

Abstract

Software Preservation Network (SPN)

Much of our digital heritage is at risk due to file format obsolescence. With the combined efforts of collecting organizations, archivists, developers of virtualization and emulation as a service platforms, archivists and other information professionals are beginning to seriously consider software preservation as a core practice of digital preservation. For if we preserve digital objects, should we not also preserve the means (the software) by which we can experience them?

To build the foundation of an organization that can preserve the means to experience our digital heritage, the **Software Preservation Network (SPN)** project requests IMLS support for a one-year planning project to organize and host a Software Preservation Network forum to solicit community input and build consensus around future steps; identify potential partners willing to participate in the implementation of the software preservation network; draft SPN's legal and licensing and organizational framework; draft a business plan to submit as part of the implementation grant for sustaining the project beyond the life of grant funding; gather empirical evidence of the need for a software preservation network within the cultural heritage community; and draft technical architecture specifications that consider integration with existing national preservation infrastructure. These activities will all be in service of building community consensus on the best organizational model for software preservation, which the project team intends to then implement after this proposed planning phase.

Archivists, museum curators, librarians, and other information professionals will be the primary audience for this project, especially those professionals who need legacy software to appraise, describe, or provide access to born-digital content stored in proprietary file formats. At the end of the project, this audience will be convinced of the need to preserve legacy software and likewise invested in pursuing a concrete plan for implementing a Software Preservation Network organization.

Progress towards these project goals will be measured by SPN forum attendance, survey participant response rate, and agreements (tentative or otherwise) established with software copyright holders. Tangible deliverables produced by the project include a roadmap for the establishment of an SPN organization, draft organizational vital documents, and a peer-reviewed journal article empirically demonstrating institutional need for software preservation.

Zach Vowell, Digital Archivist at California Polytechnic State University, is the lead applicant for the Software Preservation Network project. Mr. Vowell's formal collaborators for the Software Preservation Network project are **Jessica Meyerson**, Digital Archivist at the Dolph Briscoe Center for American History, and **Carlos Ovalle**, a professional technologist and doctoral candidate affiliated with the University of Texas School of Information, who brings his expertise regarding fair use and copyright to the team.

Narrative

Software Preservation Network (SPN)

1. Statement of Need

The Software Preservation Network (SPN) project team sees software preservation as a critical task towards accomplishing the archival profession's mission to appraise, process, preserve and provide access to primary resources. As early as 1984, archival discourse acknowledged digital objects as software-dependent (Peterson, 1984, p.386). As digital objects have grown increasingly complex, such as building information models (BIM), they can only be accessed in the same software that created them (Mitchell, 1984, p.202-3) or "some closely related software" (Rothenberg, 1999, p.8). The last thirty years have been punctuated by several discrete efforts to collect and preserve software. However, previous efforts have almost exclusively focused on preserving software as information-in-itself, rather than as a dependency for archival appraisal, processing, preservation, and access of born-digital collection material.

In 1986, the Computer History Museum, Smithsonian Institution and the Charles Babbage Institute commissioned David Bearman to write a report that addressed the challenges of creating a national software archives. Limited research interest in software at the time led Bearman to conclude that the development of a software archives and the growth of a dedicated "user community" would be co-determined (Bearman, 1987, p.41). Bearman describes several possible user groups including: 1) scholarly research on distribution, legal protections, and the social norms of developer communities; 2) researchers interested in causal explanations of why certain applications succeeded or failed; 3) "internalist" historians of science or intellectual history; and 4) "system designers" who can not fully understand a system by merely examining its constituent parts (Bearman, 1987, p.41-46). While each of these groups have distinct implications for archival work, Bearman's report overlooks archivists themselves as a possible user group.

