How Does It **Really** Work? Software Preservation and Emulation in the Archives

Lauren Work and Tracy Popp
SAA Archives*Records
8/5/2019

Session Structure

- Introduction and overview of the current landscape of software preservation
- Use Case: UI
- Use Case: UVA
- Use case comparison, shared conclusions & opportunities discussion

Definitions for today

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

Emulation: Use of software to recreate computer hardware and operate computer systems and software on host infrastructure

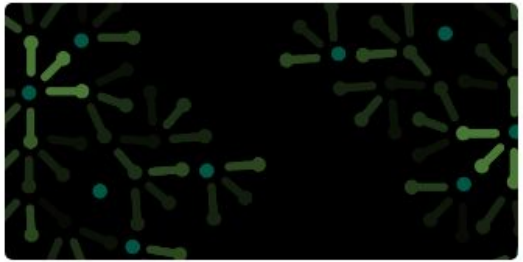
Curatorial: responsibility and practice of appraising, acquiring, describing and providing access to materials

Software
Preservation
Network

Current Affiliated Projects

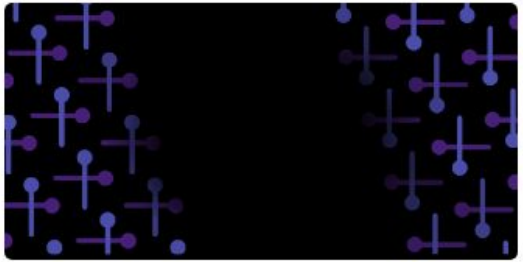
EaaSI

Scaling Software Preservation and Emulation Infrastructure (EaaSI) is developing software to enable organizations running the bwFLA Emulation as a Service (EaaS) software to network with each other and share resources. This allows participating node-hosts to share software and pre-configured software environments that others in the network can then use to ensure access to their preserved digital data.



FCoP

Fostering a Community of Practice: Software Preservation in Libraries, Archives, and Museums (FCoP) centers around the experience of a cohort of six organizations as they design and implement their own software preservation projects. Each organization created a Project Team consisting of staff roles that represent critical perspectives in the implementation of programmatic software preservation and emulation.



BPs for Fair Use

The Code of Best Practices in Fair Use for Software Preservation (BPs for Fair Use) helps ensure that the subjects, products, and tools of scholarship will continue to be accessible despite evolving technology. The code will express a consensus view of how fair use—the legal doctrine allowing the use of copyrighted materials without permission from the copyright holder under certain circumstances—applies to core, recurring situations in software preservation.

Fostering a Community of Practice

Software **Preservation** Network

- Guggenheim Museum
- University of Arizona
- Living Computers Museum & Labs
- Georgia Tech
- University of Illinois at Urbana Champaign
- University of Virginia

University of Illinois at Urbana-Champaign

- Digital Preservation Services situated within the Preservation Service unit
- Serving collecting units from a central location
- Most often work with special collections units

Digital Preservation @ University Library

- Born-digital media reformatting lab
- Collections content recovery from fragile, obsolete computer media
- Wide variety of content recovered - each requiring respective rendering software

Initial Engagement with Emulation

- Common roadblocks to implementing emulation:
- Resource intensive
  - Dedicated staff time
  - Storage and access infrastructure
  - Security concerns
- Not particularly scalable
- Despite challenges, still very beneficial
  - May offer user experience more closely linked to creator's intent
  - Can process content in situ without modifying source environment

