

Software Sustainability And Preservation: Implications For Long-term Access To Digital Heritage

Jessica Meyerson
Briscoe Center for American History
University of Texas at Austin
2300 Red River, Austin, TX 78712
j.meyerson@austin.utexas.edu

David Rosenthal
Stanford University
518 Memorial Way
Stanford, CA 94305
dshr@stanford.edu

Euan Cochrane
Yale University
344 Winchester Avenue
New Haven, CT 06520
euan.cochrane@yale.edu

Zach Vowell
California Polytechnic State University
San Luis Obispo, 1 Grand Avenue,
San Luis Obispo, CA 93407
zvowell@calpoly.edu

Natasa Milic-Frayling
UNESCO PERSIST Programme
Computers Science Department
University of Nottingham, UK
psznm@nottingham.ac.uk

ABSTRACT

Digital content and data require software for interpretation, processing, and use. This requirement raises the issue of sustaining software functionality beyond its prime use, when it is fully supported and maintained. Virtualization and emulation are two techniques that can encapsulate software in its functional form; furthermore, emulation has recently gained traction as a viable option for long-term access to digital objects. At the same time, archivists, librarians, and museum curators have begun concerted efforts to preserve software that is essential for accessing the digital heritage. In this context the members of the panel will discuss relevant work that they have been involved in to address the goal of software sustainability and preservation.

General Terms

Digital preservation frameworks; Infrastructure, systems, and tools; Domain-specific challenges; Case studies, best practices

Keywords

Software preservation, distributed digital preservation networks, partnerships, digital cultural heritage, fair use, copyright

1. THEME

1.1 Software sustainability

As demand for a specific software application declines, it is typically not economically viable for the software vendor to continue maintaining it for use within contemporary computing environments. Yet, the software may be needed at the later time to access digital content that requires it or which is most authentically rendered using it.

Virtualization and emulation can hold software static in time from a maintenance perspective, minimizing economic burden to the software vendor, while also enabling the preservation of access to the software in a form that is usable in contemporary computing environments.

David Rosenthal's 2015 report to the Mellon Foundation acts as a watershed moment for the viability of emulation, and simultaneously articulates the challenges faced by emulation practitioners [6]. Among them are technological and legal aspects that the UNESCO PERSIST project attempts to address [10], and equally complex issues of standards and established practices that need to be revisited in view of new technical capabilities provided by emulation tools and frameworks such as the bwFLA project from the University of Freiburg [3], the Olive project from Carnegie Mellon University [5], and the JSMESS work being advocated for and implemented by the Internet Archive [8].

1.1.1 Technical feasibility

The 2009-2012 KEEP Emulation Framework project [2] provided the first "simple" framework for libraries, archives and museums to use in providing access to content via pre-configured emulated or virtualized computers. Since 2012 the bwFLA Emulation as a Service (EaaS) project has been demonstrating remote access to emulated and virtualized environments, accessible via a simple web browser. In addition to a "generic Application Programming Interface (API)" for a variety of emulation and virtualization tools, the team has implemented emulation and virtualization as part of operational archival and library workflows. Through the use of sophisticated virtual disk management and "derivative environment" management, the bwFLA EaaS framework can support a highly distributed software preservation ecosystems that may be attractive to software IP holders. The bwFLA team has also demonstrated automated large scale migration of content using emulated software to perform migration via a "simple" interface.

Mahadev Satyanarayanan demonstrated the feasibility of installing, searching, and streaming VMs with executable content, making it easy to share and instantiate legacy digital artefacts [4]. Natasa Milic-Frayling [1] demonstrated the hosting of legacy software in a contemporary (commercial) cloud platform. Combined with the scalable format transformation services, built as part of the SCAPE EU project [7], software virtualization provides the full range of cloud capabilities for rendering digital content, from 'authentic' content using original software applications to migrated content using contemporary applications.

1.1.2 Content preservation practices

New capabilities are prompting memory institutions to revisit current practices and standards in content preservation. This is essential for both the quality of preservation and the development of a market for supporting services. With the increased volume and computational complexity of digital artefacts, it is expected that combined emulation and migration approaches will become a common practice. Moreover, the increased acknowledgment of the feasibility and scalability of emulation tools and services is beginning to shift how cultural heritage institutions approach their preservation strategies. In particular, the economics of emulating content when necessary now presents an attractive alternative to the policies of migrating all digital content over time. Thus, it is important to reflect on the implications of emulation viability for memory institutions, including preservation standards and the development and support of new services.

1.2 Software preservation

Memory institutions and software vendors possess software collections that present valuable digital heritage and require due care. Furthermore, a growing number of digital objects are software-dependent, i.e., software is essential for their faithful rendering and use. Through research and informal discussions with various stakeholders, the Software Preservation Network (SPN) project [9] has demonstrated and verified that information professionals are confronting such software dependence now.