In 1990, Margaret Hedstrom reported on her attendance to the symposium on the "Preservation of Microcomputing Software" in which conference participants' interest in software preservation was motivated by two primary goals: 1) the ability to "retrieve and process historical machine-readable data" and 2) to preserve software as a unique source of information (Hedstrom and Bearman, 1990, p.10). Major issues related to these goals included the feasibility of software collection/preservation in light of the required technological resources and "increasing legal constraints" (Hedstrom and Bearman, 1990, p.10). The symposium, hosted by Columbia University's School of Library Service and held at the Arden House Conference Center in the Catskills, was an attempt to "chart a responsible and practical course for the preservation of microcomputing software" and to determine the viability of a "national software collecting consortium" (Bowling in Lowood, 1990, p.5) Topics covered included software issues; machine issues; preservation and format; management, operation and governance; and archival access, outreach, and services (Lowood, 1989). One of the background papers entitled Governance, Management, Operation of a Microcomputing Software Preservation Center, (Jones in Lowood, 1990, p.2) poses a key question to the

contemporary archival community: “When and under what conditions would ‘private’ software become available for...public use?” And although Hedstrom wrote that one of the major goals of software preservation for conference attendees was to access historically valuable, “machine-readable” primary resources, these insights from Jones and Hedstrom have remained heavily deemphasized.

Twenty-three years later, a similarly comprised group of scholars and practitioners came together for “Preserving.exe: Toward a National Strategy for Preserving Software” summit (National Digital Information Infrastructure and Preservation Program, 2013). Unique to the Preserving.exe summit was the inclusion of several projects currently addressing distinct pieces of and overlapping solutions to software preservation and born-digital access challenges including the Olive Executable Archive, a partnership between IBM Research and Carnegie-Mellon University which seeks to “establish a robust ecosystem for long-term preservation of software, games, and other executable content” via a virtual machine streaming platform (Olive Executable Archive, 2015). The project team has identified this and other existing projects as potential SPN partners.

The Internet Archive continues to make significant contributions to software preservation through the work of Jason Scott, Curator of the Historical Software Collection (Internet Archive, 2015) and lead contributor to the development of the JSMESS browser emulation service (GitHub, 2015). In 2013, Scott unveiled JSMESS as a tool for experiencing legacy software systems, allowing a user to select a software title and trigger loading of a disk image of the original software installation media in the browser-based emulator.

The bwFLA: Baden-Wurttemberg Functional Longterm Archiving and Access project (bwFLA, 2015) has developed an Emulation-as-a-Service platform that provides “a standard environment to evaluate and appraise obsolete software”. Steered by Dirk von Suchodoletz, researcher in the Department of Computer Science at the University of Freiburg, Germany, bwFLA has published on their progress towards a proof-of-concept implementation since 2012. Euan Cochrane, Digital Preservation Manager at Yale University Library, has described his local experimentation with V/EaaS platforms in the Library of Congress *Signal* blog, concluding that software licensing agreements present the primary limitation “in our ability to provide access to EaaS” (Cochrane, 2014).

Emerging access platforms discussed at Preserving.exe face the same challenge: their success relies upon the availability of software executables to be run on top of the emulation or virtualization platform. Titles in the Internet Archive’s Historical Software Collection (Internet Archive, 2015) are not packaged with licenses for use, and while the Cabrinety Collection at Stanford (Online Archive of California, 2015) and the National Software Reference Library (NSRL) (National Software Reference Library, 2015) both represent large collections of proprietary software titles, neither of these collections are explicitly or intentionally slated for reuse.

As implied above, the preservation of software presents complex legal challenges, largely dealing with the uncertainty surrounding the copyright implications of preservation activities. That software itself is

copyrightable is assumed based on statute and past court cases. Software is protected by copyright law in the United States and more generally protected by international treaty, which require similar protections be granted to foreign works. Determining the copyrightable elements of a given piece of software can be complicated. The software code may be copyrightable, as can any expression of the software program distinct from the code. Furthermore, some software contains layers of copyrighted content (such as audiovisual elements) with separate copyright owners.

