Abstract

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

California Polytechnic State University, in collaboration with the University of Texas at Austin and Educopia Institute, propose a three-year project (2017-2020) entitled Fostering a Community of Practice: Software Preservationists and Emulation Experts in Libraries and Archives. The findings and community consensus drawn from the Software Preservation Network project (IMLS grant number LG-73-15-0133-15) demonstrated that a broad range of cultural heritage organizations have or will have need for access to legacy software, but that the resources necessary for implementing a software preservation or emulation program do not currently exist. While no one project can supply those resources to all organizations, the currently proposed project will make significant contributions to the pool of resources.

Co-primary investigators, Zach Vowell and Jessica Meyerson, will form a cohort of six cultural heritage professionals interested in piloting software preservation and emulation programs in their organizations. The project will lend support to this work through a Program Coordinator, installation of a web-accessible emulation-as-a-service sandbox, and through the facilitation of cohort communication and interaction. In turn, the community of six will act as a support mechanism for the broader cultural heritage community. The project will also address both the need for formal emulation documentation (tailored for broader training efforts) based on the experience of the cohort of librarians, archivists, curators, while at the same time producing materials which will be used by complementary projects (addressing legal, industry, and technical issues) directly advancing the vision of long-term access to software-dependent cultural heritage.

The pilot projects of the cohort participants will demonstrate measurable progress towards facilitating cultural heritage work and public access to software-dependent cultural resources. Beyond the organizations represented by the cohort participants, the PIs, and Program Coordinator will present and publish about their experiences, as well as conduct at least one emulation training webinar, activities whose impact can be measured by the number of people who attend, read, and participate. Brief surveys will also be used with audiences and participants to uniformly measure the impact of this project on normalizing software preservation and emulation within existing digital preservation programs. Curricular resources developed by the project will be measured by the number of iSchool students that use them. Finally, the impact of the generalized use case of cultural heritage software dependence developed by a graduate student researcher will be measured by the number of projects and organizations that adopt the generalized use case for a template to document their local cases.

Zach Vowell, Digital Archivist at California Polytechnic State University, is the lead applicant for this project. Mr. Vowell's formal collaborators are Jessica Meyerson, Digital Archivist at the Dolph Briscoe Center for American History, and the Educopia Institute, an organization which builds networks and collaborative communities to help cultural institutions achieve greater impact.

Narrative

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

1. Statement of Need

On August 1, 2016, Jessica Meyerson and Zach Vowell (the presently proposed co-primary investigators) convened the Software Preservation Network Forum in Atlanta, GA. There, a group of 36 cultural heritage practitioners (professionals charged with preserving and providing meaningful access to cultural material, often employed by a library, archive or museum) participated in a day-long discussion on actionable next steps for preserving software. As part of the Software Preservation Network project (IMLS grant number LG-73-15-0133-15), the Forum resulted in a 6 year community roadmap[1] that, among other things, prioritized efforts to expand the number of libraries, archives and museums building software preservation and emulation - an area deserving of more ideas and more diversified perspectives - into the everyday practice of digital preservation.

The Software Preservation Network project also produced a bibliography of resources, a survey, and semi-structured interviews, which all helped to confirm that libraries, archives and museums have been collecting and purchasing digital resources for the last forty years that are software-dependent, requiring a stack of software libraries and related dependencies. Most of the survey respondents, interview subjects, and Forum attendees at least implicitly acknowledged (while others explicitly stated) that providing access to electronic resources requires a vast support system of people, processes and machines that exist upstream and downstream of any one cultural heritage institution.[2]

Meanwhile, as emulation and virtualization solutions have matured (see the progress made from Jeffrey Rothenberg's (1999) report on digital preservation[3] to David Rosenthal's recent Mellon-funded report on emulation[4]), and as archival materials, scholarly research, and creative works have become increasingly dependent on their native software, libraries, archives and museums have increasingly emphasized a mode of preservation that requires 1) broader and deeper technical competencies by librarians and archivists that enable them to engage with creators of complex digital objects, 2) a more inclusive notion of the digital preservation

---

[1] For this and all subsequent references within the Narrative, please consult the "Web Links to Relevant Materials" document enclosed with the grant application.

community, and a willingness, an eagerness, even, to partner with cognate communities of practice outside of the cultural heritage sector, such as information and communication technology specialists, 3) and a stronger alignment among libraries, archives and museums that allow cultural heritage practitioners to address scale and to partner more effectively with colleagues in cognate communities of practice.

