



EaaS*i*

# Supporting the Reproducibility of Software Dependent Research

**The EaaS Program of Work**

<http://www.eaasi.info>

PV2023

Euan Cochrane, Yale University Library

Yale



# Hello!

I am **Euan Cochrane, Head of Digital Preservation,  
Yale University Library**

**Co-Principal Investigator** on the **EaSI Program of  
work**

You can find me at <https://digipres.club/@euanc>

EaSI's website is available at: <http://www.eaasi.info>

# Overview

1. Why software preservation and emulation matter for reproducibility
2. Emulation as a Service (EaaS)
3. The Emulation as a Service Infrastructure (EaaS) program of work
4. The Future for EaaS





# **Why Software Preservation and Emulation?**

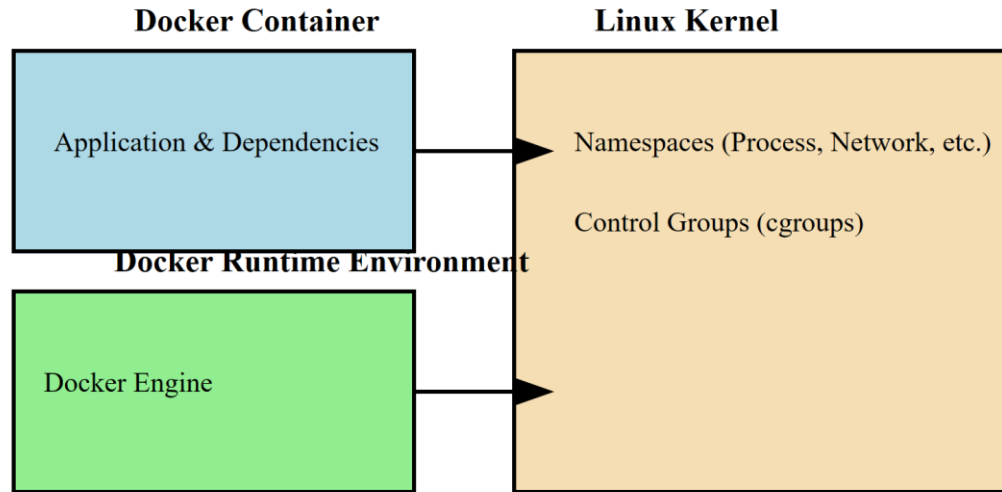


**Software Dependent  
Research Requires Software**

**&**

**Software Becomes  
Obsolete/Inaccessible**

# Containers also rely on software that obsolesces

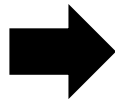


Please excuse the crudity of this diagram, it was an SVG generated by ChatGPT v4.0

```

WordStar D:\...\SC525-13\MULFUNC.DOC
File Edit View Insert Style Layout Utilities Help
Body Text Default font B I U <*> L C R J
TAPER EQUATIONS
-----
PREDICTION OF DIAMETER(d) AT ANY LENGTH(L) FROM TIP OF TREE IS :
d = 100*SqRoot((4V/(Pi*H))*
      2      3      4      5      B7
(B1(L/H) + B2(L/H) + B3(L/H) + B4(L/H) + B5(L/H) + B6(L/H) ) )
PREDICTION OF VOLUME(v) at A LENGTH(L) FROM TIP OF TREE IS :
v = (V/H)*(B1(L/2H) + B2(L /3H ) + B3(L /4H ) + B4(L /5H ) + B5(L /6H )
      B7+1      B7
      + B6(L / (B7+1)H ) )
EQUATION CODES ARE Tinn FOR INSIDE BARK function no nn.
Insert P1 L40 V7.00" C1 H0.00"