The Digital Millennium Copyright Act added an additional potential obstacle to software preservation: section 1201 anti-circumvention protection for technological copyright protection measures on software. Through the Section 1201 rulemaking procedure the Copyright Office and Library of Congress determine under what circumstances organizations and activities can be exempt from anti-circumvention laws. In the past the copyright office has made certain allowances for copying software. For example, the procedure has in the past allowed the circumvention of technological protection measures when access is prevented by use of a physical dongle when the hardware is obsolete. Similarly, in 2003 the Library of Congress allowed the circumvention of technology protection measures for video games and programs when formats were obsolete and the original hardware or media was necessary for access. However, the process is complex and proponents of such allowances must be argued each time the process occurs. A good deal of uncertainty surrounds the outcome of the rulemaking. For example, the Electronic Frontier Foundation is currently advocating that the rulemaking create a blanket allowance for circumventing protection measures for the preservation of video games. However, the measure is strongly opposed by the Entertainment Software Association, which generally opposed the circumvention of such measures for any purpose. Such debate makes the outcome of the process unclear and difficult to rely upon.

Receiving permission to preserve software from the copyright owner is a potential avenue for preservation in some situations. However, seeking permission also has its associated costs. Furthermore, determining the identity of the copyright holder of a given program and all of its copyrightable elements is not necessarily an easy task. Additionally, the problem of orphan works in which the copyright holder cannot be found, cannot to be determined, or no longer exists (such as in the case of a closed company) also adds difficulty to this process. While the Copyright Office has argued for additional legislation to attempt to address the orphan works issue, such legislation has not been forthcoming in recent years, and it is as yet unclear if such legislation would be helpful for preservation efforts.

Collaboration around software preservation activities, then, becomes necessary within the existing U.S. copyright framework. Cultural heritage organizations must seek more direct engagement with the copyright owners of software titles (most often, the software companies themselves). This engagement could take many different forms, and SPN will make the investigation and establishment of this engagement a priority of its work.

During the one-year planning grant phase, the SPN intends to address the three most immediate and pressing roadblocks for the re-use of legacy software in archival work: demonstrating the empirical basis for software preservation need (advocacy, titles, etc.) within the cultural heritage community; establishing a partnership between cultural heritage institutions, software companies and federal agencies; and articulating a business model and organizational structure through which software licensing and legitimate, ongoing non-profit re-use can operate harmoniously.

2. Impact

To meet these needs, the Software Preservation Network project team requests the support of IMLS to convene a software preservation forum that will build community consensus around the best model for a Software Preservation Network organization.

In the interest of achieving consensus and progress toward the ultimate goal of a Software Preservation Network, the project maintains three additional, though subordinate, goals: 1) produce research findings on the professional need for software preservation as a means to ensure long-term access to born-digital content, 2) produce research findings on existing organizational models for providing access to copyrighted content, and 3) build relationships with software copyright holders.

The progress made during this one-year planning grant will benefit every archivist in the country, and many more librarians and museum curators along the way. Access to information will be increased dramatically. Anyone, indeed, who needs to open a file with software that is not currently commercially available, will benefit from the work of this project. The consensus-building and research that will result from this project will build a foundation and clear path towards preserving software for future archival work and access.

Given this context and given the limited scope of the one-year planning grant, the project targets will consist in attendance to the SPN forum, research findings, and relationships with software copyright owners. The project team can measure the attendance in the number of attendees and the existing roles of the attendees' institutions within the digital preservation community. The target goal for attendance will be 50 information professionals, 3 representatives of software copyright holders, and 3 copyright legal professionals. At least half of the attendees will be employed by institutions actively involved in the digital preservation community through participation in digital preservation infrastructure, community-based digital preservation initiatives, digital preservation publications, and other modes of participation.

During the 2014-2015 academic year, researchers from Georgia Tech, Bryn Mawr, Yale University and the University of Texas surveyed members of the professional archival community regarding the access challenges associated with born-digital collection material. Of the total number of participants that responded to the question concerning obstacles to providing access to born-digital materials via virtualization/emulation services, 51% identified "access to software licenses and installation media" as an obstacle. This project's needs-based assessment attempts to build on that study by identifying resource gaps for the cultural heritage

community in terms of software collection and preservation, and determining the necessary incentives for broad participation in the network. Success of the needs-based assessment will also be measured by participant response rate and the degree to which our sample represents the full range of institutional type, size, and mission.

