Multithreading Software Preservation and Emulation

Workshop | August 2, 2019

8:30am - 4:30pm
Austin Central Library
Special Events Center (first floor)
710 W. Cesar Chavez St.
Austin Texas, 78701

Feedback Survey for the Code of Best Practices for Fair Use in Software Preservation:
https://american.co1.qualtrics.com/jfe/form/SV_bOtxhCBPUBVdzlr

Shared Meeting Notes

Table of Contents

Public workspace folder is available here. Headers link to workspace folders for each activity.

Welcome & Logistics

Link to this doc: bit.ly/swpres-emulation
Hashtag for this event: #SPNATX (use with #EaaSI or #FCOP or #software or...)

Icebreaker

[begin typing here]

Overview of software preservation & emulation

Some challenges of software

Software preservation:

- Policy and procedures for the collection, description, and maintenance of files and information necessary to operate software for the purpose of study or to provide access to digital objects
- Issue of incompatibility

Emulation:

- old computers on new computers. Emulators are software. Parts you need: Physical Environment that includes Hardware, Bootable System (as disk image), and Additional Software (as disk image). An emulation environment replaces all three components.
- **Emulation** recreating o preservation: obsolete libraries,

Questions: What is Disk Imaging? A process by which you use certain hardware and software tools to copy digital storage. It copies the files and storage infrastructure. The goal is to have a computer can read, interpret bit-for-bit. A disk image is different from a bitstream, they will have different hashes and properties.

A: Does a disk image have the key (information) to make an emulation?
The disk image doesn't give you the whole context which the emulator needs. There is additional information external to the disk image that also needs to be gathered.
Ethan answer: There is a bit of information that you may be able to extract from the disk image itself (DOS formatted, HFS, etc.) that could be useful, but these are mostly hints about the time period within which the disk/software was created. But most other components are external factors and contextual metadata that is gathered.

Why preserve and emulate software?

- (Works of art?)
- Addressing obsolescence

Legal framework (overview):

Code for Best practices for fair use: stabilizing and describing, documentation in operation, access for research, networked access, preserving source code.

Related resource: Preservationists Guide to the DMCA Exemption
(*link includes lots of additional links to guidance, blog posts, DMCA exemption itself*)

Does fair use apply to you? Collections are open to the public, lawfully acquired/licensed, implement reasonable digital security for preservation activities, have a public service mission, have trained staff/volunteers that provide LAMs services

Question: What is reasonable digital security?
Answer: DMCA Pre-reqs. If you have to break digital locks, then you have to be careful but there is wiggle room.

EaaSI Overview and Demo

- Example EaaSI OS documentation
- Example EaaSI software documentation

FCoP Overview

Participants are learning and sharing what capacity-building for software preservation through emulation looks like, documenting the impact of the activities on each stage of the data curation lifecycle within their institutions.

Outputs include various strategies to disseminate knowledge and experience gained, through reports, blog posts, community guidance, and workshops (like this one!). Working on what is needed for description and metadata, preservation planning, etc. What forms are needed? How do policies and workflows need to change? What different metadata is needed? What needs to happen locally vs. what can be done at consortial level?

(Ethan): Many of the resources I use to help guide students through using unfamiliar legacy systems are in my repository at https://GitHub.com/EG-tech/emulation-resources

Scenarios for (Re)-Use and Access

Group report out:
Resources: UVA FCOP Software Questionnaire V0.3

Group 1:
Share a scenario: Pagemaker book project file that contained a publication.
Interesting point: Wanted to check all of the boxes in the survey. Creating an intake survey for research may inform what staff does on the backend.
Creating a form for staff to use for assessment and intake. Potential hardware dependencies and consideration of artistic intent.

Group 2:
Share a scenario: Example of "Labanotation " for documenting dance. Understanding this may allow other dances to understand the movement within a particular dance.
Interesting point:

Group 3:
Share a scenario: Deck log in Navy archives. Has to be eventually transfered to NARA. These are being submitted in proprietary forms. Unsure of what software these documents are created in.
Interesting point: purpose is to provide access to these. Found that most everything is important to them as they don't know the creation environment nor future use. Little is known due to federal security concerns.

Group 4:
Share a scenario: Computer History Museum. IP Lawyers interested in seeing how a software title runs. Important to determine what it is that the attorneys are looking for in order to meet their needs. Lawyers don't know what they need either.

Interesting point: Trying to recreate an experience for a court case. How to recreate that experience when the user doesn't know what they need.