One of those cognate communities of practice is the Information and Communications Technologies (ICT) industry. In order to be full participants in this conversation, as a field of cultural heritage practitioners we must cultivate broader and deeper technical competences by librarians, archivists, and curators, and a stronger alignment among libraries, archives and museums. This project speaks to both of these goals. Additionally, broadening participation in software preservation is essential for creating a body of good practice that will lay the foundation for a robust mixed methods approach (fair use and licensing) to the legal and policy implications of adopting software preservation as a core stewardship responsibility.

Thanks in part to projects such as the CLIR Report on Digital Forensic in Cultural Heritage[5] and BitCurator[6], the cultural heritage community has established digital forensic workflows that allow practitioners to preserve bitstream representation of digital resources. However, facilitating meaningful access to bitstreams, particularly for more complex digital resources, requires the resource's native software environment, or a similar version of that environment. In contrast to migration (an object-focused access strategy which requires one to address compatibility issues for every digital object), emulation emphasizes preservation of hardware and software configurations that can be used to render many digital resources. Platforms for web-based emulation, including Javascript MESS[7] and Emulation as a Service[8], have removed the need for dedicated computers and made emulation more feasible for a broader range of cultural heritage organizations. However, more research-in-practice is needed to understand what training cultural heritage practitioners need to make use of these emulation technologies in local settings.

This project seeks to contribute to field level understanding through documentation of the pilot software preservation and emulation experiences of cohort members representing different organization types such as public libraries, university special collections, historical society and museums. The activities and documentation produced by cohort members will also serve in parallel efforts to bring software preservation and emulation into mainstream digital preservation practice (addressing specific legal, metadata and technical preservation and access challenges)[9].

2. Project Design

A cohort of six software preservationists and emulation practitioners will comprise the core of this project. Six sites provide enough variation among types of organizations that the project will have a basis for comparison and contrast, while still allowing the project to provide enough travel support that each cohort participant can attend up to 3 events where they can discuss their experiences as members of the cohort. Each participant will implement a pilot software preservation and emulation project within their six respective organizations. The cohort will be selected as a result of an open Request for Proposal (RFP) process. Before advertising, the RFP will be drafted and submitted for review to field experts experienced in successful community building including Nancy McGovern (Massachusetts Institute of Technology), Chip German (University of Virginia), and Patricia Galloway (University of Texas). The RFP will require applicants to address: likelihood of success, letters of commitment, articulation of plan (stated goals, timeline and metrics of success), and statement of meaning (interest in software preservation, stated challenges to the field, ways in which their proposed project addresses stated challenges, personal interest in participating). Once the RFP is finalized, the Project Coordinator will advertise the RFP on listservs, Twitter and the SPN website. Incentives for participation in the cohort include 1) access to the web-based emulation sandbox which requires no local installation, 2) training, communication and logistical support of the Project Coordinator, 3) support from fellow members of the cohort, cultivating professional relationships within a group of emulation project leaders, 4) formal support for problem-based learning and research on the challenges to implementing software preservation and emulation in their local organization, and 5) the opportunity to formally share their experience with their professional colleagues through journal publications, documentation, and training materials.

The Project Coordinator will manage the RFP process that will select the six cohort members, and the Coordinator will continue working with the cohort to support their pilot projects, develop documentation and training materials based on the pilot projects, plan and convene one in-person cohort meeting and otherwise foster the relationships between cohort members so that the community of practice continues to grow beyond the project end date. To facilitate the cohort projects' emulation work, the PIs will implement a web-based Emulation-as-a-Service platform. The hosted emulation sandbox will lower the barrier to participation for a broader range of institutions and allow the cohort members to focus their time on software imaging, policy development, workflows and documentation. Finally, the grant project will provide a travel award to one graduate information studies researcher to study and document two of the cohort projects.