The research data will be used to produce an argument that the project team will publish in a peer-reviewed professional journal. Upon publication, the article will serve as tangible evidence for the value of software preservation for archivists and other information professionals, evidence which may prove useful in negotiations with software companies, in garnering institutional support at the highest academic levels, and possibly even in legal settings such as the Copyright Office rulemaking procedures.

Finally, the project team will measure the relationships with software copyright owners with agreements (tentative or otherwise) towards a shared interest in establishing an SPN organization.

Additional tangible deliverables include an implementation roadmap for a software preservation solution focused on providing continual access to legacy software and drafts of vital organizational documents such as a charter, by-laws, and network membership agreements. The roadmap and model documents will not only point to a solution, but will also provide a clear plan around which the digital preservation community can rally and engage institutional support, which will be essential for economic sustainability.

The project team will sustain the progress achieved during the one-year grant project by immediately moving to an implementation phase. If necessary, project participants (which by that time will include more people than the current project team) will apply for additional grant funding to seed the establishment of a Software Preservation Network organization. If this route is taken, one of the core activities of such an implementation grant will be to develop a sustainable business model for a Software Preservation Network, so that grant funding will no longer be necessary. Again, institutional support, from colleges, universities, and federal agencies, will be vital to this economic sustainability - and the present one-year grant proposes to establish the means and opportunities to develop such support.

And ideally, software copyright holders will participate in SPN negotiating special licensing or licensing exemptions for cultural repositories currently processing born-digital collections dependent on legacy software. Such direct involvement in SPN serves as a corporate responsibility (CR) initiative. Organizational studies scholars have classified CR initiatives into functional and cross-functional categories. Cross-functional initiatives tend to address areas such as development (“building social capital”, “building infrastructure”, “improving general welfare”), environment (“reduce the production of negative environmental externalities associated with producing the organization’s good and services”), and corporate governance (“create new rules regulating how to operate as well as the generation and/or disbursement of residual profit” including “create binding guidelines outlining expected corporate behavior”) (Griffen and Prakesh, 2014). SPN participation allows software copyright holders to address each of these areas. Partnerships are one of the key

collaborative mechanisms for CR initiatives, often resulting in multi-party compacts that focus on achieving a specific objective. IMLS support will provide dedicated resources to a national outreach and advocacy effort to build such compacts. To reiterate the sentiments of digital conservators and software historians alike, 'no one institution can successfully tackle software preservation and born-digital access' (Bowling, 1990, p.5; Ensmenger, 2009, p.88-91).

3. Project Design

The Software Preservation Network project would allow the project team to assess community interest in developing each team member's local efforts into a national Software Preservation Network. Each member of the SPN project team has undertaken software preservation activities within their own organization in order to establish a series of preliminary activities and baseline metrics for a SPN organization. Currently, in their home institutions at the University of Texas at Austin and California Polytechnic State University the project team has collected a total of 985 software titles, for which they have existing campus licenses. Based on the collection, organization, and description of these software titles, as well as original research into software preservation and digital preservation initiatives, the project team presented a model for the Software Preservation Network at Society of American Archivists Annual Meeting 2014, which resulted in interest from several potential partners and ideas for alternative software licensing models.

The project team will apply previous work and experience towards SPN project goals, including: organizing and hosting a Software Preservation Network forum to solicit community input and build consensus around future steps; identifying potential partners willing to participate in the implementation of a SPN organization; draft of SPN's legal and organizational framework, based on research into existing frameworks for providing access to copyrighted content; draft business plan to submit as part of the implementation grant for sustaining the project beyond the life of grant funding; empirical evidence of the need for a software preservation network within the cultural heritage community; and a draft technical architecture specification that considers integration with existing national preservation infrastructure.

