

Fostering a Community of Practice: Software Preservationists and Emulation Experts in Libraries and Archives

Against the backdrop of a thirty-year discourse surrounding the viability of national software preservation strategy, the maturity of emulation and virtualization solutions/platforms combined with increasing software-dependence of archival materials, scholarly research/reproducibility, and creative works has raised the visibility of software preservation as an essential component of any normative digital preservation/curation strategy. Emulation's maturity is evidenced by David Rosenthal's recent report to the Mellon Foundation on emulation, as well as the bwFLA Project's development of support services for emulation installations. The cultural heritage sector's recognition of software's central importance can be seen in CLIR's Software Curation Fellowship program, National Digital Stewardship Residency program placement at Rhizome (rhizome.org), recent posting for Software Heritage Curator at Rhizome, PERSIST, the Software Preservation Network (SPN) [IMLS grant LG-73-15-0133-15], Software Heritage, and results from SPN's Software Preservation in Cultural Heritage study. However, until very recently, only a small group of individuals, organizations and grant-funded projects concentrated on these issues. In other words, while there is increasingly broad acknowledgement of the importance of software preservation and emulation services, the cultural heritage community falls short of broad adoption.

The Software Preservation Network Forum, held August 1st, 2016, was attended by representatives across libraries, archives and museums - many of them new to the issues surrounding software preservation. Attendees challenged themselves to produce a community roadmap for software preservation activities - evaluating priorities, dependencies and their institutional readiness to participate, either by translating these activities into a locally implemented program, or contributing to a national, coordinated effort to tackle some of the roadmap milestones. However, between the feedback, questions and concerns articulated during the SPN Forum and SPN's ongoing coordination with complementary/parallel projects, it has become clear that the creation of a cohort of librarians and archivists developing software preservation and emulation workflows in their local organizations can serve to broaden adoption of these activities as essential components of cultural stewardship.

Project Design: First year activities would involve recruiting a Program Coordinator to establish and coordinate the efforts of the cohort of Software Preservationist librarians and archivists. This role would build an explicit bridge between existing software preservation projects and the selected librarians and archivists in the proposed cohort. The primary investigators will set up an Emulation as a Service (EaaS) sandbox that will be hosted and maintained by the Texas Advanced Computing Center (TACC) at the University of Texas at Austin. They will also coordinate with the bwFLA team (<http://bw-fla.uni-freiburg.de/>), as the latter develops a tenant manager module for the EaaS software. The sandbox will be used by all members of the cohort. The primary investigators will continue their partnership (begun during the 2015-2016 SPN project) with the Cyberlaw Clinic at the Berkman Center for Internet and Society at Harvard Law School, to develop boilerplate Memorandums of Understanding and Service Level Agreements that can be used by members of the cohort to govern use of the sandbox. In addition, Brandon Butler and Peter Jaszi will consult on possible legal implications as the project develops. Once the Program Coordinator has been recruited, likely by the end of second quarter, Year 1, they will begin documenting the EaaS sandbox. The Program Coordinator will then make use of prior work by Euan Cochrane at Yale University to create an emulation training toolkit that members of the cohort will test drive in the first quarter of Year 2. Additionally, the project call for applications to the Software Preservation Network community-of-practice program will be released during the third or fourth quarter of Year 1, with notifications scheduled to go out at close of Year 1.

Second year activities would begin by convening an inaugural, in-person, 2-day meeting where each member of the cohort is asked to review reporting requirements, determine ways projects can complement one

another, and review the emulation training toolkit. After the meeting, the first half of Year 2 will give all members time to complete initial components of the projects, including: documenting their local use case and mapping out their projects in greater detail. The cohort will then complete their projects during the third and fourth quarters of Year 2 and the first and second quarters of Year 3. Meanwhile, the Project Coordinator will conduct briefings with individual members of the cohort and conduct monthly Skype hangouts with all cohort members for shared reporting. Year 3 will see the completion of the cohort members' projects and consist largely of sharing the results of the work completed in presentations and publications. Sharing will take place across relevant domains: computer science, software development, networking, security, computing infrastructure (via TACC and bwFLA) and archival studies and library science (via the project cohort members). Additionally, the primary investigators will seek to partner with information studies educators to develop a version of the emulation training toolkit that can be used in iSchool curriculum, to train the next generation of digital preservation researcher-practitioners. The primary investigators plan to make documentation for all phases of the project available online.

Diversity Plan: The project team intends to recruit for the Software Preservation Network cohort as widely as possible to attract a diverse group of librarians, archivists, and museum curators. In addition, a central goal of the project will be to continue the network cultivation that was started in the 2015-2016 SPN project, building a community membership inclusive of cultural heritage repositories (of all types), software copyright holders, and federal agencies currently engaged in digital preservation, software development and software collection.

National Impact: The project will result in several highly impactful, concrete outcomes. The first will be the cohort of librarian and archivist software preservationists themselves, who will serve as champions of software preservation, practical points of peer reference, and case studies for organizational advocacy. The second concrete outcome will be the emulation training toolkit, to be used by archivists and librarians in their work, and translated by iSchool faculty for instruction in graduate coursework. The toolkit will be strengthened by use case documentation developed by the cohort in their initial activities. Finally, an online, hosted, multi-tenant Emulation as a Service sandbox will have far-reaching impact in lowering the bar of access to emulation technology. This multi-tenant instance to be installed and hosted by the Texas Advanced Computing Center will also provide guidance for administering a multi-tenant EaaS instance at scale.

Other perhaps less tangible outcomes will nevertheless have high national impact, including the expansion of the Emulation as a Service (EaaS) platform documentation. Cohort member selection criteria will aim to produce a cohort of librarians and archivists that vary in terms of their organizational mission and type. As such, the project will also continue the work of the 2015-2016 SPN project by articulating the range of necessary services and support that the Software Preservation Network should provide in order to sustain software preservation and emulation programs in a growing number of libraries and archives.

Budget: The primary investigators estimate that a grant of \$248,555 will be needed to accomplish project objectives (if selected for full proposal, costs may vary based on a more detailed budget). The primary investigators expect 41% will fund a Program Coordinator position to develop training documentation, coordinate with cohort participants, communicate between the participants and the broader profession, and work with the EaaS sandbox developers, 12% will contribute to the costs of contracting with bwFLA to develop the EaaS sandbox for this project, 8% will be allocated for the hosting of the sandbox, and 7% will go towards travel and meeting logistics (venue rental, etc.) for the primary investigators, the Program Coordinator, and the cohort participants to meet in person. Less than 3% will be used towards legal consultants and the primary investigators' compensation. The remaining funds will go towards covering the respective universities' federally-approved modified total direct cost rates.