```



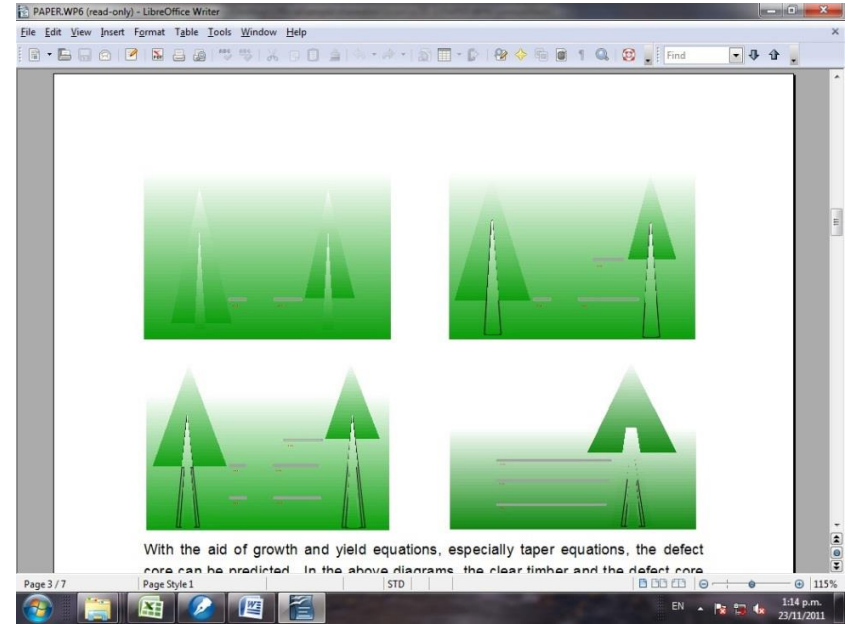
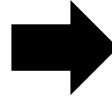
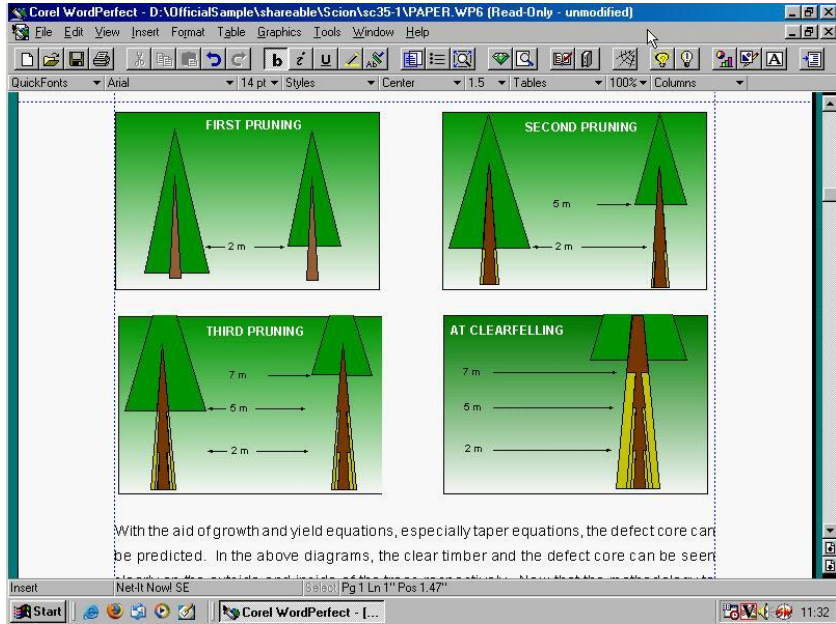
```

LibreOffice Writer
File Edit Format Table Tools Window Help
Text Times New Roman 10
TAPER EQUATIONS
-----
PREDICTION OF DIAMETER(d) AT ANY LENGTH(L) FROM TIP OF TREE IS :
_d = 100*SqRoot((4V/(Pi*H))*
      2      3      4      5      B7
(B1(L/H) + B2(L/H) + B3(L/H) + B4(L/H) + B5(L/H) + B6(L/H) ) )
PREDICTION OF VOLUME(v) at A LENGTH(L) FROM TIP OF TREE IS :
|
      2      3      4      5      6      5
_v = (V/H)*(B1(L/2H) + B2(L /3H ) + B3(L /4H ) + B4(L /5H ) + B5(L /6H )
      B7+1      B7
      + B6(L / (B7+1)H ) )
EQUATION CODES ARE Tinn FOR INSIDE BARK function no nn.
COLUMN NUMBER_DESCRIPTION
-----
2- 5 EQUATION CODE
7-31 Description
32-39 B1 * 100,000 Normally the first column of
40-47 B2 * 100,000 each of these will be blank

```

Original content in original software  
(WordStar Version 7 for MS-DOS)

Original content in newer software  
(LibreOffice Writer)



Original content in original software  
(WordPerfect in Windows 95)

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



Layer: 0 Ortho 121'-9" 30'-6 3/32"

AutoCAD  
 \* \* \* \* \*  
 ASE  
 BLOCKS  
 DIM:  
 DISPLAY  
 DRAW  
 EDIT  
 INQUIRY  
 LAYER...  
 MODEL  
 MVIEW  
 PLOT...  
 RENDER  
 SETTINGS  
 SURFACES  
 UCS:  
 UTILITY  
 SAVE:

TOTAL FLOOR GROSS AREA IN SF  
7718 SF

TEMPLE STREET

GROVE STREET

1716A001.DWG

<b>Yale University</b> Building Assignment Plan	Bldg. Name:	109 GROVE	111: First floor plan
	St. Address:	109-111 Grove Street	1/20"=1'-0" Date: 06/30/9

Loaded menu C:\ACAD\SUPPORT\ACAD.mnx  
 AutoCAD Release 12 menu utilities loaded.  
 Command:

Software for format accessibility

# Demo: UVI

EaaS Demo UI

https://uvi.emulation.cloud/admin/#/admin/uv

Yale Admin EaaS Instances MacOS Windows Linux Computing History QEMU Software Sites Documentation Imaging Some\_blogs Test Data xkcd WDPD2019 Workshop Inspiration

- Environments
- Software
- Objects
- UVI
- Networks

---

- Import Environment
- Create Environment
- Import Container
- OAI PMH
- Settings
- Emulators

Build: 737A80CA6B  
UI-Build: 0140F83042

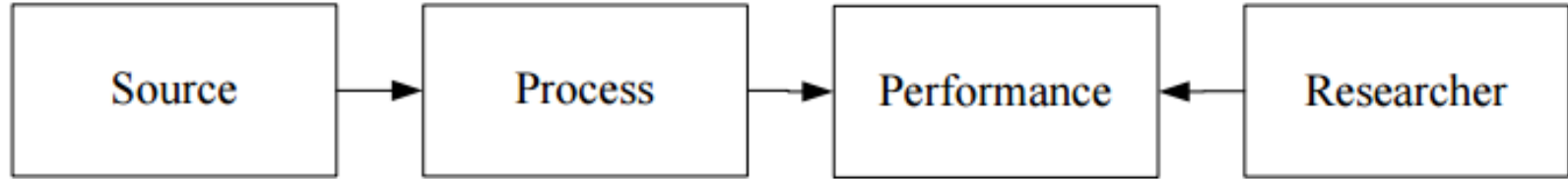
## UVI

Object upload

Upload a file to render:

Use writeable media (supports data export)

# National Archives of Australia “Performance Model”



**Figure 2: Performance model – source and process components**



# **Emulation as a Service**

(EaaS)

# EaaS simplifies access to, and provides a generic API for, various emulators *and* KVM



PCE - PC Emulator



Universal Amiga  
Emulator



KEGS - Kent's Emulated GS



# Emulation/Virtualization

- An emulation software package (“emulator”) is used to create a virtual version of one computer within another computer that has different hardware
- Old software can be run on the “emulated” computer hardware just like it was running on the original physical computer.



# EaaS provides web-based access to emulators and KVM *preconfigured* to run numerous Operating Systems

## Create Base Environment

### Choose System

- QEMU Mac OS 8/9/10.0-10.4 (PPC Qemu)
- Generic 90s PC
- Generic 2000s PC
- WinXP (IDE generic)
- (WebEmulator) v86
- Windows 98 (USB pointer)
- Apple (Motorola) (7.x - 8.x)
- Windows 7 (generic ide)

Customization of emulator/KVM parameters is also available:

### Native Config

# EaaS: Features of note

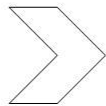
- Seamlessly move environments/containers to emulated hardware from virtualized or physical hardware as technology ages
- Sophisticated virtual hard disk management
  - Dynamically translates between HDD image formats (vdi, vmdk, raw, qcow, qcow2, etc)
  - Supports linked disk images/environments that are dependent on parent images
- Can print to PDF from any environment with a post-script printer driver (universal PDF conversion)
- Handles/Persistent identifiers available for configured environments
- Internet access within configured environments an optional feature
- Environments can be paused and resumed on demand
- Generic Application Programming Interface
  - Interact with emulation session and disk images
  - Upload/attach content
  - Download changed files
- Networked emulated environments and environment isolation and access (e.g. via proxy)





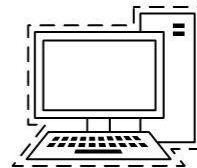
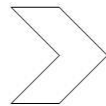
“Base”  
Environment  
(e.g. Windows XP)

5 GB  
(Compressed)



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

150MB

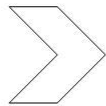


Derivative x1.1  
(e.g. Windows XP + SPSS 13 + SPSS code)

1MB

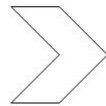
~5,151MB  
(with EaaS)

Instead of  
~15,301MB  
(without)



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

250MB



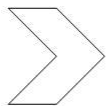
Derivative x3  
(e.g. Windows XP+ STATA 8.2 + STATA  
code)

1MB

~5,251MB  
(with EaaS)

Instead of  
~15,501MB  
(without)

# Derivatives: example



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

50MB

~5,050MB  
(with EaaS)

Instead of  
~10,050MB  
(without)

**TOTAL**    Instead of  
~5,452MB    ~30,852MB  
(with EaaS) (without)

EaaS requires  
~83% less data




# **Emulation as a Service Infrastructure**

(EaaSI)



# Why EaaS?

- Finding legacy software is hard.
  - There is a lack of comprehensive metadata describing software, its requirements, and capabilities
  - Legacy software is hard to configure
  - Integrating emulation in workflows is difficult
- 

## ***Usable software, forever***

EaaS provides technology and services for the application of software emulation across a diverse spectrum of professional disciplines, organizations, and individual use cases.