The project team has identified a series of activities required to achieve the above stated project goals. First, the team will plan and conduct an inaugural advisory committee meeting. The meeting will serve as a kick-off meeting for the project where primary investigators will review the terms of the award and solicit immediate feedback from advisory committee members regarding software industry contacts, project coordinator recruitment, forum venue, outlines of vital documents and participant recruitment for the needs-based assessment. Guidance on how best to recruit software copyright holder partners will be a particularly emphasized in this meeting.

Next, the project team will fully develop a survey tool that will profile existing born-digital collections, assess current access strategies for born-digital material, identify existing software collection and preservation efforts, and map participation in existing national digital preservation infrastructure projects. Investigators will

follow best practices for survey design. The project team will test the survey questions for reliability (consistency across measurements) and validity (degree to which the survey tool measures what it is intended to measure) by distributing a pilot survey prior to distributing the assessment across all potential respondent pools.

Based on the survey responses, the project team will conduct semi-structured interviews - the reliability of the research findings will be strengthened by this mixed methods approach. The semi-structured interview questions will be derived from preliminary analysis of survey data survey questions, and will allow the project team to explore software preservation or implications for software preservation in local contexts.

Once the survey responses and interview data have been gathered, the project team will conduct qualitative data analysis. The survey questions will provide options for users to write in a response if available multiple choice options do not address their current practice. Summative analysis/plots based on multiple choice responses will be analyzed using pandas, an open-source Python library for data analysis (pandas, 2015). All free text survey responses and interview transcripts will be analyzed using Dedoose. Prior to coding the bulk of free text responses to survey questions and interview transcripts, the project team will conduct a pilot coding exercise to test for intercoder reliability and identify any changes that need to be made to the codebook.

During the second half of the grant period, the project team will plan and convene a Software Preservation Network forum. This is the key project deliverable. The forum will take place over two days immediately before or after the 2016 Annual Meeting of the Society of American Archivists, and will bring together potential network participants to generate community participation and build consensus around a concrete roadmap for organizational implementation. By establishing connections with potential partners early in the funding term, forum programming will focus on organizing partners into sprints that address facets of each challenge identified by the needs-based assessment. Sprints will also be focused on soliciting feedback regarding drafts of vital organizational documents, including: a charter and by-laws, MOU (or other document) governing relationship with network and software copyright holders, and a SIP agreements for depositing software into the network.

Although the project team believes that all of the activities are important steps in establishing the viability of the software preservation network, the team has built in flexibility to the project design by prioritizing project activities so that the actual drafting of vital network documents (rather than mere feedback solicitation) could, if necessary, be folded into the activities of the forum as a way to engage potential project partners in tangible first steps. Aware that survey design, distribution, and analysis can take extensive time, the final research portion of the project design may forego the semi-structured interviews (which would eliminate the added transcription services cost and potentially remove the Dedoose costs).

4. Diversity Plan

The project team intends to communicate announcements of the SPN forum widely in an effort to attract as diverse a forum attendance as possible. In addition, a primary goal of SPN will be to cultivate a Network 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.

The project team will distribute the survey tool to as many cultural heritage repositories as possible. These repositories provide the expertise necessary to preserve and describe proprietary software titles they collect for the purpose of supporting access to their existing born-digital collections. The project team will depend on a representative sample for the needs-based assessment in order to make data-informed recommendations for SPN implementation activities (i.e., technical architecture, governance, business model, most needed software collecting areas). Participant recruitment for the survey and semi-structured interviews will involve snowball method as well as targeted recruitment. Mixed recruitment strategy will ensure that data gathered is representative across several axes of repository classification including: size, type, collecting focus, and number of years they have collected born-digital materials.

5. Project Resources: Personnel, Time, Budget

Zachary Vowell is the primary investigator (PI) for this grant proposal. Mr. Vowell is the Digital Archivist at the Robert E. Kennedy Library, California Polytechnic State University, and is a member of the library's faculty. Mr. Vowell will devote 8% of his time to the Software Preservation Network project, and the grant proposal requests 8% of his annual salary as compensation for the project's work.

