

Investigating Emulation as a Service for Reproducible Research at Yale

Limor Peer

Associate Director for Research, Yale ISPS

Ethan Gates

Software Preservation Analyst, YUL

Librarians Building Momentum for Reproducibility
A Virtual Conference
January 28, 2020

Institution for Social and Policy Studies



ISPS was founded in 1968 as an inter-disciplinary center to support social science and public policy research at Yale University.

The **ISPS Data Archive** provides open access to digital collections of social science experimental data, metadata, code, and associated files produced by ISPS researchers for the purpose of replication of research findings, further analysis, and teaching.

A screenshot of the ISPS Data Archive website. The header includes the Yale ISPS 50 Years logo (1968-2018) and the text "Institution for Social and Policy Studies" with a tagline "ADVANCING RESEARCH · SHAPING POLICY · DEVELOPING LEADERS". A navigation bar contains links for Research, Events, News, Team, About, and Programs. The main content area shows a breadcrumb trail "HOME > RESEARCH > DATA" followed by the word "Data" in a large font. Below this is a grey box labeled "ISPS Data Archive" and a row of four buttons: "Browse", "Deposit", "About", and "Approach". A welcome message follows: "Welcome to the ISPS Data Archive!" and a paragraph of text explaining the archive's content and access policies, including a link to the ISPS KnowledgeBase.

<https://isps.yale.edu/research/data>

Computational reproducibility refers to changes in scientific practice and reporting standards to accommodate the use of computational technology...in particular whether the **same results can be obtained from the data and code used in the original study.**

Stodden, V. (2015). Reproducing statistical results. *Annual Review of Statistics and Its Application*, 2(1), 1–19. <https://doi.org/10.1146/annurev-statistics-010814-020127>

Data Quality Review Framework



Peer, L., Green, A., & Stephenson, E. (2014). Committing to data quality review. *International Journal of Digital Curation*, 9(1). <https://doi.org/10.2218/ijdc.v9i1.317>

Software-dependent reproducibility problems

- Original code tied to legacy software
- Legacy software that is no longer available
- Proprietary software that is difficult to package with reproducibility packaging tools
- The packaging runtime (e.g. Docker, Rezip) is no longer supported on modern operating systems

The image features a large, dark blue circular background with a bokeh effect of smaller, lighter blue circles. Overlaid on this are two concentric white circles. The text 'EaaS' is centered within the inner circle. The 'E' and 'a' are white, while the second 'a' and 'S' are blue. The 'i' is a vertical blue bar.

EaaS

A Very Special Thanks to our Funders...



THE
ANDREW W.
MELLON
FOUNDATION



Alfred P. Sloan
FOUNDATION



Project Goal

Deploy and scale infrastructure and services for software emulation, including distributed management, sharing, documentation/discovery, and access.

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

What is Emulation-as- a-Service (EaaS)?

Simplifies access to various emulators



Clockwise from top: FS-UAE (Amiga), LinApple (Apple II), VICE (Commodore), Mini vMac (Macintosh Plus), SheepShaver (PowerPC Macs), BeebEm (BBC Micro), QEMU (x86 PCs), Hatari (Atari)

Enables management of persistent emulation environments

The screenshot shows a web interface for managing emulation environments. On the left is a sidebar with navigation options: 'Environments', 'Object Environments', 'Software', 'Create Base Environment', and 'Settings'. The main content area is titled 'Environments' and features a '+ New environment' button and a search bar. Below the search bar is a list of four environments, each with a title and a set of action links. At the bottom right, there is a pagination control showing page 1 of 6.

Environment Name	Actions
Apple Mac OS 7.5	[Configure environment] [Edit description] [export] [Delete] clean
Apple Mac OS 7.5	[Configure environment] [Edit description] [export] [Delete] 4
Apple Mac OS 9	[Configure environment] [Edit description] [export] [Delete] n.a.
Atari 1040ST (68000 CPU)	[Configure environment] [Edit description] [export] [Delete] n.a.

Environments

Object Environments

Software

Create Base Environment

Settings

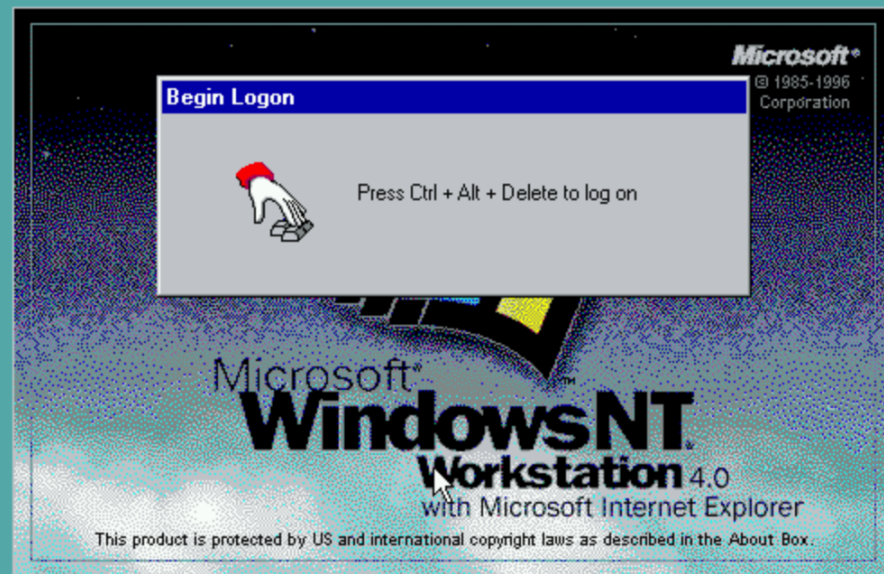
Actions

Screenshot

Restart

Stop

Save Changes



Emulation-as-a-Service

- In development by the bwFLA team at the University of Freiburg since 2011 (now commercially maintained by OpenSLX)
- Since 2017 CiTAR builds RDM workflows to repeat, replicate, reproduce or reuse software based research on top of EaaS
- Since 2018 the EaaSI project