# Core Services

- **Search & Discover**

*Find software and environments to re-use*

- **Import**

*Upload your content and software to EaaSI*

- **Create**

*Configure new computer environments*

- **Manage**

*Administer access to your node and control use*

- **Access**

*Provide end-user access to software and content*

- **Collaborate**

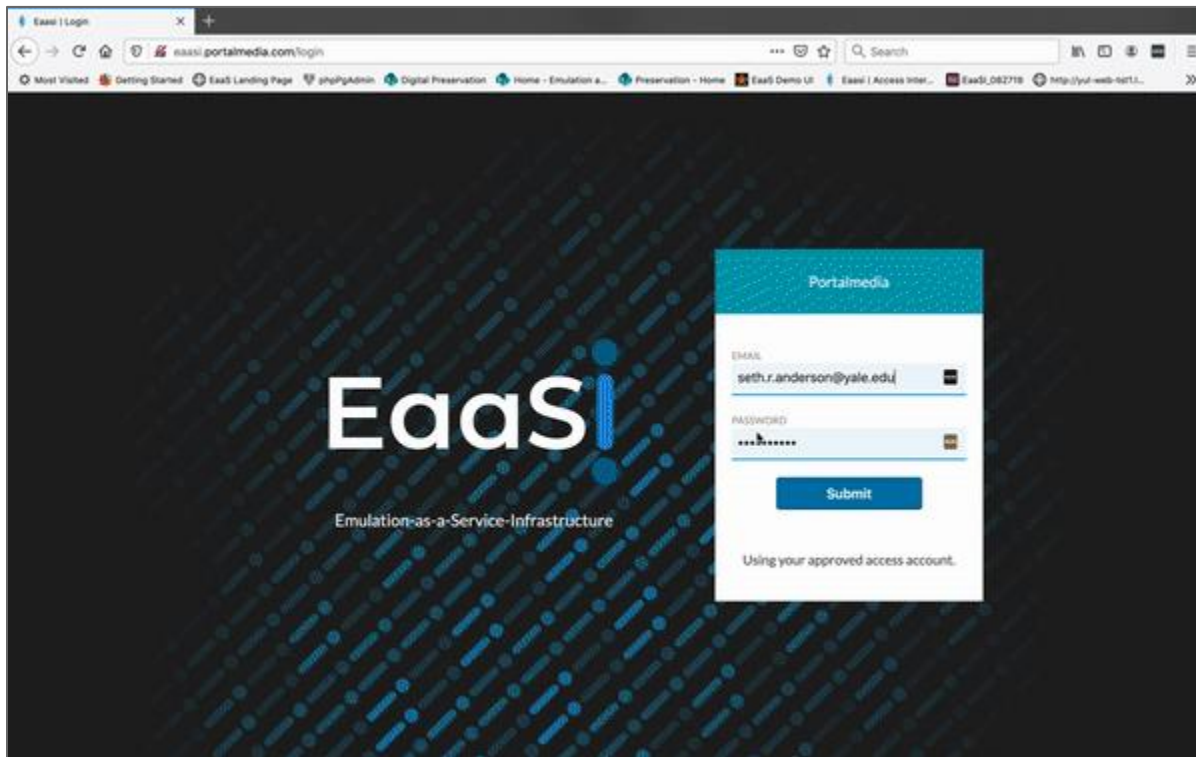
*Share software and environments and join the community of EaaSI users*



# Benefits

- Makes finding necessary software EaaS
- Supplies right emulator version and preconfigures common settings
- Enables access to emulators over the web via browser interface
- Smooths creation of new emulated computing environments
- Smooths configuration of common OS features in legacy systems





EaaS User Interface



# Examples





Environments



Software

Objects

Import Environment

Create Environment

OAI PMH

Settings

Emulators

Build: ADAC9CD452  
UI-Build: 991368F06F

# Demo: Repronzip

## Environments

Virtual machines

Object Environments

Number of Environments: 22

Page Size: 10

repro

Name ↑	ID	Own...	ObjectID	Actions
<input type="checkbox"/> Stacked Up ReproZip Example	6eaf01...	shared	ReproZip_Package-S...	Choose action ▾

[1] to [1] of [1] << Page [1] of [1] >>

Environments



Software

Objects

UVI

Networks

Import Environment

Create Environment

Import Container

OAI PMH

Settings

Emulators

## UVI

Object upload

Upload a file to render: GerberGreenLarimer\_APSR\_2008.do

Choose

Use writeable media (supports data export)

Upload additional files

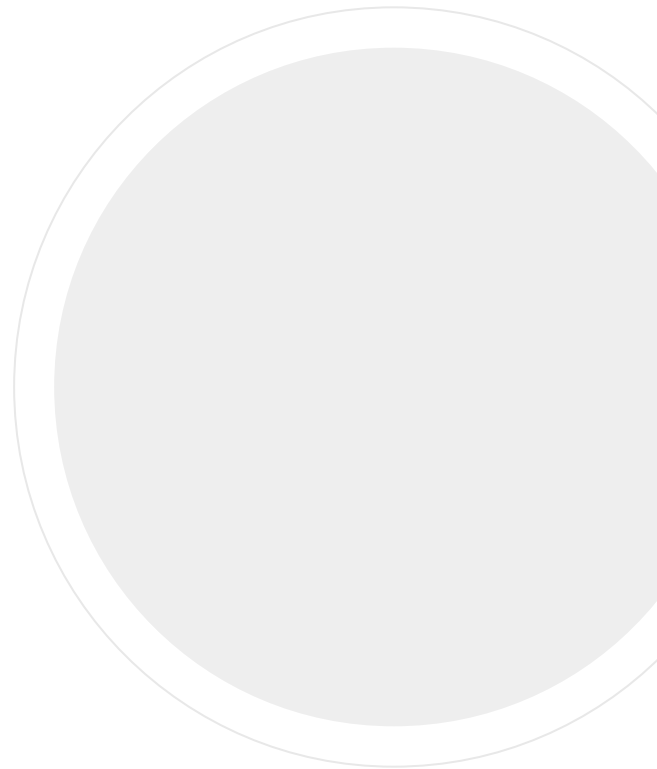
- GerberGreenLarimer\_APSR\_2008\_social\_pressure.dta
- GerberGreenLarimer\_APSR\_2008\_social\_pressure\_household\_level\_stata\_output.dta
- gerbergreenlarimer\_apsr\_2008\_social\_pressure\_household\_level\_stata\_output.csv
- GerberGreenLarimer\_APSR\_2008.log
- GerberGreenLarimer\_APSR\_2008.R
- GerberGreenLarimer\_APSR\_2008\_r\_output.txt
- GerberGreenLarimer\_APSR\_2008\_social\_pressure.csv
- 15d48af8-e38e-4dd0-ace9-62f90826963a.ddi32.xml

Upload



# Challenges

- Legal/Copyright
- Lack of Emulators
- Distributed digital objects
- Integration
- Application signatures and a software registry
- Loss of Open-Source Software Repositories