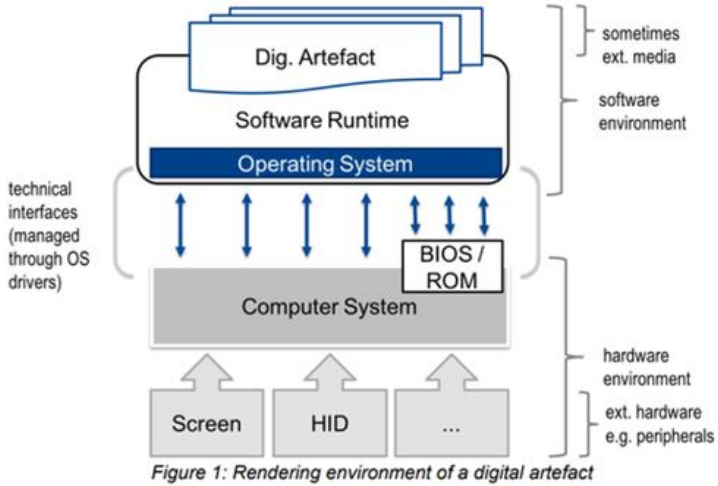


Figure 1: Rendering environment of a digital artefact

From the Tate Museum

Involvement with FCoP

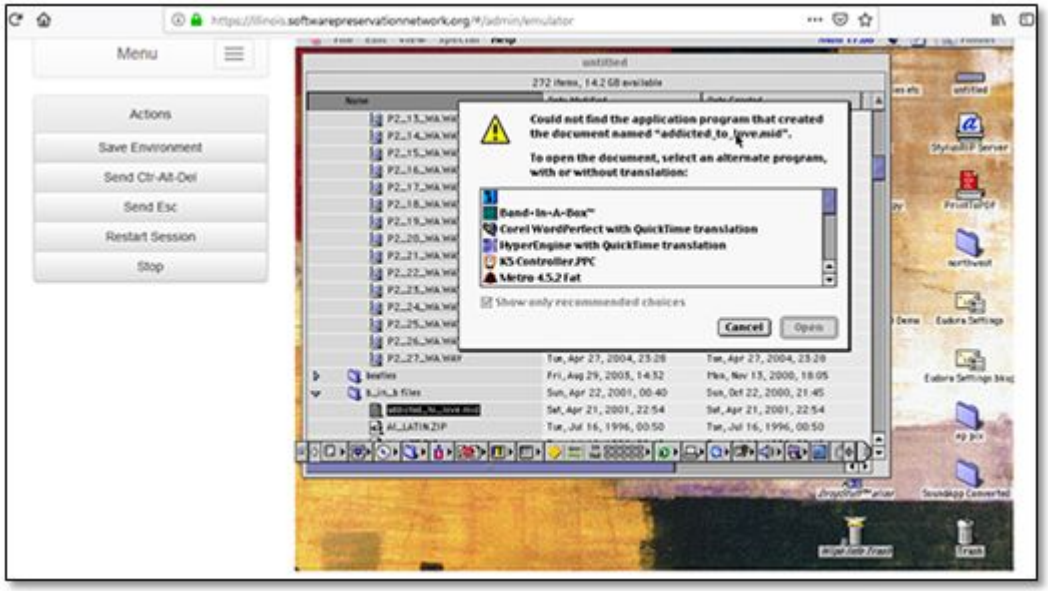
Project Goals:

- Improving preservation and access for composers collections
- Developing workflows
- Piloting and prioritizing emulation research

Collections and Rescoping

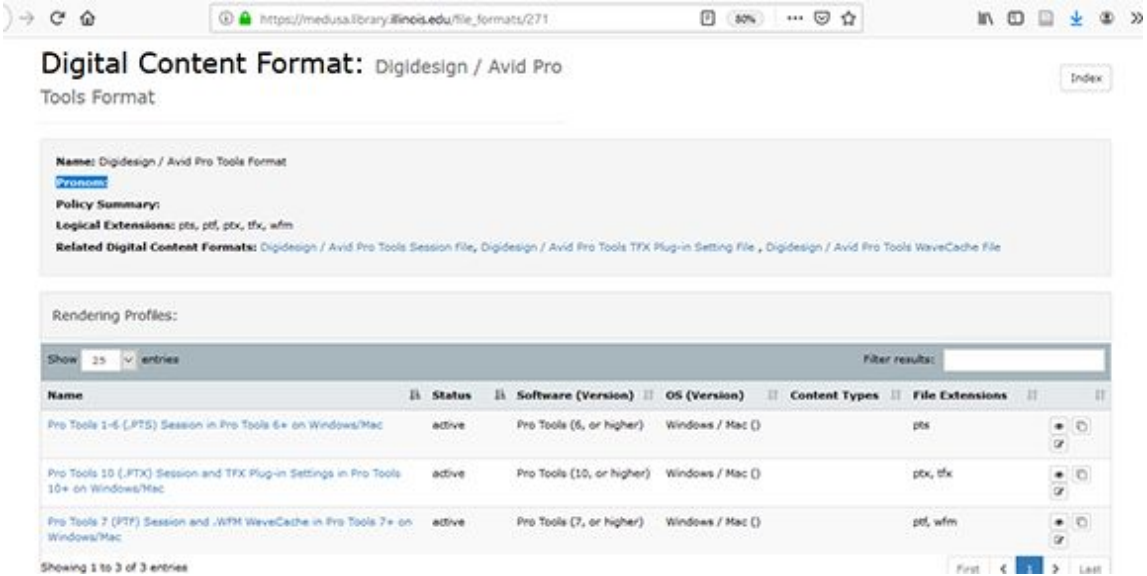
- Initially centered around born-digital collections of three Illinois composers
- Scope: 1992-2012
- Scaled back to focusing on one for EaaSI:
- Michael Manion:

  Macintosh PowerBook 3400c
- Used .E01 source disk image to generate EaaSI environment
- Emulated environment as processing tool



Scaling and Documenting Curation

- Digital Content Format Registry
- Publicly viewable:
- https://medusa.library.illinois.edu/file_formats



https://medusa.library.illinois.edu/file_formats/271

Scaling and Sharing the Workload

- Addendum to deed of gift for electronic records
- Guidelines for Donating Digital Materials
- Workflows
  - Installing and testing legacy software
  - disk imaging QA

Software Install Workflow (excerpt)

About the Software
SoftwareTitle
SoftwareVersion
SoftwareDeveloper
SoftwarePublisher
YearPublished(orDateOnLabel)
System Requirements: OS
System Requirements: RAM
System Requirements: Other
System Requirement: Free Text Notes
Estimated Time To Complete this activity:
Software Source
Where did you find the software?
Where are the installation files stored? (if you downloaded installf files that worked and weren't malware-laden or buggy, please upload them to Box)
Install Files Structure (extracted directory, ISO disk image etc):
Is an authentication method required?
If so, what authetication method is required (key, physical dongle, key and "phoning in" to an authetication server, etc.)
Do you have the correct authentication method?
Were you able to successfully authenticate?
Estimated Time To Complete this activity:
LocalInstall: HostComputerEnvironment
Host Computer Installation Tag:
Host Computer Name:
Host Computer Operating System:
Estimated Time To Complete this activity:

Use Case: University of Virginia

Local Context

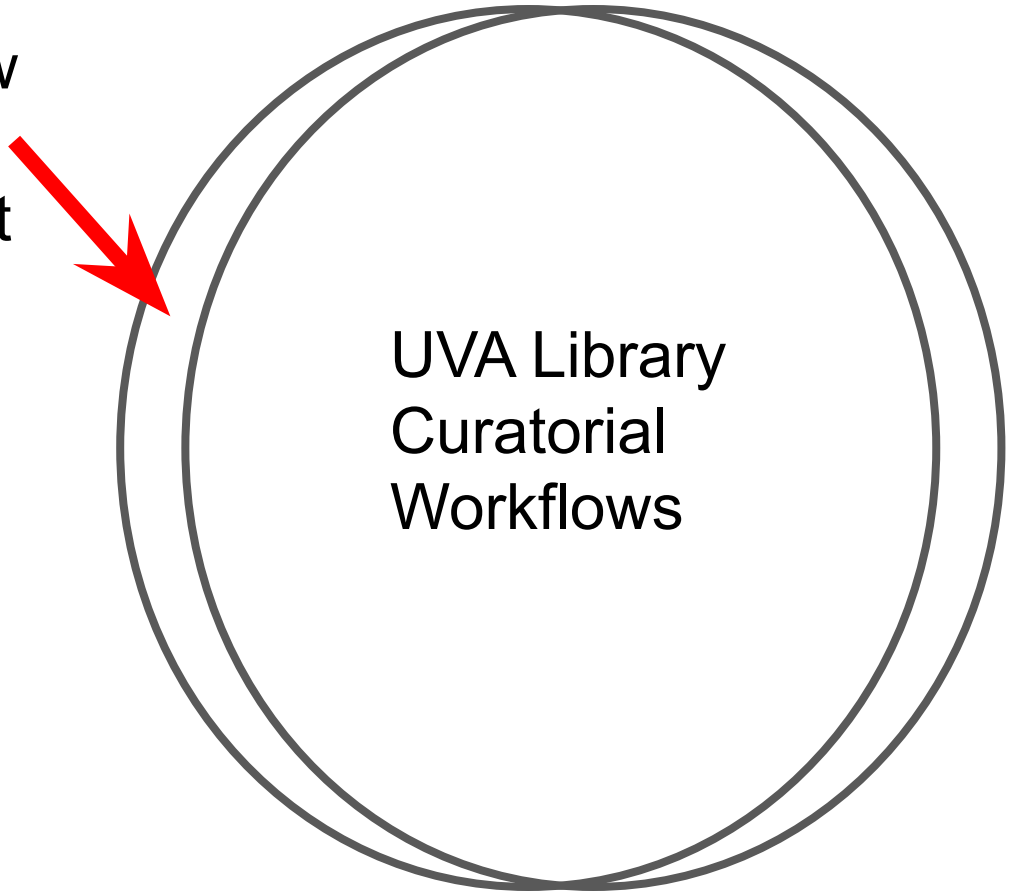
FCOP Grant Goals

- Curatorial workflows and documentation
- Documentation and development of technical approach
- Develop and share archival description approaches for software
- **Bonus!** Access & user tests and documentation

