

Initial Findings from the 2021 Software Preservation Network Hosted Emulation Services Pilot



Ethan Gates
Software Preservation Analyst, Yale University Library
User Support Lead, EaaS! Program of Work



The Emulation-as-a-Service Infrastructure (EaaS!) grant-funded program of work is spending 2021 investigating the feasibility of emulation as a scaled, high-quality, cloud-hosted service using the EaaS! platform.

Exclusive access to a pilot hosted service was offered to organizational members of the Software Preservation Network, with SPN acting as a consortial “customer” and first point of contact while EaaS! oversees deployment, training, support, and data synthesis.

The pilot remains ongoing, but initial findings and feedback have already provided critical insight for future service design and implementing emulation in digital preservation workflows.

Primary EaaS! Staff:

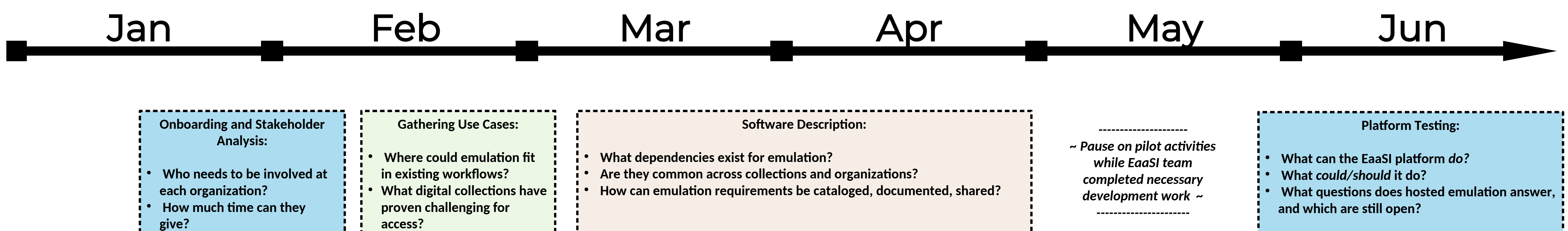
- Community Training and Evaluation (Educopia Institute)
- User Support Lead (Yale University Library)
- Chief Programmer (OpenSLX)
- UX/UI Implementation (Dual Lab)

SPN Community:

- 14 participant organizations (see right)
- 18 organization “leads” (primary point of contact)
- 40 further individuals identified to represent organizations with testing and feedback

Participant Organizations:

- Australia’s Academic and Research Network (AARNet)
- Computer History Museum
- Cornell University
- Georgia Tech
- Harvard Graduate School of Design
- Harvard Library (Houghton)
- Harvard-Smithsonian Center for Astrophysics
- Hesburgh Libraries of Notre Dame
- National Archives of Australia
- New York University
- Penn State University
- Stanford University
- University of Texas at Austin
- University of Virginia



1 Emulation is broadly applicable across collections



Cited use cases for emulation included:

- Design files
- Digital artwork
- Early text-based/word processor formats
- Data reproducibility
- Research software
- Augmented/virtual reality
- Games
- Media archaeology/retro-tech research
- Born-digital government records
- Appraising and/or migrating digital objects – of any kind

2

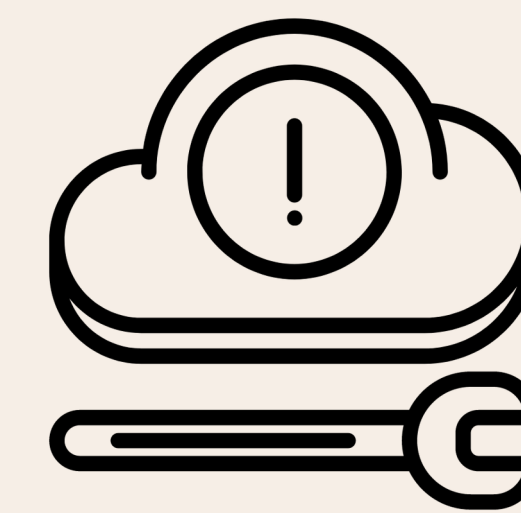


To put emulation services into production at their organization, participants reckoned they would have to consult with:

- University/library/department IT
- Digital Library services
- Legal council
- Collection management services
- Public services, reference desk staff, and/or subject liaisons
- Data services
- Research and scholarly services
- Preservation services
- Metadata/cataloging services

5

Computing is flexible; staff capacity, less so



Preliminary Google Cloud deployment statistics (May 26 - June 15):