# Copyright

CODE OF BEST PRACTICES IN  
FAIR USE FOR  
SOFTWARE PRESERVATION

<http://www.arl.org/component/content/article/6/4630>

Released 24<sup>th</sup> September 2018

**Brandon Butler**, Director of Information Policy  
University of Virginia Library

**Krista Cox**, Director of Public Policy Initiatives  
Association of Research Libraries

**Peter Jaszi**, Professor Emeritus  
Program on Information Justice and Intellectual Property  
Washington College of Law, American University

FUNDED BY

**The Alfred P. Sloan Foundation**

AN AFFILIATED PROJECT OF

**The Software Preservation Network**

**EaaS*i***



# **EaaS Futures**

Menu



Actions

Screenshot

Download Print Jobs

Save Environment

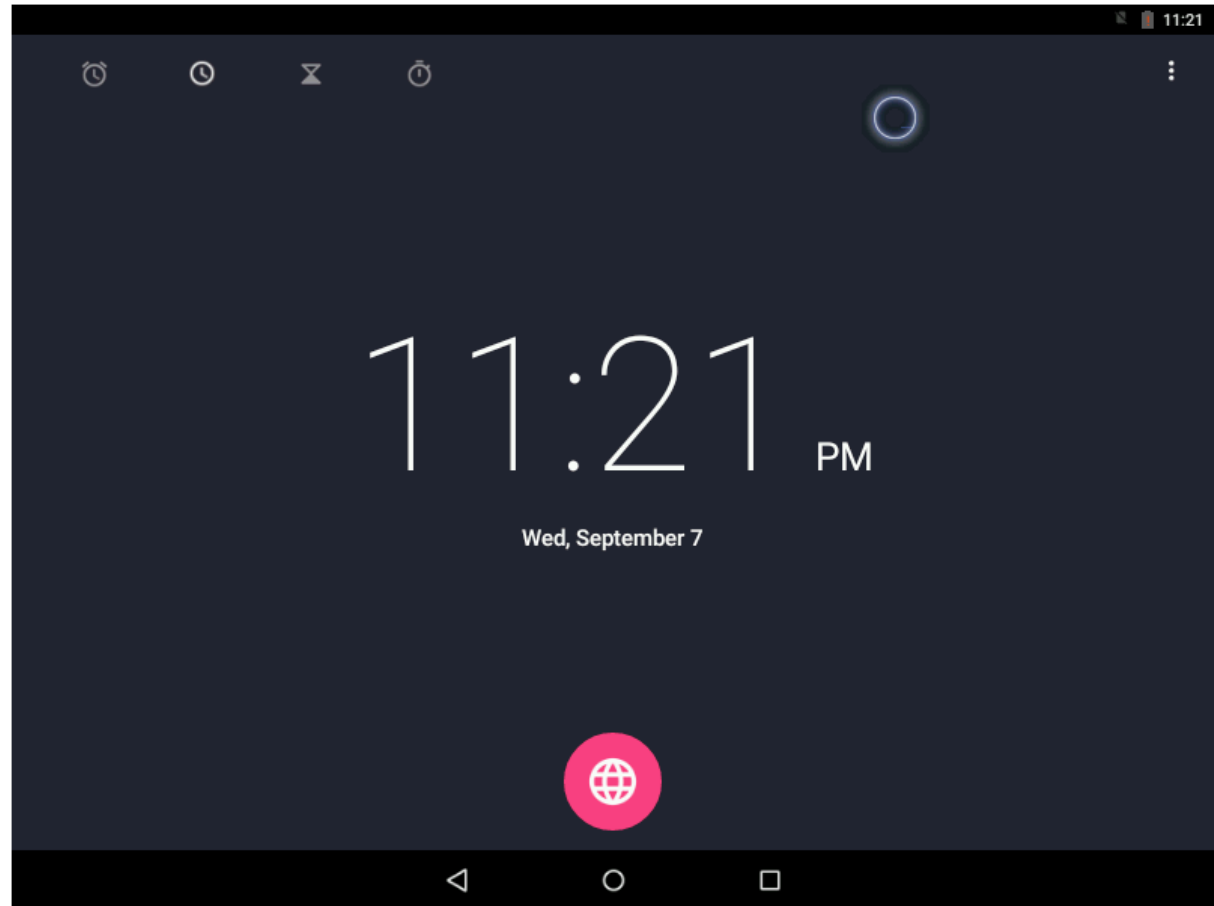
Send Ctr-Alt-Del

Send Esc

Restart Session

Connection Info

Stop



Menu



Actions

Screenshot

Download Print Jobs

Save Environment

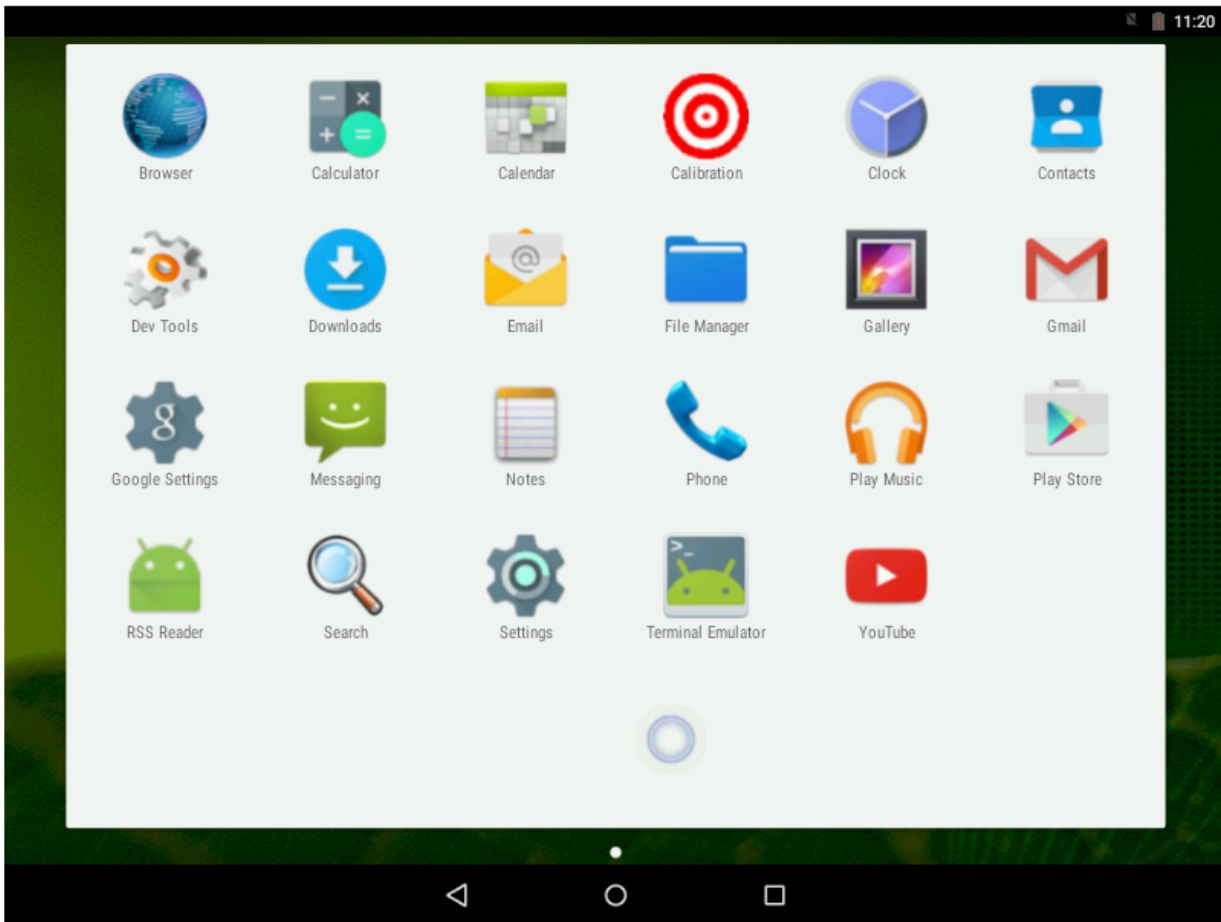
Send Ctr-Alt-Del

Send Esc

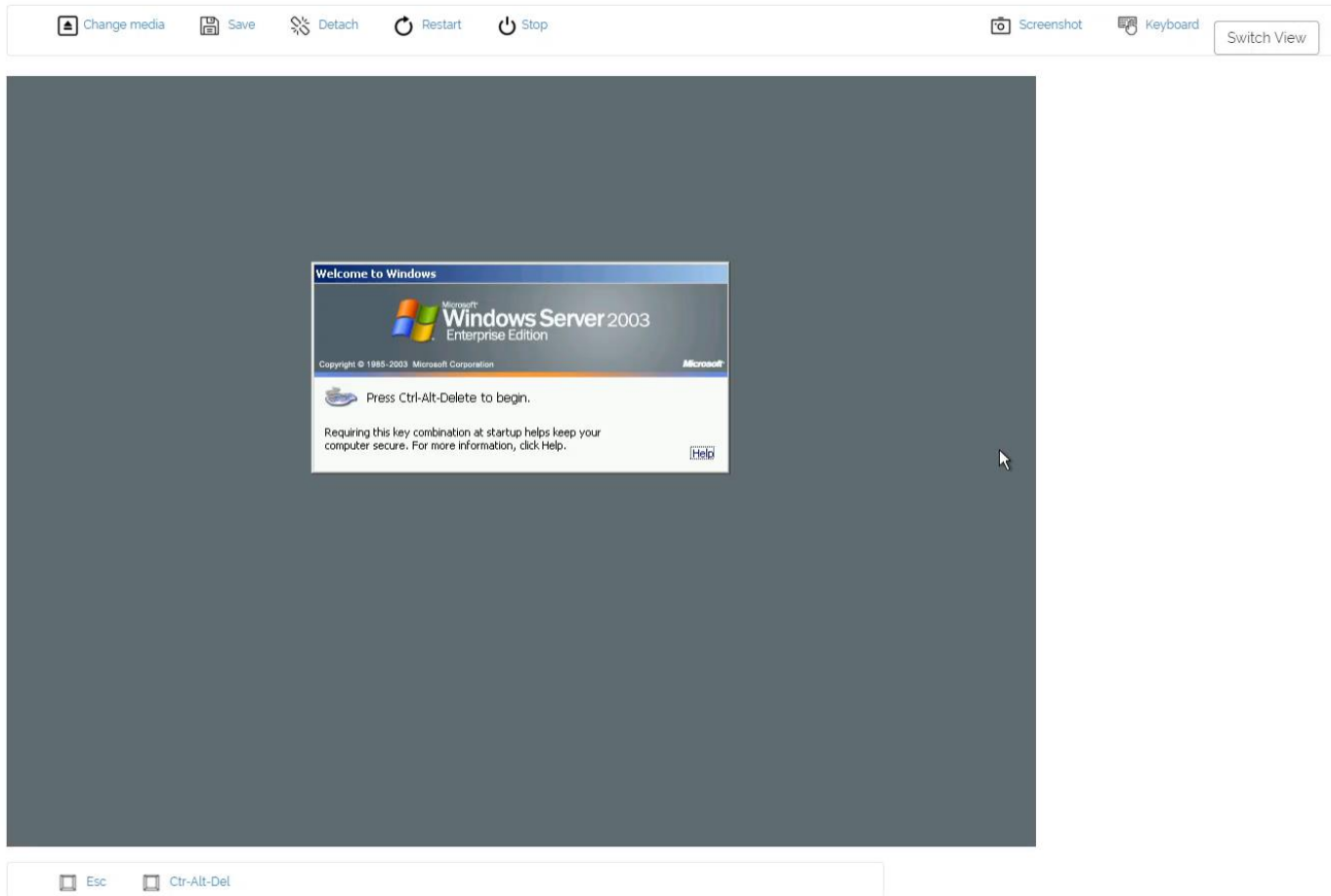