Jessica Meyerson is the Digital Archivist at the Briscoe Center will serve as co-primary investigator throughout the project term. Ms. Meyerson will devote 8% of her time to the Software Preservation Network project, and the grant proposal requests 8% of her annual salary as compensation for the project's work.

The primary investigators' role in the project will consist of overall project management, shared supervision of the graduate research assistant coordinator position, designing, convening, and leading the Software Preservation Network forum, developing survey questions, and writing journal articles and other communications based on the results of the forum and the findings of the project's research.

The primary investigators have chosen to hire a graduate student to serve as the Graduate Research Assistant Coordinator. The primary investigators believe this coordination role is appropriate for a graduate student because 1) the level of responsibility expected from this position and the required skills provide fertile professional development for a graduate student and 2) the activities and goals of this project could be complementary to existing research areas across cognate fields including information studies, history, and public policy. The Graduate Research Assistant Coordinator will be responsible for Software Preservation

Forum planning logistics, such as comparing venue quotes for room reservations and catering during the two-day forum, as well as contribute to the analysis of qualitative data gathered during the need-based evaluation for proprietary software preservation within cultural heritage organizations. Travel arrangements for our advisory committee members and event promotion will also be the responsibility of the graduate student coordinator. The primary investigators propose to pay the graduate research assistant a monthly base salary of \$1,833 for 20 hours per week. In addition, and as a component of hiring a graduate research assistant, the primary investigators request the cost of tuition for full-time enrollment during the fall, spring and summer semesters for the 2015-2016 academic year.

Travel

As the SPN forum, the key project deliverable, will seek to bring together people invested in software preservation, the project requests travel support for the project's primary investigators (Vowell and Meyerson), the Graduate Research Assistant Coordinator, and three of the project's advisory committee members.

The presence of the project's advisory committee members (Mr. Hanson, Dr. Lee, and Dr. Lowood) is vital to establishing the forum's ability to draw a large attendance of information professionals interested in solving the dilemmas of software preservation. One co-primary investigator (Carlos Ovalle) will not receive funding for travel support, because he is offering his consultant services on a pro bono basis.

Supplies, Materials, and Equipment

The Software Preservation Network forum will take place over two days. The project team requests support to adequately accommodate 50-75 participants, provide audio-visual support, and offer snacks and refreshments during the course of the forum. In preparation for the SPN forum, the project team will design print materials including posters, programs, summary documents, and sign-in sheets.

Another deliverable for this project will be an article that the primary investigators hypothesize will establish once and for all the archival community's need for software preservation as a core component of its work. Publishing the article will provide highly visible, tangible evidence for the value of software preservation for archivists and other information professionals. The project team requests support for the article processing costs associated with open access journals, which the team has chosen to target to ensure that there is no fee associated with accessing the summative analysis and findings that result from the project's research.

Survey data and semi-structured interviews will be the two data sources upon which the peer-reviewed article will be based. In order to effectively perform content analysis on the transcripts from the semi-structured interviews, the project team requests support for transcription services.

Once transcribed, the primary investigators can begin analyzing the data. Based on experience in a previous study, the primary investigators request support to use Dedoose, a qualitative data analysis software created by social science researchers in the 1990s at the UCLA Center for Culture in Health.

Bibliography of References

Software Preservation Network (SPN)

Bearman, D. (1987). Collecting Software: A New Challenge for Archives & Museums. *Archives and Museum Informatics* (Archival Informatics Technical Report [August 1985]), 1(2), 1-73.