To ensure the development of a functional and sustaining community of practice, the cohort will meet in person prior to their project start date. The Cohort Kick-off meeting will be structured to allow time for the cohort members to become familiar with one another personally and to understand the ways in which their project goals and outcomes overlap. Throughout the duration of the cohort project terms, the Project Coordinator will organize regular monthly meetings and bi monthly discussions on topics pertaining to their work, challenges and updates. Project documentation developed by each cohort participant will be subject to peer review. This approach provides additional encouragement for the cohort participants to be in dialogue with one another and ensures the quality of documentation shared with the broader cultural heritage community. Cohort participants will be given free access and training on how to use several communication tools[10] for the duration of the project in order to encourage them to communication with one another outside of structured or facilitated interactions.

This project also provides travel and registration support for each of the cohort members to attend conferences where they will present, facilitate discussion and actively solicit interest from libraries, archives, and museums. Webinars, screencasts and facilitated discussions resulting in the documentation of new software-dependent use cases will all be recommended as a means of disseminating knowledge gained during the duration of the cohort projects. The project will also produce a Getting Started with Software Preservation and Emulation Guide that can be adapted for use in graduate information science courses. The Getting Started Guide would introduce students to software preservation as a lens through which to examine the more nuanced aspects of metadata, policy, legal and technical challenges inherent in providing long-term access to digital cultural heritage.

In Year 1, the Project Coordinator will be recruited to manage the efforts of the cohort of Software Preservationist librarians, archivists, and curators and assist the PIs with the Cohort Kick-off meeting logistics, implementing a communication strategy for the project and developing a first iteration of the Getting Started Guide for the Cohort Kick-off meeting. The Project Coordinator role will also build an explicit communication bridge between existing software preservation projects and the selected librarians, archivists, and curators in the cohort. The PIs 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. The Fostering a Community of Practice RFP will be finalized and advertised. The selected applicants will be notified in the third quarter of Year 1.

Second year activities would begin by convening the in-person, 2-day Cohort Kick-off meeting where each member of the cohort is asked to review reporting requirements, determine ways

projects can complement one another, and review the Getting Started Guide. After the meeting, the first half of Year 2 will give all members time to complete initial components of the projects, including the documentation of 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 bi monthly discussions with all cohort members for shared reporting. And the PIs will award a travel support grant to a graduate student researcher who will conduct an ethnographic analysis of two cohort pilot projects in progress.

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 archival studies and library science (via the project cohort participants). Cohort participants will also create a training webinar for an audience of local stakeholders (administration, staff, faculty, students, etc.). Additionally, the primary investigators will seek to partner with information studies educators to develop a version of the emulation Getting Started Guide 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 through Open Science Framework (https://osf.io/).

Throughout the project, primary investigators will continue their partnership (begun during the 2015-2016 SPN project) with Chris Bavitz, (Supervising Attorney and Co-Director of the Cyberlaw Clinic, Harvard Law School Professor) and the Berkman Center for Internet and Society at Harvard University, to develop boilerplate Memorandums of Understanding and Service Level Agreements that can be used by members of the cohort to govern use of the emulation sandbox. In addition, the primary investigators will continue to engage with Brandon Butler (Director of Information Policy, University of Virginia Libraries) and Peter Jaszi (Professor of Law Emeritus, American University) as they explore the fair use implications of preserving software for the purposes of long term access to software-dependent cultural heritage.

The primary goal of this grant project is to empower librarians, archivists, and curators to address the key challenges of providing long-term access to software-dependent cultural heritage through the development of a community of practice representing a range of organization types in libraries, archives, and museums. This community of practice will support one another in the development of embedded technical competencies in software preservation and emulation. Another goal of the project is the dissemination of the knowledge necessary for

these competencies - in the form of professional development training, internal organizational advocacy and recommended graduate information science curriculum resources. The project's third goal is publication of a scholarly research article which uses the cohorts as the site of research. Success will be defined through achievement of these goals.

