Software Preservation Workshop Proposal
Prepared by Sellam Ismail, Software Curator
March 14, 2005

DRAFT

Synopsis

This proposal outlines a Software Preservation Workshop to be held at the Computer History Museum in the Fall of 2005. The purpose of the workshop will be to discuss issues of historic software collecting and preservation, including:

- Software Collecting Criteria
- Software Archiving and Access
- *Transcoding* Issues: Problems and Solutions
- Intellectual Property Laws and Historic Software
- Examples of Software Archives

The nature of the discussions will be both technical as well as philosophical. The intent of the planned discussions is to establish best practices for institutional software collecting and to also tackle technical software preservation issues.

Invitations will be sent to individuals and groups throughout the world who are currently examining software preservation issues and developing pertinent strategies and solutions. The target audience for both presenters and attendees is:

- Museum curators
- Computer historians
- Historians of technology
- Corporate historians
- Software developers
- Intellectual property specialists
- Interested members of the hobbyist community active in these issues*

Other likely participants might include major computer companies around the Valley, as many would find an interest in beginning to think about how they are preserving their software heritage. Suggestions

* The major work done to date in the area of software preservation has been carried out by hobbyists

include IBM, HP, Sun, Apple, Oracle, Microsoft, SGI, etc.[†] The event would also be of interest to members of the FOSS (Free/Open Source Software) community. For example, Linus Torvalds might be invited as a participant since preserving Linux should become more relevant as that operating system approaches its 15th year.

Goals

The primary goals of this workshop are as follows:

1. Begin to establish best practices for the preservation of software over the next century.
2. Provide a recurring forum for individuals and groups with an interest in the area of software preservation to meet and exchange ideas for preservation and future research, and to present and discuss solutions.
3. Establish the Museum as a key institution in the area of historical software preservation and research.
4. Set an agenda for a more formal software preservation symposium in 2006.

Format

The workshop is intended to be a formal event but with a limited attendance of perhaps 25-50 individuals. The ideal length of this event would be one full day and one half day, with 5-6 presentations on the first day and 2-3 presentations on the second day.

The Hahn Auditorium would provide an ideal environment for the workshop. A banquet lunch could be served on the first day, with no requirement for any food service on day two.

The total estimated budget for the event including staff time and food is $5,000.

Proceedings

Each presentation should be captured on audio or videotape for posterity. The tapes can be transcribed and the transcriptions can be

[†] These participants could also be potential sponsors of the event

published on the Museum's website as proceedings of the event. We could also make the taped presentations available from the Museum's website for streaming download.

Topics

The following are proposed topics for the workshop.

Software Collecting Criteria

The reasons for collecting and preserving software, and the criteria by which an institution selects and collects software will be discussed.

- What should be collected/preserved and why?
- Should the endeavor be to collect and archive every software title ever published or just significant titles?
- How does the collection and preservation of software differ from the collection of more "traditional" artifacts?

Software Archiving and Access

- Discussions on how to catalog, archive, and provide access to software, with an emphasis on physical storage issues.

- Discussion on the tools and techniques needed to assist the software history researcher. Different research requires different ways of accessing software. For example, someone researching the evolution of coding styles may not need the same information as someone researching the evolution of real-time software control.

- Discuss ideas for "proactive preservation". How has the collection and preservation of software been neglected, and will this affect our ability to research early software? What steps can be taken by current software developers to ensure their software is properly preserved for future historical research?

- What are the costs of "proactive preservation" to a corporation? Do the benefits of preserving the history of a corporation's software development outweigh the costs? What steps, if any, have corporations or organizations taken to ensure preservation of their software assets for future study? Has retention of

Software Archiving and Access (Continued)

software beyond the commercial interests of the organization (e.g. preservation) been incorporated into the software life-cycle? What are the incentives and/or disincentives for organizations to incorporate historical preservation issues into software development?

- Do changes in the software marketplace (e.g. "Open Source") inherently facilitate preservation and future study?

Transcoding Issues, Problems, and Solutions

- What problems have been encountered and solutions developed to preserve software stored on media that is either deteriorating or obsolete? Discussions will include specific examples of tools and techniques from individuals experienced in the practical issues of data and media conversion.

- Technical discussions regarding the interpretation of storage media and formats. Knowledge of various file and filesystem formats must be re-learned and preserved in order to be able to interpret and recover software and data from various media. Is it important to preserve the physical and logical structure of the media itself, or is just the content important?

- Technical discussions regarding the characteristics of various storage media and implications for the preservation of software, both in terms of how to deal with existing media (e.g. properties of various types of magnetic tape, discs, paper) and the characteristics and longevity of archival media.

- How do we ensure today's software is preserved for posterity? What methodologies need to be developed to ensure the migration of software archives to newer media as current archival media becomes obsolete? What are the criteria for software archival media?

Intellectual Property Laws and Historic Software

- What challenges and obstacles to software preservation do copyright and intellectual property laws present for the software archivist?

- What constitutes "abandonware"? What is the legal status of software developed by a company no longer in business where the ownership rights are not at all certain?

Ideally, an individual with direct experience with these issues such as Brewster Kahle, Lawrence Lessig or an attorney with the Electronic Frontier Foundation would be available to present on this issue.

Examples of Software Archives

Existing software archives throughout the world will be discussed, including at institutions (museums, universities, etc.), in private collections (corporate and individual), and online repositories.