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
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.

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.

Data Quality Review Framework

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, Reprozip) is no longer supported on modern operating systems
A Very Special Thanks to our Funders...
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
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
Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment

ISPS Data Archive

<table>
<thead>
<tr>
<th>Browse</th>
<th>Deposit</th>
<th>About</th>
<th>Approach</th>
</tr>
</thead>
</table>

D001
Field Date: 2006
Archive Date: 2010

<table>
<thead>
<tr>
<th>DATA FILE NUMBER</th>
<th>DESCRIPTION</th>
<th>FILE FORMAT</th>
<th>SIZE</th>
<th>FILE URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>D001F01</td>
<td>Dataset (individuals)</td>
<td>Excel .csv</td>
<td>3274403B</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F02</td>
<td>Dataset (individuals)</td>
<td>.dta</td>
<td>41293806</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F03</td>
<td>Dataset (households)</td>
<td>Excel .csv</td>
<td>7077888</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F04</td>
<td>Dataset (households)</td>
<td>.dta</td>
<td>14365491</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F05</td>
<td>Program file</td>
<td>.do</td>
<td>1509</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F06</td>
<td>Output file</td>
<td>.txt</td>
<td>16076</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F07</td>
<td>Program file</td>
<td>R (2.9.1) .R</td>
<td>4022</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F08</td>
<td>Output file</td>
<td>.log</td>
<td>10240</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F09</td>
<td>Treatment materials</td>
<td>Adobe Acrobat (8.1) .pdf</td>
<td>808960</td>
<td>Download file</td>
</tr>
<tr>
<td>D001F13</td>
<td>Metadata (DDI 3.2)</td>
<td>.xml</td>
<td>197766</td>
<td>Download file</td>
</tr>
</tbody>
</table>
Approach #1:
Manually “Rebuild” Computing Environment
1. **ID:** 2018265-97e7-4578-94aa-bc0b4ab0ca9
2. **Name:** YARD-ISP5 ID D001 Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment
3. **Handle:** --- create

**Emulation system settings**

**Configured Drives**

**UI options**

**Networking**

**Configured software**

- ab5bba37-70fc-4020-a267-1d35c75bc6e6
- 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://tesc.yale.edu/research/data/d001
Approach #2: UVI
Approach #3: Container/Package Import
Provided by User:

- File system + user files
- Runtime configuration
- External dependencies

Depend on:

- Vendor/Tool Runtime
- Linux OS
- Hardware

Provided by combination of vendor and user - unstable
After EaaS Import:

EaaS!

File system + user files
Runtime configuration
External dependencies

Depend on:

Provided by EaaS – stable

Generic Runtime
Linux OS
Emulator
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