The PIs will measure success and evaluate outcomes in a variety of ways. Benchmarks will be established at the outset of each pilot project including: a scan of software-dependent resources in their collections, documentation of required software applications, and the creation of emulation environments that leverage economies of scale for their organization's software-dependent resources. The quantity and proportion of software-dependent resources to which cohort projects successfully provide meaningful access is one way to measure project success. However, in order to measure the overall impact of broadening experience with software preservation and emulation, the PIs will use the structure of the RFP (goals, project plan, meaning, concerns, perceived challenges) to develop tools to evaluate individual and organization learning outcomes over the course of the cohort projects. The PIs want to contribute to the long-term goal of enabling meaningful access to software-dependent resources, and would argue that a more nuanced understanding of the local challenges facing repositories attempting to implement software preservation and emulation is essential for shaping a national software preservation and access strategy.

The empowerment of the six cohort participants will be multiplied by the number of colleagues with which they can share their newly acquired competencies. The project will produce knowledge related to software preservation and emulation, and the PIs and Project Coordinator will measure knowledge dissemination through attendance, readership, and web statistics. For conference presentations and webinars, presenters will document the number of attendees. For journal publications, authors will record the subscription base and the readership of the journal. For less formal dissemination methods, such as Twitter, less formal, anecdotal, accounts of the knowledge dissemination will be noted. Unfortunately, tracking the impact of the knowledge dissemination through citations and altmetrics will not be possible within the time frame of the grant.

The success of the project in fostering the academic and professional development of the graduate student researcher will be evaluated based on acceptance of an original work of scholarship that uses cohort site observations and analysis, to a peer reviewed journal or conference proceeding. Success in student support will also be evaluated based on a feedback worksheet prepared by the PIs and the Project Coordinator for the student researcher to complete at the end of the research experience. The research travel award should be sufficient

to visit two of the cohort participants for 3 days total (2 days for the first visit, 1 day for the second), and the research produced should reflect the time spent.

The Project Coordinator is expected to carry out his/her tasks at a 70% FTE time allotment. However, this amount of time will be sufficient to address the scope and scale of the project due to economies of scale gained through subcontracting the Project Coordinator services to Educopia. The Educopia Institute already employs three Community Managers currently assigned to work with preservation and library publishing communities. Not only does the Educopia staff have expertise in project management and community coordination, but the cohort members will benefit from the profession-level perspective afforded to Educopia project managers. This profession level view will raise visibility for cohort members' work and project outcomes across the cultural heritage landscape, connecting our project with complementary communities of practice including digital forensics, distributed digital preservation and library publishing. Additionally, the project will benefit from the Educopia communications apparatus which has a solid understand of social media and web content strategies that solicit interest and broaden engagement.

Educopia is well positioned to promptly recruit a Project Coordinator, building on their existing reputation within the cultural heritage ecosystem. Educopia's Director of Strategic Initiatives, Christina Drummond, will commit a portion of her time to cover gaps until a Project Coordinator is hired, and to continue administrative oversight of the Project Coordinator, as well as provide expert guidance in project management, cohort management, and facilitation through the duration of the project. The 70% time also acknowledges the natural ebbs and flows of the project timeline - phases where the Project Coordinator will be heavily involved in project logistics, and phases where the cohort members are implementing their project proposals and the Project Coordinator is primarily facilitating communication between members.

The project PIs will dedicate a percentage of their workweek to this project. Mr. Vowell will use 50 hours per year, while Mrs. Meyerson will contribute 8% of hers. They will spend this time recruiting and onboarding the Project Coordinator, developing the RFP for cohort participants, evaluating and selecting RFP proposals, assisting in installing and testing the Emulation-as-a-Service sandbox platform, logistics and programming for the 2-day cohort kick-off meeting, support and assistance to the Project Coordinator for weekly meeting with the cohort participants, developing criteria for the Research Travel Award, evaluating and selecting Research Travel Award applications, assisting the development of emulation training documentation and sharing activities, reflections and instruction through the project's website and Open Science Framework presence.