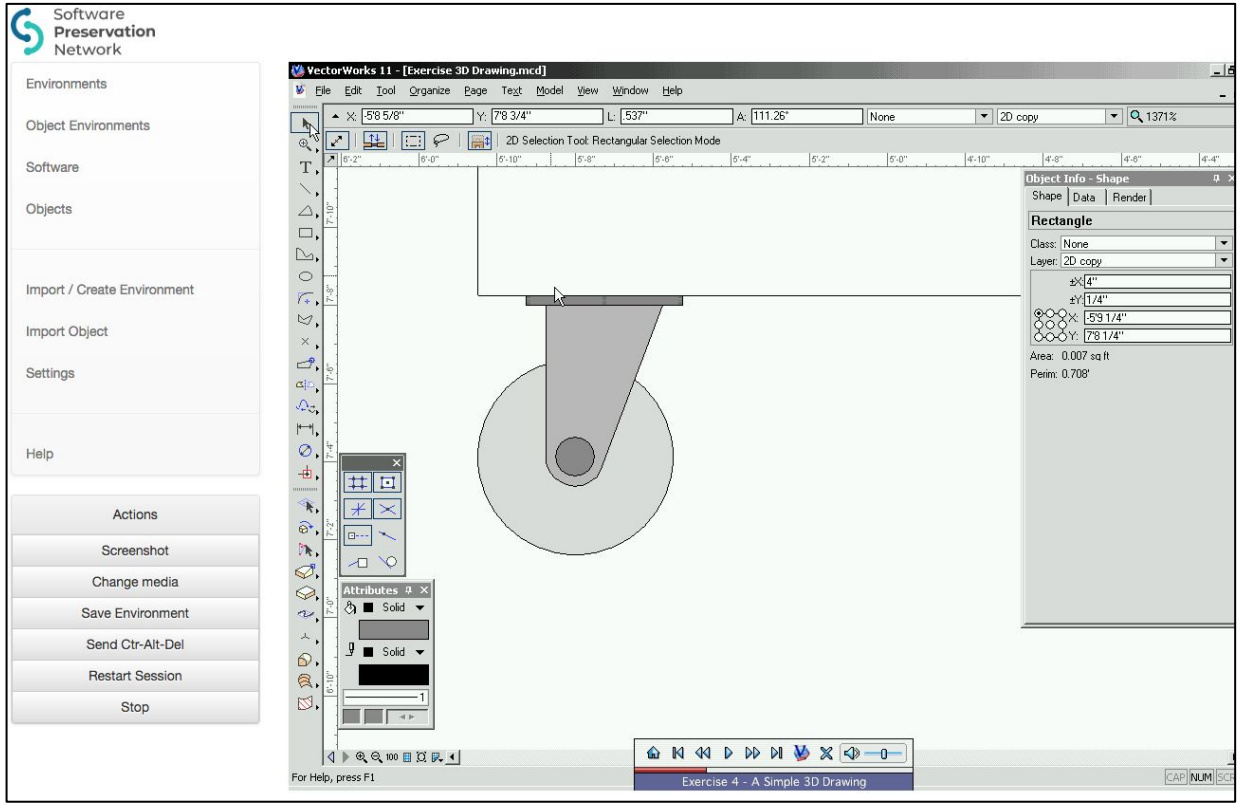
UVA Context

Not pictured: Elizabeth Wilkinson

Goal for unique/new software and software-dependant digital materials workflow



UVA Project: Emulation in the Archives



EMULATION IN THE ARCHIVES

bit.ly/EIA2019Notes

UNIVERSITY of VIRGINIA LIBRARY

- hardDrive
  - _6LQI000~004~
  - _AE7J000~004~
  - _P2R2000~004~
  - _RE25000~004~
  - .DS_Store
  - 5th & Water St
  - Accessability Symbol
  - AFG Letter text
  - AIMR
  - Airport Specs Template
  - Albemarle Winery
  - AlbemarleCountyZoningOrdinance
  - Applegate
  - ARCH STAMP
  - autorun.inf
  - Bampict
  - Barr
  - Battig
  - Blanton
  - Bluestone Pattern v8
  - Brooks
  - Bryan Freebee
  - Cabin Handrail v9
  - Campbell-Paris
  - CBD Downloads
  - Chandler Franklin Obryan
  - Cherry
  - CHO ACT MasterPlan.For Home