The image features a large, dark blue circular background with a bokeh effect of smaller, lighter blue circles. In the center, the text "EaaS" is displayed in a white, sans-serif font. The letter "i" is replaced by a solid blue vertical bar. The entire composition is framed by two concentric white circles.

EaaS

D001

Field Date: 2006

Archive Date: 2010

Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment

ISPS Data Archive

[Browse](#)

[Deposit](#)

[About](#)

[Approach](#)

DATA FILE NUMBER	DESCRIPTION	FILE FORMAT	SIZE	FILE URL
D001F01	Dataset (individuals)	Excel .csv	32744038	Download file
D001F02	Dataset (individuals)	.dta	41293806	Download file
D001F03	Dataset (households)	Excel .csv	7077888	Download file
D001F04	Dataset (households)	.dta	14365491	Download file
D001F05	Program file	.do	1509	Download file
D001F06	Output file	.txt	16076	Download file
D001F07	Program file	R (2.9.1) .R	4022	Download file
D001F08	Output file	.log	10240	Download file
D001F09	Treatment materials	Adobe Acrobat (8.1) .pdf	808960	Download file
D001F13	Metadata (DDI 3.2)	.xml	197766	Download file



Approach #1: Manually “Rebuild” Computing Environment



Environments



Software

Objects

Import Environment

Create Environment

OAI PMH

Settings

Emulators

Edit Run Add Software

ID: 2c0826f5-97e7-4576-94aa-6c09460a0ca9

Name: YARD-ISPS ID D001 Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment

Handle: --- [create](#)

Description



Emulation system settings



Configured Drives



UI options



Networking



Configured software

- a95bba37-70fc-4820-a287-1d35ef75bd66
- 7925-Microsoft_Office_Professional_Enterprise_Edition_2003
- Stata_10
- Adobe_Acrobat_Reader_9.2
- R_2.9.1_Windows_32_bit

Revision history

Revisions:


- Copied data from CD to desktop.
Put shortcut to data folder in the startup folder.

<https://isps.yale.edu/research/data/d001>



A graphic consisting of four concentric circles of varying diameters, centered on the left side of the slide. The circles are light gray and overlap each other.

Approach #2: UVI

- 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 

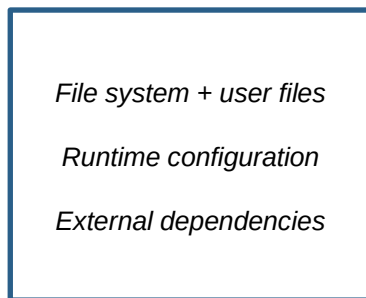
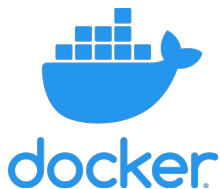
Upload

Build: 737A80CA5B
UI-Build: 0140F83042

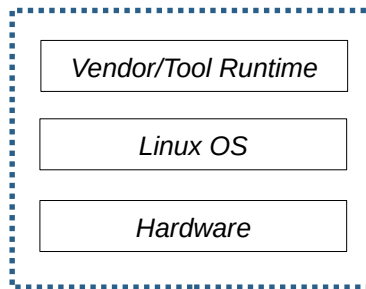


Approach #3: Container/Package Import

Provided by User:



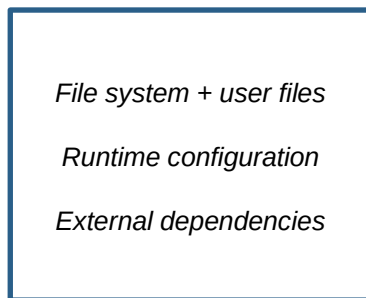
Depend on:



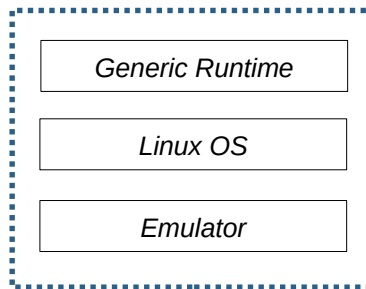
***Provided by
combination of
vendor and user -
unstable***

After EaaS Import:

EaaS!



Depend on:



Provided by EaaS –

stable

- Environments
- Software
- Objects
- Detached Networks
- Import Environment
- Create Environment
- Import Container
- OAI PMH
- Settings
- Emulators

Create new Container

Choose Origin Runtime
docker

Docker

Image Format
EaaS can process Docker images from Docker registries. Registry hosted images are referenced via name and tag. Additionally, paths for input and output directories are required.

Specify container name and tag:

Name:
Image Name (e.g phusion/l)

Tag:
Image Tag (e.g 0.9.22)

Input path
/input

Output path
/output

Cancel Start

Thank you!

ethan.gates@yale.edu
@The_BFOOL

limor.peer@yale.edu
@l_peer

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

<https://isps.yale.edu/research/data>