The PIs have identified the technical, human resource, and project management risks to the success of the project. First, the performance of the Emulation-as-a-Service sandbox could prove inadequate for the purposes of the cohort's work. By partnering with TACC we have access to an entire datacenter full of hardware, including hundreds of nodes devoted to virtualization, so in most cases some functional piece of virtualized hardware can be easily substituted. If non-virtualized hardware becomes an issue, TACC has on-site vendor hardware support. The reliability of the chosen emulation platform poses a potential risk which can be addressed through a conservative approach to deploying updates, and testing any updates or customizations in a development environment. Secondly, the time and additional training it may take to both 1) survey and 2) forensically image software titles passively collected by each of the cohort's organizations could disrupt the sequence and timing of the project's timeline. In short, in may take longer than we anticipate for cohort participants to prepare for their pilot emulation projects. To accommodate this contingency, the first two quarters of Year 3 are reserved for extra time needed by each cohort participant to complete their pilot project.

Additionally, the RFP process may not garner six proposals, or it may not attract the diversity the PIs desire. Similarly, during the pilot project phase, the PIs will make concerted effort to avoid privileging well-resourced and regularly sponsored organizations.

3. Diversity Plan

In order to diversify the pool of applicants for the software preservation and emulation cohort, the RFP will include specific language to encourage members of underserved communities to submit a pilot project proposal. In addition, the PIs and Project Coordinator will personally invite archivists, librarians, and curators that work in organizations affiliated with groups such as the Historically Black Colleges & Universities Library Alliance (HBCU) and The Association for Specialized and Cooperative Library Agencies (ASCLA) to apply to the RFP process. Both of these organizations focus energies on advocating for information professionals that identify as members of historically underrepresented populations.

The PIs will also incorporate universal design for learning principles into the installation of the emulation-as-a-service sandbox platform. Zach Vowell will consult with Cal Poly's Center for Teaching, Learning, and Technology and the Disability Resource Center to ensure that sight-challenged users and (information professionals and the general public alike) of the platform, as well as those who have non-visual learning styles, will not have to depend on screen representations to interact with the platform. These consultations will take place in Year 1 of the project, so that feedback can be incorporated into the implementation and use of the emulation sandbox during Years 2 and 3.

4. National Impact

The cultural heritage professions are currently in a particularly vulnerable moment with respect to software preservation and digital preservation more generally. While consortia have joined the ecosystem as an effort to leverage economies of scale and collective buying power for digital preservation, the PIs have also encountered skepticism on the part of library administrators regarding line item memberships to organizations and consortia, that, from certain perspectives, look very similar. These vulnerabilities and the means we choose to address them can be informed by libraries' struggle with the growing costs of scholarly publishing. John Wentzer recently characterized the biggest challenges facing libraries and access to electronic resources:[11] "The fundamental hurdle that prevents academic libraries from enjoying the full economic benefits enabled by digital technology is the challenge of collaborating across traditional institutional boundaries. Network technology promises to reduce costs primarily by eliminating the need to build duplicate collections of redundant materials at hundreds of different locations." In order for software preservation and emulation to be normalized as standard digital preservation and access activities, we have to identify and acknowledge our proprietary software collections as a "public good shared by the [cultural heritage] community as a whole" (Wentzer, 2016).

As cultural heritage professions contemplate their future in the face of massive political and economic shifts affecting the pool of state and federal resources available to ensure free, public access to information, we must ask ourselves "On what basis do we have a mandate to preserve cultural heritage?" The answer is trust. Trust is earned through demonstrated competency at the work that you have been entrusted to do. We, the cultural heritage practitioners and organizations, only serve to undermine trust with our publics when fear and uncertainty pushes us to relinquish our privilege and responsibility to preserve.

While software preservation and emulation are only one piece of the digital preservation puzzle, both activities are at the forefront of what libraries, archives, and museums are *thinking* about doing. The PIs want to push that further. The proposed cohort of Software Preservationists and Emulation Experts will learn and lead - helping to normalize this practice by determining what additional work is needed to make emulation just another "tool in the toolbox" and software preservation part of digital preservation business-as-usual. Through its funding of the Software Preservation Network Project, the IMLS has demonstrated its investment in understanding and addressing the challenges for software dependent cultural heritage. Fostering a Community of Practice builds on our previous work[12], and empowers libraries, archives and museums to act as "strong community anchors"[13] and to cultivate trust -

with our users and with ourselves - through further exploration into strategies that facilitate meaningful, long-term access to software-dependent cultural heritage.