Bowling, M. (1990). A Center for the Preservation of Microcomputing Software: Access and Outreach. *Henry Lowood Papers (SC1187)*. Dept. of Special Collections and University Archives, Stanford University Libraries, Stanford, California.

bwFLA (2015). *bwFLA — Emulation as a Service*. Retrieved from <http://bw-fla.uni-freiburg.de/>

Cochrane, Euan (2014). Emulation as a Service (EaaS) at Yale University Library. *The Signal: Digital Preservation*, August 20, 2014. Retrieved from <http://blogs.loc.gov/digitalpreservation/2014/08/emulation-as-a-service-eaas-at-yale-university-library/>

Ensmenger, N. (2009). Software as History Embodied. *IEEE Annals of the History of Computing*, 31(1), pp. 88–91. doi:10.1109/MAHC.2009.16

GitHub (2015). *JSMESS GitHub Wiki Home*. Retrieved from <https://github.com/jsmess/jsmess/wiki>

Griffin, Jennifer and Prakesh, Aseem. (2014). Corporate Responsibility Initiatives and Mechanisms. *Business & Society*. 53(4), 469-70

Hedstrom, M., & Bearman, D. (1990). Conferences. *Archives and Museum Informatics*, 4(1), 10–15. doi:10.1007/BF02770069

Internet Archive (2015). *The Internet Archive Software Collection*. Retrieved from <https://archive.org/details/software>

Jones, C. (1990). Governance, Management, Operation of a Microcomputing Software Preservation Center. *Henry Lowood Papers (SC1187)*. Dept. of Special Collections and University Archives, Stanford University Libraries, Stanford, California.

Mitchell, W. J. (1996). Architectural archives in the digital era. *American Archivist*, 59(2), 200–204.

National Digital Information Infrastructure and Preservation Program (2013). Preserving.exe: Toward a National Strategy for Software Preservation. Retrieved from http://www.digitalpreservation.gov/multimedia/documents/PreservingEXE_report_final101813.pdf

National Software Reference Library (2015). *Welcome to the National Software Reference Library (NSRL) Project Web Site*. Retrieved from <http://www.nsrl.nist.gov/>

Olive Executable Archive (2015). *Olive Executable Archive*. Retrieved from <https://olivearchive.org/>

Online Archive of California (2015). *Guide to the Stephen M. Cabrinety Collection in the History of Microcomputing, ca. 1975-1995*. Retrieved from <http://www.oac.cdlib.org/findaid/ark:/13030/kt529018f2/>

pandas (2015). *Python Data Analysis Library*. Retrieved from <http://pandas.pydata.org/>

Peterson, T. H. (1984). Archival Principles and Records of the New Technology. *The American Archivist*, 47(4), 383–393. doi:10.2307/40292703

Rothenberg, J. (1999). Commission on Preservation and Access, Council on Library and Information Resources, & Digital Libraries. *Avoiding technological quicksand: finding a viable technical foundation for digital preservation : a report to the Council on Library and Information Resources*. Washington, DC: Council on Library and Information Resources.

Original Preliminary Proposal

The Software Preservation Network

Much of our nation's digital heritage is at risk due to file format obsolescence. Recent archival literature has emphasized documentation strategy or significant properties as the most viable approaches to preserving content stored in proprietary file formats. However, with the combined efforts of collecting organizations, developers of virtualization and emulation-as-a-service (EaaS) platforms, archivists and other information professionals are beginning to seriously consider software preservation as a core practice of digital preservation. For if we preserve digital objects, then should we not also preserve the means (the software) by which we can experience them?

Software preservation is best accomplished through a **network**, in part, because the software we aim to preserve represents a network of machines, people and processes, and in part because "the universe of potential holdings is far too large for a single repository to be expected to manage." The co-primary investigators (Jessica Meyerson, Carlos Ovalle, and Zach Vowell) are requesting National Forum Grant support for the first phase of The Software Preservation Network (SPN), whose activities could ultimately include collaboration on software collection, preservation strategies for binaries and source code, descriptive schemas for software, and preservation planning for cloud-based subscription software/SaaS. Furthermore, SPN could combine the work of V/EaaS researcher-practitioners with a distributed preservation model bolstered by formal agreements and relationships between software companies, member repositories, and federal agencies. The National Forum Grant project represents the first steps towards these ultimate outcomes, allowing SPN to pursue a social, economic, legal and preservation infrastructure which will benefit society by contributing to the long-term access of our digital heritage.