Restart Session

Connection Info

Stop



# Demo: Networked environments









Home



Machines



Networks



Images



User Assets



Software



Session Data



Automation



Settings



Help



Logout

# Networks

Available Networks

Running Networks

alfresco ad

Create Network

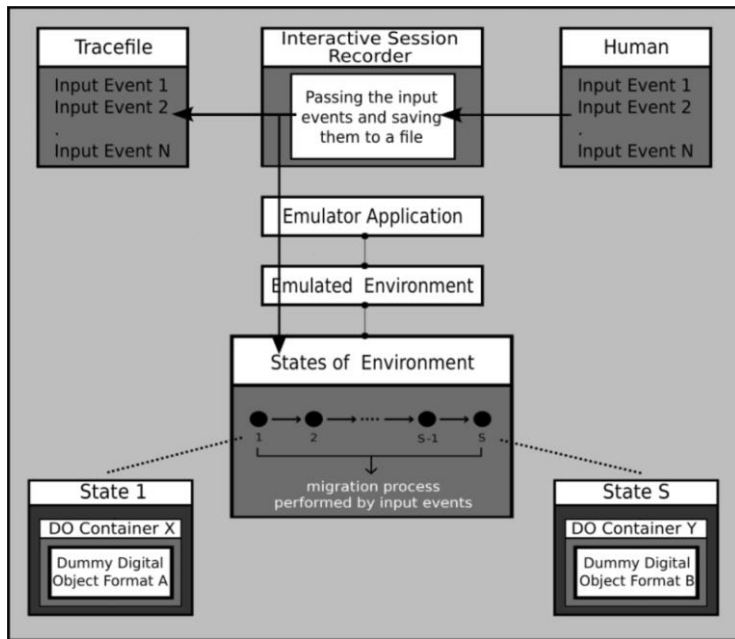
ID	Name	
7052b848-7b6b-4e60-b310-110371614cc5	Alfresco Additional Files	select ... ▼

Page Size: 25 ▼

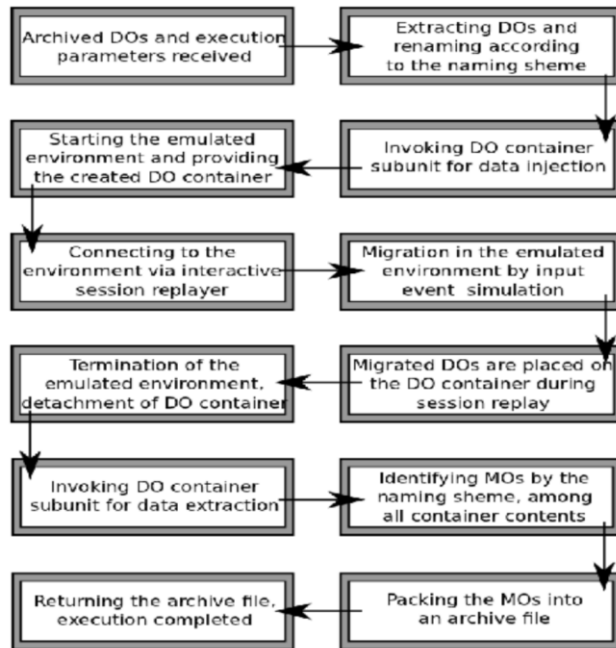
[1] to [1] of [1]