Schedule of Completion

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

Year 1 (2017-2018)

	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
Recruit Program Coordinator	■	■	■									
Develop Request for Proposal (RFP)			■	■	■							
Install emulation sandbox environment				■	■	■						
Develop "Getting Started" Guide							■	■	■	■		
RFP finalized and advertised						■	■					
RFP responses reviewed								■	■			
Cohort participants chosen through the RFP process are notified									■			
Project communication strategy is implemented				■	■							
Program Coordinator and PIs introduce cohort participants to existing software preservation project leaders and conduct meetings to solidify connections										■	■	
Plan and program Cohort Kick-Off meeting										■	■	■

Year 2 (2018-2019)

	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
Plan and program Cohort Kick-Off meeting (continued)	■	■										
2-day Cohort Kick-off meeting			■									
Cohort pilot projects are launched			■									
Cohort participants' local use cases are fully documented			■	■	■	■						
Research Travel Award is announced and advertised							■	■				
Research Travel Award applicants reviewed									■			
Research Travel Award recipient schedules pilot project site visits										■		
Research Travel Award recipient makes first visit to two pilot project sites (2 days each)											■	
Program Coordinator conducts weekly or bi-monthly virtual meetings with cohort participants			■	■	■	■	■	■	■	■	■	■
Cohort pilot projects in progress					■	■	■	■	■	■	■	■

Year 3 (2019-2020)

	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May
Cohort pilot projects continue and draw to a close	■	■	■	■	■	■						
Cohort, PIs, and Program Coordinator develop the emulation "Getting Started" Guide based on project experience	■	■	■									
Cohort participants disseminate the results of their projects through articles, presentations, and posters			■	■	■	■	■	■	■	■	■	
Research Travel Award recipient makes second visit to two pilot project sites (1 day each)			■									
Program Coordinator conducts a proof-of-concept webinar on piloting an emulation program									■			
Research Travel Award recipient submits original work of scholarship and develops an overview of common elements									■	■	■	■
PIs work with iSchool faculty to develop curricular content on emulation and software preservation			■	■	■	■	■	■	■	■		
PIs make any of the project's unpublished documentation available on Open Science Framework											■	■

DIGITAL PRODUCT FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions
You must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

PART I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

The digital products created during this project will be the Emulation Getting Started Guide; the journal articles, presentation slides, and posters from the cohort pilot projects; the emulation training webinar; and the graduate student researcher's scholarship based on cohort site visits. The copyright will remain with creators of these products, but in the spirit of building a community of practice around software preservation and emulation, the products will be licensed with a Creative Commons Attribution International 4.0 (CC BY 4.0) license. Any scholarship produced by the project will published through open access journals.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

No ownership rights will be asserted over the digital products by California Polytechnic State University, or by its formal collaborators the University of Texas at Austin, or the Educopia Institute.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

The graduate student researcher scholarship will take the form of an ethnographic analysis of two of the cohort participants. As such, there may be some privacy concerns, both in the observations of the participants themselves, and in regards to the collection materials the cohort participants are working with. To address these issues, the cohort participants that agree to host the graduate student researcher will be given an informed consent agreement that outlines their rights to privacy and notifies them that they can decline to participate at any time.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

The digital products created during this project will be the Emulation Getting Started Guide; the journal articles, presentation slides, and posters from the cohort pilot projects; the emulation training webinar; and the graduate student researcher's scholarship based on cohort site visits. The Getting Started Guide will be a document, and the webinar will be recorded and disseminated as a digital video. The graduate student's scholarship will most likely take the form of a journal article and a use case document.