Group 5:
Scenario: Working with an artist from Puerto Rico to understand the project. THinking of a broad range of users. Challenges in that data is backed up to a number of carriers. Variety of tape storage and CD-Rom. Incomplete information about creation operating system and browser environment.
Interesting Point: Identifying slices of the enviornmnets to use and/or recreate.

Group 6:
Scenario: a researcher interested in studying the history of video games. Reviewing a broad variety of formats to gather contextual information. Hardware associated with running software. Trying to figure out what it is the researcher is interested in. If emulating, how far do we take emulation? Emulation also serves as a quality assurance/verification that the local instance is operating as expected. Emulation is one tool in access and preservation.
Interesting points: Batteries have changed slightly over the years. Documenting these slight changes.

Group 7:
Scenario: software-based artwork. Stakeholders are museum conservators and users/museum visitors. The work is a software-based audio work. Goal is to preserve the software and maintain the functionality as intended by the artists. Lots of moving parts.
Interesting Point: Lack of a workflow for accessioning this type of artwork. There are lapses in documentation for understanding this artwork, source code, etc.

Software Inventory

Group 1:
System requirements, hardware dependencies. Local notations of "boxed software" e.g. software that one could purchase and pull from the shelves. Additional contextual information. Mission of the institution will determine what is captured about the object. E.g. one institution might be interested in physical space that an item occupies compared to another institution might not be as interested in maintaining the item.

Group 2:
Discussed context. Prompted discussions around extent of questions. E.g do you really need information about all of the information such as does the software require a

“pointing device”. This will vary depending upon the institution's needs. Might need the extent to which a title/versions were produced.

Group 3:
Sometimes carriers are overwritten with information different from what the media indicates. Questions relating to donor use. Is the “thing” labeled on the disk/carrier actually representative of what is stored on the carrier?

Group 4:
Have a lot of contextual information. Has a serial number. Doesn’t have any information about the intellectual context. Common visual cues (e.g. headphones) may not be understandable in the future.

Group 5:
The tensions of what information you need based on the use. For example, making the software available for emulation - you may need additional information that you did not initially have in order to run the software. If possible, getting a user to illustrate how they used the software.

Group 6:
Example given of an item arriving without much information included with the item. Persisted to uncovering additional information due to frustration, “being mad” about not having enough contextual information.

Group 7:
Covered a lot that has been covered by other tables. Talked about looking for last date title was supported. Some obsolete titles are still being supported. WOuld be helpful to know about known bugs and issues about software. Learning about these issues via anecdote or other communities. Questions about what to retain - what represents the authentic representation of the work.

Envisioning Future Services

Group 1:
Shared observation:
Status Quo: reading room access

Challenges: Advocating for it to be a library priority. Maintaining the staff support necessary.
Solutions: Work with collaborators

Group 2:
Shared observation:
Current state - able to preserve collection material and software via Archivematica
Challenge: No way to provide access to any of the software (EaaS may help a solution with that). Institutional challenges - even if you have go ahead from management, getting it to be a priority is a challenge, esp tech challenges and need for IT support
Solutions: Collaborating within the digpres community and EaaSI project as well as greater software literacy would help

Group 3:
Shared observation:
Current state/Status quo: Needs to be approved by IT and administrators - can't be too burdensome on budget or infrastructure
Challenge: EaaSI infrastructure needs to expand so more can join
Solutions: Have something approved that gets past those barriers, and join consortium that helps do the heavy lifting for them.

Group 4:
Shared observation:
Current state/Status quo: Not being able to provide access, but wanting to provide access
Challenge:
Solutions: Need to present a proof of concept to allow administrators locally to see and understand the service and what it looks like institutionally. Starting small might also be helpful.
(Note: lots of these workshop exercise templates, and other exercises/webinars/etc. are available on the SPN website**)**

Group 5:
Shared observation:
Challenges: Can't show demand because no one can get to the collections, but you can't get to the collections to show demand. Catalog out of date, need to demo need to get funding/budgets
Solutions: Demo shared collaborative work across the community, do proof of concept even if speculative
Future state: world where everyone wants what we have!