< < Page [1] of [1] > >

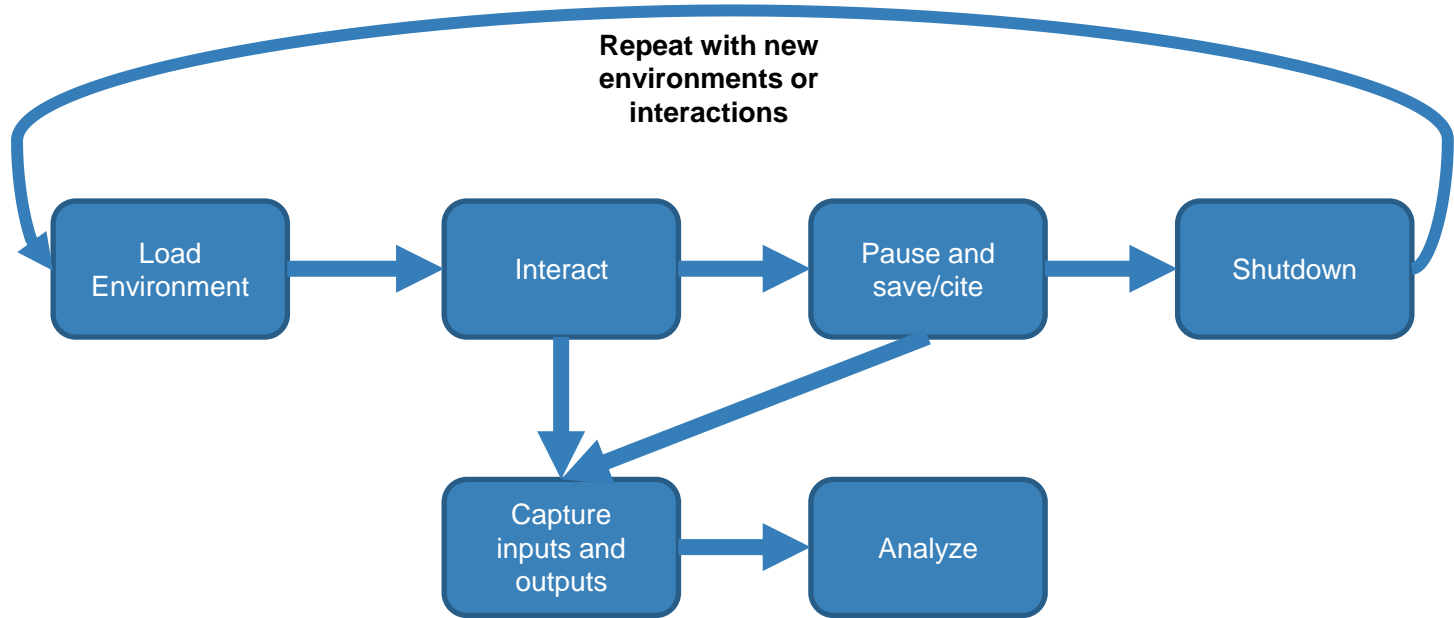
# Migration via Emulation (Planets Project)



<http://eprints.rclis.org/16263/>



# Analyzing software environments at scale



# Applications of machine learning to Software Preservation and Access

- Distant reading of software
- Distant reading of digital objects in context
- Learning standard interactions with similar types of objects
  - e.g. “would you like to try this interaction that most users have previously performed (and stuck with)?”
- Learning how to document software
  - Analyze file systems and interactions with the active software or objects opened within the software and
- Match objects to appropriate interaction environments automatically
- Match software to emulated hardware configurations
- Match disk images to emulated hardware configurations (e.g. from BitCurator).

# Establishing an Software Metadata- Registry

- Built on the Wikibase platform
- Able to uniquely and persistently Identify all software versions
- Able to be queried in federation with other wikibase instances, e.g. Wikidata
- A platform for recording software signatures

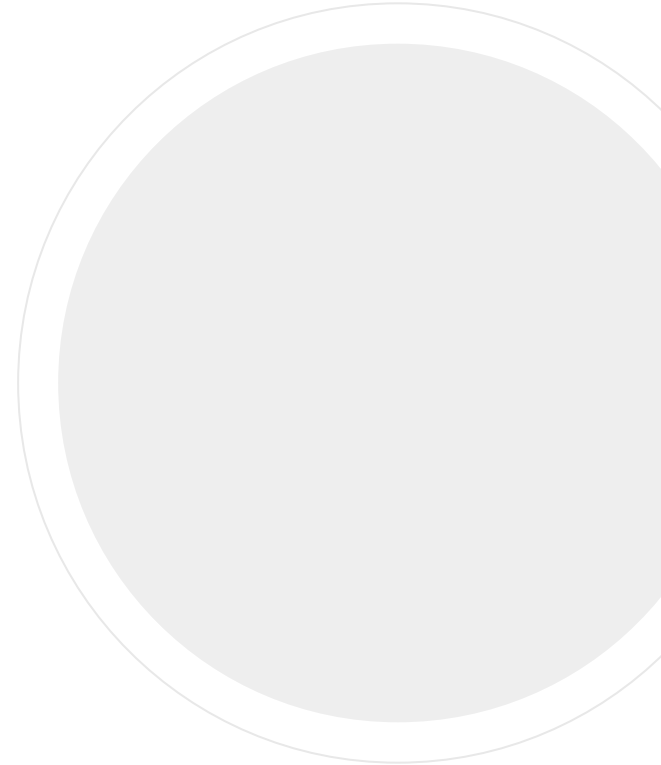
# Establishing an Open-Source Repository “Archive”

- Open-source repositories disappear
- Open-source repositories that exist are slow to access
- Changing repository configurations in preserved software environments is complicated.



# Summary

- We *must* preserve software and make it accessible forever to preserve access to software dependent research
- EaaS is providing a platform to make this feasible
- While there are challenges still ahead the future applications of EaaS are extremely promising





# A Very Special Thanks to our Funders...



Alfred P. Sloan  
FOUNDATION



# Thank you!

You can find me at

- [ewan.cochrane@yale.edu](mailto:ewan.cochrane@yale.edu)
- <https://digipres.club/@euanc>

Learn more at

- <http://www.eaasi.info>