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

The Emulation Getting Started Guide, journal articles, and use case document will only require modern word processing software and a modern desktop computer. Presentation slides and posters will require presentation software. The webinar will require a video conferencing software such as WebEx, which can export to a digital video format once recorded.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

The Emulation Getting Started Guide, journal articles, use case document, presentation slides, and posters will take the form of an accessible PDF-A file. The recording webinar digital video will be an MP4 file with a H.264 codec

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

All digital products from this project will be peer-reviewed among the cohort participants, the Project Coordinator, and the PIs. The Emulation Getting Started Guide, in particular, will undergo several rounds of iteration (see Schedule for Completion document) at different stages of the project, with different perspectives contributing to the refinement of the document.

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

All digital products will be deposited in California Polytechnic State University's institutional repository, which is backed up by the repository vendor (bepress) as well as on the Cal Poly library's Amazon S3 storage service. Document and file formats have been selected to minimize the need for specialized file migration and normalization, and a non-proprietary audio-only (WAV) version of the digital video will be preserved alongside the MP4/H.264 master version to minimize the risk of intellectual content loss inherent in digital video formats. The digital products will also be stored and made available on the Open Science Framework platform (http://osf.io).

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

The digital products will be described using Dublin Core metadata

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

The export of metadata from the institutional repository has been tested, and the PIs will make certain that only exportable Dublin Core fields are used in describing the digital products.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

The instituuional repository supports OAI-PMH harvesting, and the PIs will integrate the digital products' presence on the Open Science Framework into the project's overall communication strategy.

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The digital products will be made available openly both on the California Polytechnic State University institutional repository (digitalcommons.calpoly.edu), which uses the DigitalCommons software. The digital products will also be made openly available on the Open Science Framework website. Standard web browsers and standard internet connection will be sufficient to access the digital products.

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

One example of a journal article made available on the California Polytechnic State University institutional repository can be found here: http://digitalcommons.calpoly.edu/lib_fac/103/

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL:

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

The ethnographic analysis of the cohort by a graduate student may involve interview data. The methods will be determined the graduate student, and the approximate dates will be April - August 2019.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

Since the ethnographic analysis will involve human subjects, IRB approval will be necessary. The research has not yet been approved, because neither researcher or subject has been identified yet (both will be selected through a competitive process during the project period). A stipulation of the Travel Award will require the graduate student to obtain IRB approval before commencing research.

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

Personally identifiable information will likely be collected during the ethnographic analysis, but due to the small size of the cohort (six cohort members) and the hight visibility of the project, it will be impossible to anonymize the data. However, the graduate student will be required to obtain an informed consent from the two participants in the research, and the participants will be free to leave the ethnographic study at any time.

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

The graduate student will obtain informed consent from the two research participants when scheduling the first visit. At that time, the graduate student will review the informed consent form with the participant and ask them to read over the entire document. If individuals are willing to continue with participation in the research, they will sign the informed consent and those forms will be kept in a physially secure location. Each consent agreement can be traced back to the ethnographic transcripts, notes, and audio recordings.

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

The specific methods will be determined by the graduate student who receives the Travel Award, but it is likely that the data will be generated first by audio recordings, transcribed, and imported into a qualitative data analysis program.

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

The graduate student will not need to create any additional documentation related to the ethnographic analysis of the cohort.

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

Ethnographic research data will be archived in the California Polytechnic State University institutional repository, as well as the Open Science Framework platform. The graduate student will be required to submit the necessary metadata and a data paper to make the data accessible for future users. As with the project's other digital products, the research data will be licensed under a CC-BY 4.0 license, so that future researchers and the emerging software preservation community of practice may freely consult and re-use it. The California Polytechnic State University institutional repository will preserve the data indefinitely and is committed to responsible and sustainable management of submitted data as well as associated descriptive and administrative metadata, by employing a comprehensive digital preservation strategy.

A.8 Identify where you will deposit the dataset(s):

Name of repository: Digital Commons @ Cal Poly

URL: http://digitalcommons.calpoly.edu

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?

The data management plan will be reviewed every 12 months, and again when the recipient of the Travel Award is selected (approximately February 2019). Implementation of the plan will be conducted by the Project Coordinator and the graduate student after the first set of cohort site visits (April 2019).