Preservation Network

Environments

Software

Objects

Import Environment

Create Environment

Settings

Emulators

Help

Actions

Screenshot

Change media

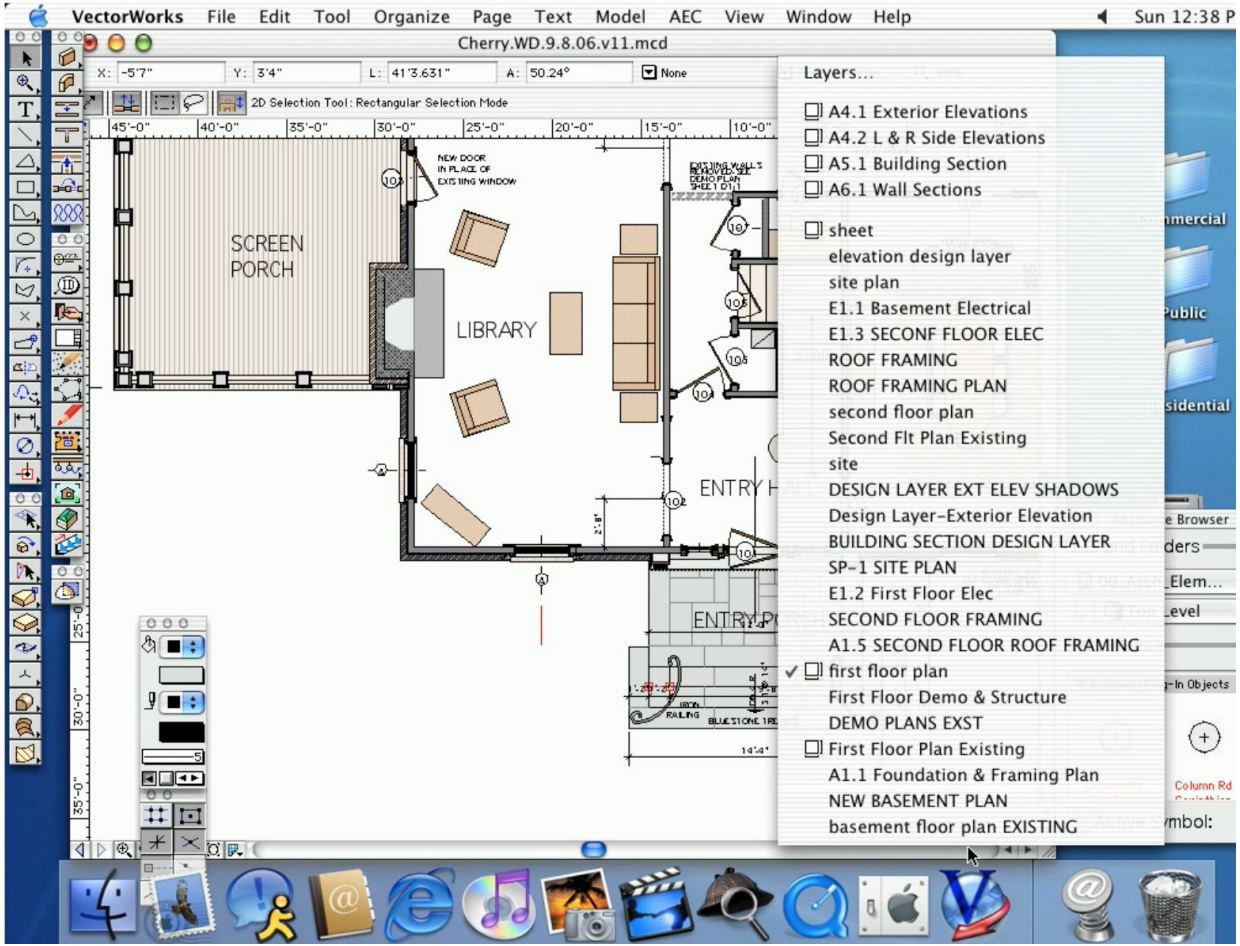
Save Environment

Send Ctr-Alt-Del

Send Esc

Restart Session

Stop



Goal: Curatorial Workflows

Digital Donor Checklist

Digital Donor Checklist (V.03)

Created by Work, Lauren (lw2cd), last modified on Jul 12, 2019

Prior to acquisition or preliminary assessment prior to donation

Questions:

- Do we want to build this out for holistic collection review? (e.g. analog and digital materials)
- Yale form structure (checks in with processing workflow, aspects of work that need to be done, with some UCLA sprinkled in)

Contextual - external/donor Qs

- [ ] Statement about context and content of collection (what is the history behind the collection? where did it come from, how was it created, etc.)
  - [ ] Analog materials as well as digital transfer? Additional contextual information to provide? (might take away this question if we decide this will be a "digital" checklist only)
- [ ] Approximate years of use of the machine? (Date of purchase could be helpful, but even years of use, and how it was used (for personal and for work purposes? Work only?)
- [ ] Knowledge of sensitive personal info (self or otherwise)
- [ ] Encrypted/protected files?
- [ ] Are there other materials that may exist on the transfer hardware that shoul
- [ ] Do you know of any material that may be under copyright on the machine?
- [ ] Have you recently moved or rearranged files in preparation for donation?

bit.ly/EIA2019Notes

Software Questionnaire

Software Questionnaire (V.03)

Created by Work, Lauren (lw2cd), last modified just a moment ago

Meant for collections that have been flagged as potentially having software and/or some types of software dependent material in need of early collection & preservation attention via the initial donor collection inventory & checklist

Goals: Know enough about the context/content of work using software & software itself as to make acquisition, appraisal, arrangement, description, preservation, and access judgements without being overly onerous to donors/SC&A staff

Note: This particular questionnaire is in relation to collections that are archival in nature/collection focus that are part of our FCOP grant. Depending on the collection (particularly for University Archives) this particular questionnaire may be related to/support - but does not necessarily replace - questions and structure around scientific data and software for research.