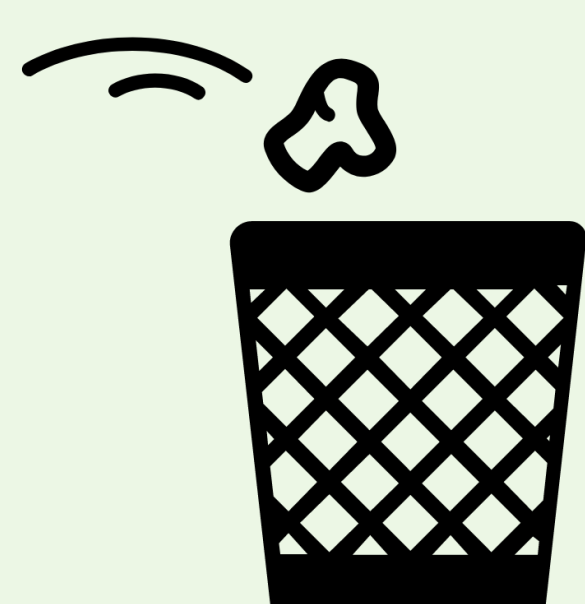
- 1 x 16-core CPU
- 32 potential simultaneous emulation sessions
- 500 GB solid-state storage available
- ~300 unique emulation sessions run
- 65 “Public”, 54 “Private” Environments (~66 GB) saved
- 109 digital objects uploaded (~26 GB)

Challenges to offering a hosted emulation service, identified by the EaaS! team:

- Active time zone spread (Australia to U.S. East Coast) prevents clear maintenance windows
- Large majority of User Support Lead and Chief Programmer’s time directed to hosted service debugging, documentation, and service requests; limited availability for other areas of EaaS! program and/or non-hosted EaaS! users
- Balancing requests for new features with infrastructure/“invisible” improvements

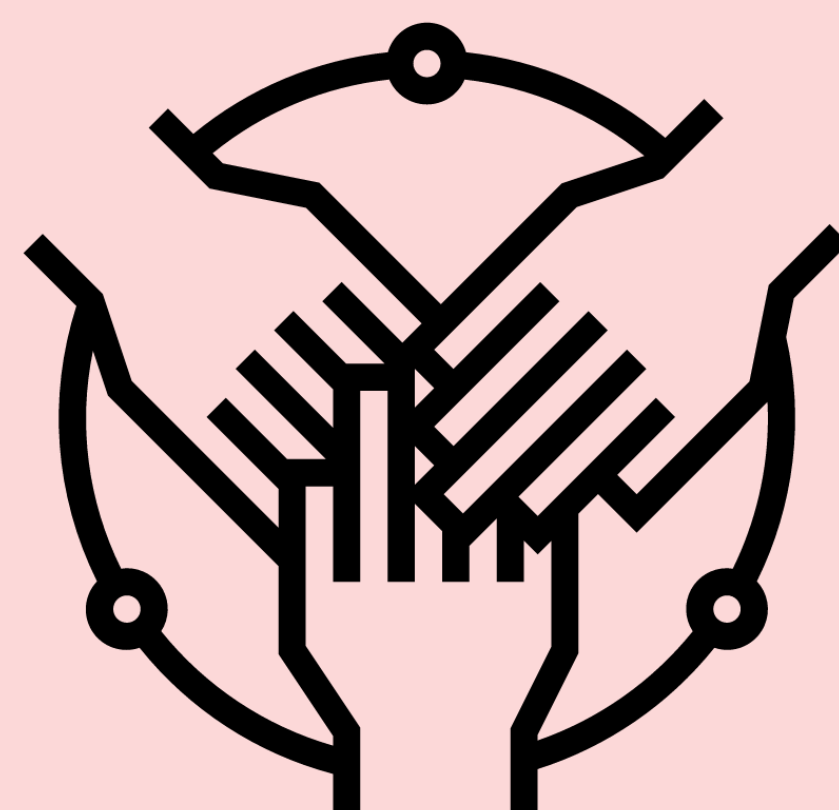
3

Get ready for trial and error



Many sources of documentation were identified as necessary to test emulation services – and not all of them were readily available:

- System requirements and dependencies
- Contributor credits for complex digital objects
- User manuals, README files, or other instructions from the creators of software and born-digital objects
- Photographs of associated media
- Model/training data or benchmarks
- Correspondence related to development, testing, and/or distribution of software
- File format specifications
- Finding aids



Consensus grew on the value of coordinated efforts that are not necessarily in-scope for the EaaS! platform itself:

- Defined and shared acquisition policies for utilitarian software (dependencies) to avoid redundant collecting
- Mapping emulation-related metadata to existing schema and standards – e.g. PREMIS, CodeMeta, MODS, RDA
- Enriching related third-party services with regard to treating software as objects of collection and preservation: e.g. Zenodo, Wikidata

4

More collaborative software preservation efforts are required

Contact

ethan.gates@yale.edu

<https://forum.eaasi.cloud>



Yale