEMU (http://wiki.qemu.org/Main_Page) an emulation and virtualization tool.

It is another great candidate for the service as this version of the content can no longer be accessed on contemporary operating systems and the emulated version enables users to play through the texts and hear them read just as though they were using the CD on their local machine. The ability to easily export content from the emulation service will be added in a future update and will enable this content to become even more useful.



(<https://blogs.loc.gov/digitalpreservation/files/2014/07/hebrew-texts.jpg>)

Screenshot virtual machine used to access CD-ROM that wouldn't play in current OS.



(<https://blogs.loc.gov/digitalpreservation/files/2014/07/finance-data.jpg>)

Accessing legacy finance data through a Windows 98 Virtual Machine.

hadn't degraded and the software still worked. A Windows 98 virtual machine was then created for the student to use to extract the data. Exporting the data to the host system was a challenge. The simplest solution turned out to be having the researcher email the data to himself from within the virtual machine via Gmail using an old web browser (Firefox 2.x).

Finance Data, Circa 1998/2003: Use Case Three

A Finance PhD student needed access to data ([inter-corporate ownership data](http://www.statcan.gc.ca/ads-annonces/61-517-x/index-eng.htm) (<http://www.statcan.gc.ca/ads-annonces/61-517-x/index-eng.htm>)) trapped within software within a CD-ROM from the general collection. Unfortunately the software was designed for Windows 98: "As part of my current project I need to use StatCan data saved using some sort of proprietary software on a CD. Unfortunately this software seemed not to be compatible with my version of Windows." He had been able to get the data out of the disc but couldn't make any real sense of it without the software: "it was all just random numbers."

We have recently been developing a collection of old hardware at the Library to support long-term preservation of digital content. Coincidentally, and fortunately, the previous day someone had donated a Windows 98 laptop. Using that laptop we were able to ascertain that the CD

We were also able to ingest the virtual machine into the emulation service where it can run on either VirtualBox or QEMU.

This is another great candidate for the emulation service. The data is clearly of value but cannot be properly accessed without using the original custom software which only runs on older versions of the Microsoft Windows operating system.

Other uses of the service

In exploring these predictable use-cases for the service, we have also discovered some less-expected scenarios in which the service offers some interesting potential applications. For example, the EaaS framework makes it trivially easy to set up custom environments for patrons. These custom environments take up little space as they are stored as a difference from a base-environment, and they have a unique identifier that can persist over time (or not, as needed). Such custom environments may be a great way for providing access to sets of restricted data that we are unable to allow patrons to download to their own computers. Being able to quickly configure a Windows 7 virtual machine with some restricted content included in it (and appropriate software for interacting with that content, e.g., an MS Outlook PST archive file with MS Outlook), and provide access to it in this restricted online context, opens entirely new workflows for our archival and special collections staff.

Why we couldn't use bwFLA's EaaS

In all three of the use-cases outlined above EaaS was not used as the solution for the end-user. There were two main reasons for this:

1. We are only in possession of a limited number of physical operating system and application licenses for these older systems. While there is some capacity to use downgrade rights within the University's volume licensing agreement with Microsoft, with Apple operating systems the situation is much less clear. As a result we are being conservative in our use of the service until we can resolve these issues.
2. It is not always clear in the license of old software whether this use-case is allowed. Virtualization is rarely (if ever) mentioned in the license agreements. This is likely because it wasn't very common during the period when much of the software we are dealing with was created. We are working to clarify this point with the General Counsel at Yale and will be discussing it with the software vendors.

Addressing the software licensing challenges

As things stand we are limited in our ability to provide access to EaaS due to licensing agreements (and other legal restrictions) that still apply to the content-supporting operating system and productivity software dependencies. A lot of these dependencies that are necessary for providing access to valuable historic digital content do not have a high economic value themselves. While this will likely change over time as the value of these dependencies becomes more recognized and the software more rare, it does make for a frustrating situation. To address this we are beginning to explore options with the software vendors and will be continuing to do this over the following months and years.

We are very interested in the opportunities EaaS offers for opening access to otherwise inaccessible digital assets. There are many use-cases in which emulation is the only viable approach for preserving access to this content over the long term. Because of this, anything that prevents the use of such services will ultimately lead to the loss of access to valuable and historic digital content, which will effectively mean the loss of that content. Without engagement from software vendors and licensing bodies it may require law change to ensure that this content is not lost forever.

It is our hope that the software vendors will be willing to work with us to save our valuable historic digital assets from becoming permanently inaccessible and lost to future generations. There are definitely good [reasons](http://news.bbc.co.uk/2/hi/6265976.stm) (<http://news.bbc.co.uk/2/hi/6265976.stm>) to [believe](http://www.info.apple.com/support/oldersoftwarelist.html) (<http://www.info.apple.com/support/oldersoftwarelist.html>) that they will, and so far, those we have contacted have been more than willing to work with us.

Posted in: [Digital Content](#), [Tools and Infrastructure](#)

[1 Comment](#) | [Add a Comment »](#)

One Comment

1. **Matthew McKinley**

August 20, 2014 at 3:07 pm

Love the idea of a browser-based emulator as a “virtual reading room” for restricted archival material. We have a VRR here at UC Irvine but it’s more of a conceptual solution (treating DSpace user account registration as a form of Reading Room registration) than a technological one.

RE: Licensing, I saw a very interesting presentation at SAA last week regarding a proposed Software Preservation Network focused both on collecting and maintaining software as well as negotiating licensing agreements for the purpose of providing access to digital materials. The presenters were Zach Vowell at Cal Poly and Jessica Meyerson & Carlos Ovalle of UT Austin. They’re still in the very early planning and grant-finding stage, but it seems like a fruitful avenue for conversation & collaboration.

Disclaimer

This blog does not represent official Library of Congress communications.

Links to external Internet sites on Library of Congress Web pages do not constitute the Library's endorsement of the content of their Web sites or of their policies or products. Please read our [Standard Disclaimer](#).