Commercial software (produced for purchase by an organization, not the donor)

Purpose and use *(may aid SC&A/P with determination of appraisal, users and research interests)*

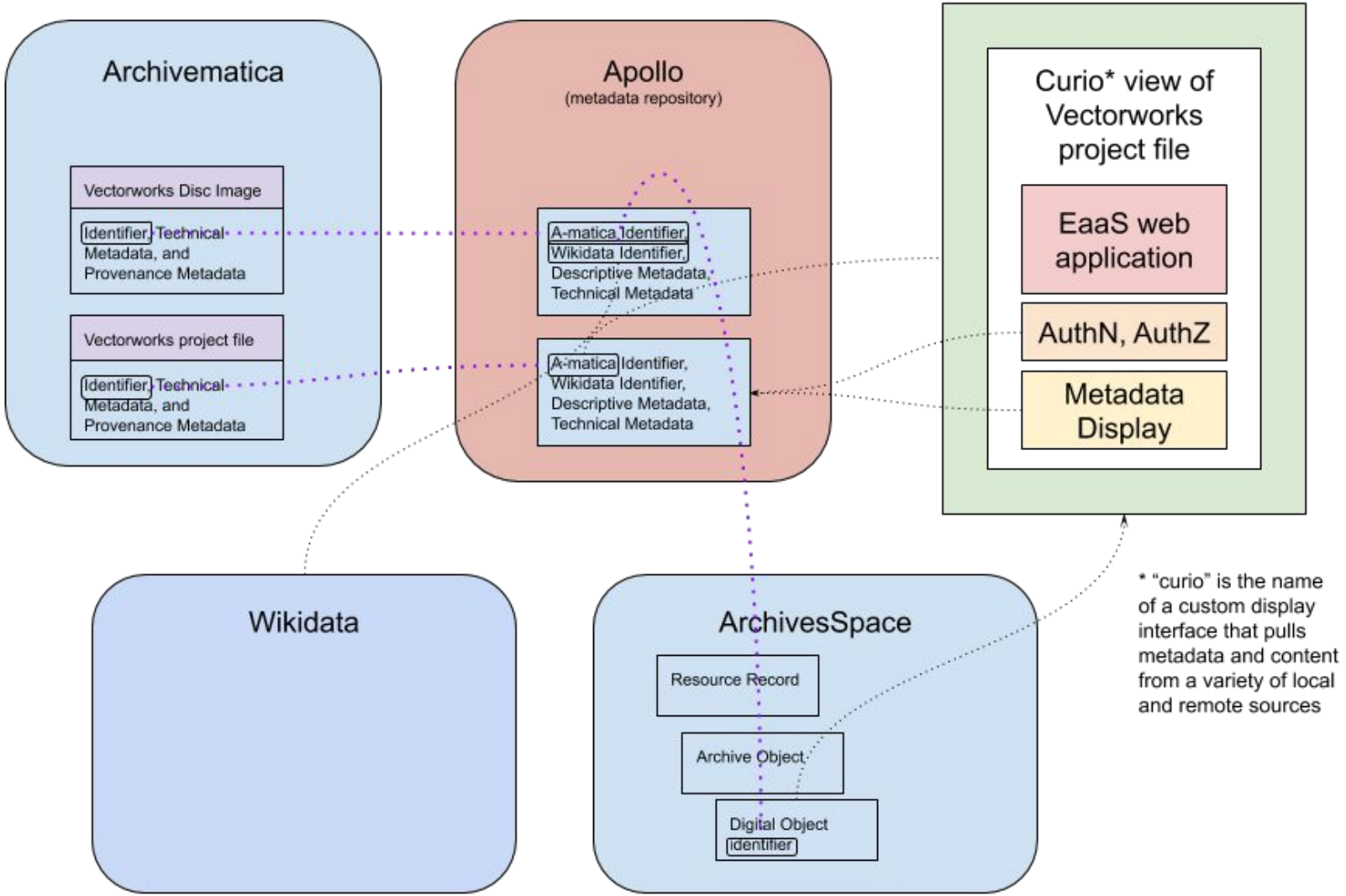
- [ ] What is this software primarily used for?
- [ ] How did you use this software in your work?
- [ ] What operating system did you use with this software?
- [ ] Is there other software that can also render the files you may have produced with this software?

bit.ly/EIA2019Notes

Carriers/Transfer of software (*will aid preservation and access steps*)

- [ ] Do you still have the installation disks for the software?

