Saving Software and Using Emulation to Reproduce Computationally Dependent Research Results

Euan Cochrane
Yale University, euan.cochrane@yale.edu

Limor Peer
Yale University, limor.peer@yale.edu

Ethan Gates
Yale University, ethan.gates@yale.edu

Seth Anderson
Yale University, seth.r.anderson@yale.edu

Saving Software and Using Emulation to Reproduce Computationally Dependent Research Results

Euan Cochrane, Seth Anderson, Ethan Gates
Yale University Library
https://www.softwarepreservationnetwork.org/eaasi/

Limor Peer
Institution for Social and Policy Studies, Yale University
https://isps.yale.edu/research/data

Motivation

Using digital data necessarily involves software. How do we ensure long-term usability and computational reproducibility? How do institutions address usability challenges such as, license restrictions, legacy software, code rot, and dependencies? At Yale University, a team in the Library is working with the ISPS Data Archive to look into the application of a novel approach to emulation as a potential solution.

Software-dependent research 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

Traditional solution

Author or archive re-writes the code; requires updating the scholarly record with modified code
Or, end-user modifies the code to use new function, if available

EaaSI Funders

EaaSI solution

Keep older software accessible!!!

EaaSI for data curation

- Select existing software environments or build one from software installation media and online software sources
- Add code and/or data, and/or packages and save a derivative “content environment”
- Test the reproducibility
- “Publish” (publicly or privately) the content environment and get a handle and DOI
- Add link to archive’s discovery tool

EaaSI for researchers

- Access and reproduce legacy research using original code
- Seamlessly access *proprietary* legacy software
- Export data into open or more modern formats
- Access legacy databases in real time via secure proxy

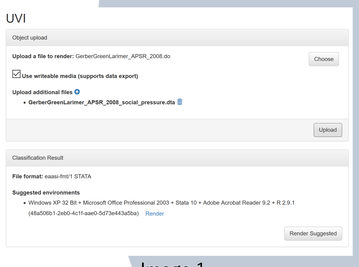

Image 1

- Use **the matching API** from EaaSI’s **Universal Virtual Interactor** (**UVI**) to automatically identify compatible emulated computers for interacting with code and data (image 1)
- Once matched to one or more emulated computers use the **UVI’s interaction API** to automatically execute the code against the data using the original software within a web browser interface (image 2)

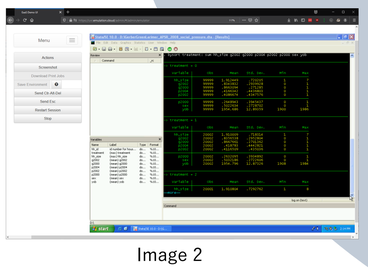

Image 2

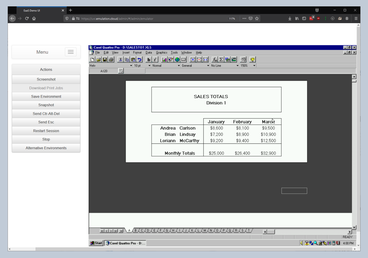

Image 3

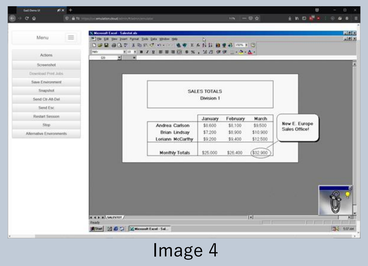

Image 4

- Without original software **even simple data objects** like this Excel v3.0 (.xls) workbook **can be distorted or have missing content**
- In Quattro Pro running in Windows 95 (image 3) the workbook is missing the annotation explaining the data, an annotation that is visible when opened in Excel 97 in Windows 98 (image 4)
- Use the **UVI’s matching API** to confirm archived digital objects have the original software available in the EaaSI network

ISPS Archived Study: Gerber, Alan S., Donald P. Green, and Christopher W. Larimer, 2008, Replication Materials for “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment.” running in Stata 10 within in Windows XP
https://isps.yale.edu/research/data/d001