Group 6:
Shared observation:
Status quo: Group responsible is a small part of an org - limited funds, limited staff, services, etc**.**

Challenges: Giving up ownership, staff time and expertise
Solution: Open knowledge sharing
Future State: Workflows, emulation embraced and used as a preservation method at the highest organizational levels

Status quo:

- Currently preserving information in proprietary systems
- No one is doing this at home - the need to become so evident to everyone
- Want assessment of libraries who need this service
- ASpace and cataloging system are all seperate discovery systems
- Digitization
- Reading rooms and physical materials
- Policy framework does not support most types of objects
- No long term funding
- Service limited to a few campuses
- Document delivery via digital library platforms and 1-on-one email reference
- Sharing Git repository for long term preservation in insitutionla repository
    - Working it into their process, educational process
- External services; Outsourced to unit's IT group
- Lack of institutional will
- Student interest
- Preserving student work (games)
- Service providers have NOT provided services
- Project-driven ad-hoc efforts make work stop and go
-

Future state:

- Preservation of museum documentation systems that can we used on a workstation in the reading room
- Users are accessing information in human-readable formats
- Reading room terminlas with reference and reading assistnace
- vendor/cloud consortial infratructure
- Digipres basics + tech access training
- Using our services to view digital art and art websites
- Training: catalogers, referene staff, reading room staff, all proficient in software preservation and emulation
- Users actively using emulated enviornments to do primary research
- We are comfortable providing access to emulated software and files. We know who can have access, when and to what
- Flexible - not so different solutions for so very similiar objects and tasks

- Library digital systems provided and maintained by librarians & IT jointly. Highly collaborative and responsive.
- Allows collaboration
- Uses technology but low echnical barrier to entry
- The ability to build and evolve
- A centralized service that is provided by the library with the input of all the repositories on campus
- Students can use the service with no training
- Tech for capture and access may be outsourced
- Access via networked, web-based software-as-a-service
- Organizational policy explicitly committeed to permanant, ongoing access to software and the objects that depend on it (as a matter of the public trust)
- The software developer/creator makes their platform + codebase acessible and non-proprietary so that submitting for
-

Challenges:

- Policy or regulatory environment
- Service provided needs to be commercially viable or well-funded
- Shared union environments
- Skills using available environments
- Vendor relationships and proprietary control
- Lack of diversity
- Digital literacy - decreased knowledge, day to day apps are more automated
- Funding
- Objects that don’t fit on a reading room workstation
- Scalability
- Interoperability

Solutions:

- Educate ourselves and policymakers about our rights and neds
- Teach digital literacy
- User-centered design with diverse users
- Emulation as a service
- Outreach to vendors
- Develop standards and best practices
- Design with affordability in mind
-

Very good q about membership and subscription service realities out of looking at solutions from these exercises. Educopia doing some work on "membership fatigue" and what these means for those working on the consortial side to better support/identify/combine services to help alleviate some of these very understandable concerns. Economy of scale is desirable, but resources are still finite. EaaSI definitely thinking of this quite a bit as they move forward with their work

Interested in exploring the idea of an endowment around some of this work. We need different ideas to aid with this work.

Six Month Plan

Examples from 6 month plans:

-Planning a digital preservation conference around software!
-Figuring out what formats you have, and process and emulate 1 small collection to demonstrate value
-Funding to start a digital preservation workstation - going to collaborate with other depts to formulate a policy, figure out a collaborative arrangement with other depts
-Software preservation not necessarily a priority at a small repository, but plan to talk to other colleagues, play with sandbox, figure out your needs, and plan to get higher on the priority list
-Goals to ID stakeholders: at the institution convince folks to be stakeholders as well. Advocacy institutionally. Also a goal to get environments up to demo.
-Install locally an EaaSI instance - design a wrapper for the landing page, get a kiosk set up for the "middle age" of computing. Not all fun and games - need to have marketing, sell gen admin tickets (museum use case), curators will make a sign, test software in EaaSI
-Personal 6 month example: Research what games have been preserved, get to know folks in the field who are preserving and archiving video games, ask questions about games and context for preservation

Wrap up
Delta, Plus, Minus Exercise: Plus/Deltas are typically done in last 10 minutes of a meeting or activity. Projects and organizations that really look to maximize their learning from the practice also use it mid-way

through a longer meeting or work session. It can be particularly helpful as a coaching tool when a meeting is going offtrack – it will help recalibrate the team around the meeting's objective and potentially help them identify a new objective.

What worked really well?
-Being able to play
-Facilitation really great - broad overview in the morning, then progressively more intense activities =)
-Table group were diverse both in career and also from different institutions.

Deltas/Things to improve
-Needs to be some kind of working definition thing - terms used in every day practice that facilitators used completely differently
-Would be nice to hear about implementation in a collection and to hear their use case.
-Sometimes words on the slides and words on the worksheets that aligned with the slides were different

Would people attend a forum on this in 4-6 months? (Fair amount of hands for "yes")