Goal: Technical Infrastructure Workflows


Archivematica
Vectorworks Disc Image
Identifier, Technical Metadata, and Provenance Metadata
Vectorworks project file
Identifier, Technical Metadata, and Provenance Metadata
Apollo
(metadata repository)
A-matica Identifier, Wikidata Identifier, Descriptive Metadata, Technical Metadata
A-matica Identifier, Wikidata Identifier, Descriptive Metadata, Technical Metadata
Curio* view of Vectorworks project file
EaaS web application
AuthN, AuthZ
Metadata Display
Wikidata
ArchivesSpace
Resource Record
Archive Object
Digital Object Identifier
* "curio" is the name of a custom display interface that pulls metadata and content from a variety of local and remote sources

Goal: Descriptive Workflows

Archival Description Strategies for Emulated Software

Case Study

UVa. SPN-FCoP Emulation in the Archives Team
Updated: July 2019

Introduction

The Sheeran Architects records (1980-2008) include drawings and office files relating to projects by Peter L. Sheeran and members of his firm, financial records, correspondence, notes and sketches. The collection includes both digital and analog materials.

bit.ly/EIA2019Notes

(Bonus!) Access & User Testing

Bringing it together: Current Shared Practice & Future Opportunities

	Similarities	Differences
Institutional Context	Large public research institutions Org. structure 1 preservationist	Number of special collections and curators, different silo styles
Collections	Large special and research collections, ability to support grants	Size and focus of collections, budget lines/sources, collection policies
Digital Collections Infrastructure	Fairly mature structures	In house vs mix of open source

Workflow	UI more mature in their workflows. Scale of big importance. Documentation very important	Ownership of workflow/systems, development time, UVA still building
Metadata/ Description	Strong focus on standards & interoperability, desire to use Wikidata	UVA focus on DACS and archival description, UI focus on flexible approach
Access	Able to employ emulation sandbox, discovery via finding aid, reading room only	Staffing, training, responsibilities, researchers

Shared practice

Archival principles are still the bedrock of this work

Collaboration & cooperation are key both institutionally but also in community (including FCOP & EaaS)

Software preservation, emulation and access will take longer and take more troubleshooting at first: be kind to yourself, ask questions, share documentation, and take small steps

Opportunities and Shifts in Practice

- Curator/Donor interaction; early, early, early
- Legal environment
- Description
- User and Access studies
- Iterate: Share your documentation, work, and testing before everything is perfect.
- **The big picture: The preservation of software and software-dependant digital materials are more than technological problems**

Select Resources:

- **About EaaSI:** https://www.softwarepreservationnetwork.org/eaasi/ (accessed 8/04/19)
- **About FCoP:** https://www.softwarepreservationnetwork.org/fcop/ (accessed 8/04/19)
- **Code of Best Practices for Fair Use in Software Preservation:** https://www.softwarepreservationnetwork.org/bp-fair-use/ (accessed 8/04/19)
- **EaaSI Handbook:** https://eaasi.gitlab.io/eaasi_user_handbook/ (accessed 8/04/19)
- **EaaSI Sandbox**: https://www.softwarepreservationnetwork.org/eaasi-sandbox/ (accessed 8/04/19)
- **On the Rushdie project** from Personal Digital Archiving 2017: “one of the biggest lessons learned noted by Waugh was the need to document everything the software engineers do as their work is just as ephemeral as the born digital information they wished to preserve”: https://ws-dl.blogspot.com/2017/04/2017-04-17-personal-digital-archiving.html (accessed 8/04/19)
- **Software Preservation Network:** https://www.softwarepreservationnetwork.org/ (accessed 8/04/19)
- **SPN EaaSI SAA 2019 Multithreading Software Preservation Workshop notes**: bit.ly/swpres-emulation (accessed 8/04/19)

Questions and Comments?

lw2cd@virginia.edu
tpopp2@illinois.edu

- **SPN EaaSI SAA 2019 Multithreading Software Preservation Workshop notes**: bit.ly/swpres-emulation
- **UVA EIA 2019 Workshop notes:** bit.ly/EIA2019Notes

Holding slide for video/demo. Also could encourage folks to peruse the public EaaS sandbox