Both researchers and practitioners have engaged in projects and initiatives to create essential resources and establish effective practices in software preservation, from metadata frameworks to technical capabilities required to create software images from obsolete storage media. In addition, Yale University, SPN, and the Society of American Archivists' CAD/BIM Task Force have pursued relationships with software rights holders in order to resolve legal impediments to preservation practices. Generally, the preservation community continues to evolve their practices and strive for more comprehensive and complete technical registries to support and coordinate software preservation efforts.

2. PROGRAM

The panel will be structured to include a brief introduction of the overall topic, followed by panelist reports on software preservation initiatives over the past 2 years. The reports will be followed by a discussion on the topic with the audience, moderated by Maureen Pennock, Head of Digital Preservation at the British Library. Both the reports and discussion will entice the audience and panelists to reflect and take part in the discussion of several aspects of software preservation and legacy software services:

Community coordination. How to leverage ongoing developments towards a coordinated effort to collect and preserve software essential to access our digital heritage?

Legacy software licenses. How to approach legal issues related to commercial and orphan legacy software?

Economic sustainability. What evidence is required for market viability? Can cultural heritage institutions make a business case to rights holders for preserving software?

Technology infrastructure. Implementation, management, and access to legacy software services.

Standards and best practices. Development of guidelines for cultural heritage institutions that need to re-use software.

3. PRESENTERS AND PANELISTS

Zach Vowell is the Digital Archivist at the Robert E. Kennedy Library, California Polytechnic State University, San Luis Obispo, and co-primary investigator on the Software Preservation Network (SPN) project. Vowell will discuss the input from the research and cultural heritage community, gathered by the SPN team, and the implementation roadmaps developed at the SPN forum in August 2016.

Euan Cochrane is the Digital Preservation Manager at Yale University Library. He has worked with emulation tools since 1990s and collaborated with the University of Freiburg on digital forensics since 2011. Cochrane is currently working on the citation framework for complex and interactive digital objects,

funded by DFG/NEH. He will present on emulation use to establish legal access to CD-ROMs from the Yale Library and preservation of canonical versions of installed software environments.

Jessica Meyerson is the Digital Archivist at the Dolph Briscoe Center for American History, University of Texas at Austin, and co-primary investigator on the SPN project. Meyerson will address the legal issues confronted by the SPN team, as well as the implementation roadmaps developed at the SPN forum in August 2016.

David Rosenthal is co-founder and Chief Scientist of the LOCKSS Program at the Stanford Libraries. He will discuss report conclusions and activities following its release [6].

Natasa Milic-Frayling is Chair of the Technology and Research Workgroup of the UNESCO PERSIST Programme and Prof. and Chair in Data Science at the University of Nottingham, UK. She will present on the PERSIST Programme and the plans to create a platform for hosting operational installations of legacy software.

4. ACKNOWLEDGMENTS

Outcomes and results from the Software Preservation Network project were funded by the Institute for Museum and Library Sciences (IMLS), grant number LG-73-15-0133-15.

5. REFERENCES

- [1] Abe, Y., Geambasu, R., Joshi, K., Lagar-Cavilla, A., and Satyanarayanan, M. "vTube: Efficient Streaming of Virtual Appliances Over Last-Mile Networks". Proc. of the ACM Symposium on Cloud Computing 2013, Santa Clara, CA
- [2] "About the Emulation Framework". *KEEP Emulation Framework [EF]*. The KEEP Project, 2011. Web. 13 July 2016. <<http://emuframework.sourceforge.net/about.html>>
- [3] *bwFLA Emulation as a Service*. University of Freiburg, n.d. Web. 13 July 2016. <<http://eaas.uni-freiburg.de/>>
- [4] Milic-Frayling, N. "Sustainable Computation—Foundation for Long term Access to Digital". Workshop on Open Research Challenges In Digital Preservation, iPRES 2013, Lisbon, Portugal
- [5] *Olive Archive*, Olive Archive, 2015. Web. 13 July 2016. <<https://olivearchive.org>>
- [6] Rosenthal, D. S.H. "Emulation and Virtualization as Preservation Strategies". Report prepared for The Andrew W. Mellon Foundation, October 2015, New York, NY
- [7] *SCAPE: Scalable Preservation Environments*. The SCAPE Project and the Austrian Institute of Technology, 2014. Web. 13 July 2016. <<http://www.scape-project.eu>>
- [8] Scott, Jason. "Still Life, With Emulator: The JSMESS FAQ". *Internet Archive Blogs*. The Internet Archive, 31 Dec. 2013. Web. 13 July 2016. <<https://blog.archive.org/2013/12/31/still-life-with-emulator-the-jsmess-faq/>>
- [9] *Software Preservation Network*. The Software Preservation Network Project, 2016. Web. 13 July 2016. <<http://www.softwarepreservationnetwork.org>>
- [10] van Gorsel, M., Leenaars, M., Milic-Frayling, N., and Palm, J. "Evaluation and Strategies of Digital Preservation & UNESCO's Role in Facing the Technical Challenges". 2nd Ann. Conf. of the ICA, UNESCO-PERSIST, 2014, Girona, Spain