Pre-Grant Progress

Each of the project's primary investigators has initiated software preservation efforts within their own organization in order to establish a preliminary activities and baseline metrics for other repositories interested in participating in the SPN's software preservation activities. Currently, in their home institutions at the University of Texas at Austin and Cal Poly State University the project's primary investigators have collected a total of 985 software titles, for which they have existing campus licenses. Members have also established a survey tool to focus collecting efforts, and which captures information on software-enabled digital object creation and storage practices.

In August 2014, the primary investigators presented a preliminary model for the Software Preservation Network at the 2014 Annual Meeting of the Society of American Archivists. Following the presentation, the primary investigators compiled a list of interested professionals representing the spectrum of possible participant groups:

- Member repositories that will provide the expertise necessary to preserve and describe proprietary software titles they collect for the purpose of supporting access to their born-digital collections.
- Software companies that can provide support by sharing legacy software titles and negotiating special licensing or licensing exemptions for cultural repositories currently processing born-digital collections whose contents cannot be ascertained through any other means.
- Federal agencies currently supporting digital preservation and software authenticity efforts can provide a central hub within the network for authenticating disk images of proprietary software titles.
- Practitioner-researchers that will contribute case studies and investigations into the issues surrounding the operation, acquisition activities, and use of a viable software preservation network, thus strengthening the knowledge base on which the digital preservation community depends.

Acknowledging the broad participation on which a successful implementation of SPN would rely, the primary investigators have assembled an advisory committee which will supply diverse perspectives representing many of the participant groups above:

- Cal Lee - Associate Professor, School of Information and Library Science, University of North Carolina at Chapel Hill
- Henry Lowood - Curator for History of Science & Technology Collections and Film & Media Collections, Stanford University Libraries
- Amy Stevenson – Archives Manager, Microsoft

The advisory committee will play a central role during the planning grant, and have already provided valuable feedback and support for the primary investigators. In addition, the primary investigators have identified the following activities and outcomes which grant funds would allow them to undertake and achieve.

National Forum Grant Activities

An early activity would be the organization of an inaugural advisory committee meeting. Soon thereafter, we would conduct an empirical study that would evaluate the diversity of need for proprietary software preservation within cultural heritage organizations, resulting in evidence-based recommendations for network implementation (including legal/licensing and organizational framework). The primary investigators plan to disseminate findings from the study through presentations at a digital preservation conference - which also serves as a venue for engaging professionals from within archival studies, as well as cognate fields within the information studies community, about participation in SPN. Indeed, we are requesting the support of IMLS, in part, to convene a software preservation summit in conjunction with such a digital preservation conference. The summit would consist of working sessions that would focus participants' energies toward drafting SPN's organizational framework, vital documents and technical architecture which the empirical study's recommendations will launch. We will also submit a journal article to publications such as *The American Archivist* (which has solicited a manuscript submission based on the primary investigators' presentation at SAA), providing a more thorough articulation of the software preservation challenge and SPN's answer to that challenge. Additional activities to be undertaken include establishing a public communications strategy and drafting documents vital to the network's operation, such as a charter, by-laws, and network membership agreements.

Outcomes

1. Formal report detailing findings and subsequent recommendations from an empirical assessment of software preservation needs within the cultural heritage community
2. Journal article articulating the cultural heritage preservation challenges that SPN is meant to address
3. Identification of all organizational partners willing to participate in the implementation of the software preservation network
4. Software preservation summit resulting in vital organizational documents, including charter by-laws and membership agreements, and a white paper focused on implementation
5. Detailed budget for pilot SPN (to be undertaken during a later implementation grant)

Budget

The SPN primary investigators anticipate that a grant of \$80,000 will be needed to accomplish project objectives. The primary investigators expect 48% will fund a graduate student assistant to work on data gathering, analysis, and project management, 8% will contribute to the costs of convening the summit, while 5% will enable the primary investigators to travel to the summit. The remaining funds will go towards covering the respective university's federally-approved modified total direct cost rates.