



**Carnegie Mellon University**

# Olive: An Archive For Executable Content

A new Computer Science/Library Partnership

*Presenters*

**Gloriana St. Clair**

Dean of University Libraries

**Daniel Ryan**

Curator of Executable Content

# Overview of Olive

Olive seeks to close a gap in the ability of academic libraries to meet the preservation requirements of constituents producing executable content.

- There is a high demand for access to and execution of early computer applications and games
- Preservation of data or code alone is insufficient – users must be able to actually manipulate this content in its original environment
- Olive will provide for precise replication of research operations

# Olive since 2011

**We began thinking about these challenges in 2011.**

At that time, Olive was not actively being developed. We had:

- No funding
- Repurposed software
- Vague details
- Hand waving
- No web presence

We're here to tell you about how all of this is changing.

# IMLS Leadership Grant

We received an IMLS National Leadership Grant beginning in October 2012.

- Award: \$497k over 2 years
- Goal: To understand what types of content can be ingested into Olive and to determine how an executable content archive can fit into existing trusted repository standards such as OAIS, OCLC, CRL and JISC

## **Olive Team**

- Gloriana St. Clair, Carnegie Mellon University
- Jerome McDonough, University of Illinois at Urbana-Champaign
- Anita de Waard, Reed Elsevier Inc.
- Mahadev Satyanarayanan, Carnegie Mellon University
- Daniel Ryan, Carnegie Mellon University



# Sloan Foundation Grant

We received a grant from the Alfred P. Sloan foundation beginning in January.

- Award: \$400k over 2 years
- Goal: To develop the technical framework for Olive which supports the long term preservation of executable content; To plan for an effective organizational structure to sustain the archive and provide access to executable content

## **Grant Team**

- Mahadev Satyanarayanan, Carnegie Group Professor of Computer Science
- Gloriana St. Clair, Dean of University Libraries
- Daniel Ryan, Curator of Executable Content
- Jan Harkes, Principal Project Scientist
- Benjamin Gilbert, Project Scientist

# So what are we doing?

## **Technical Work**

- Establish a curation/publication workflow
- Provide cloud execution support for non-linux users
- Improve native client VM execution
- Separate curation/publication from access
- Looking ahead

## **Organizational Work**

- Investigate an organizational model for sustainability
- Organize a summit of interested parties to discuss options

## **Initial Content**

- Ingest a variety of content to aid in development

# Technical Overview

## Phases

### Alpha

- Single curator, private cloud, native (linux) clients only, invitation only
- Private cloud
- Native linux client support only
- Access content by invitation only

### Beta

- Introduce more curators to test workflow
- Invite more users to test system
- Begin support for cloud execution

### Public Release

- Implement established access restrictions, licensing requirements, etc.

# Initial Publication Workflow

## **VM Construction & Testing**

Done on native client only, using VMNetX software built by Satya.

## **VM Upload to Olive**

Development team will implement a web interface for VM uploads.

## **Publication**

- Curator must agree to legal notice/licensing requirements
- Automated VM testing before publication
- Built-in support for xml metadata inclusion
- Final submission check before publication

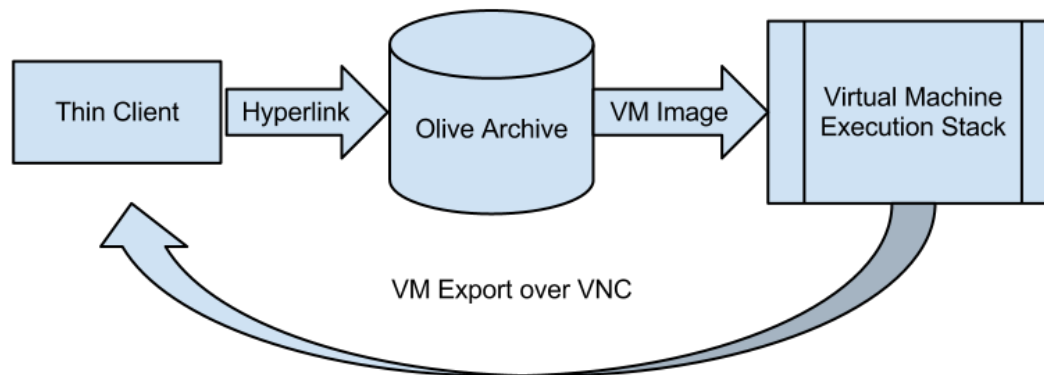


# Providing Cloud Based Execution

Goal: To provide a Youtube-like interaction experience for end-users.

## Technical Details

- Clicking a link will cause a VM to execute in the cloud
- GUI to the VM will be exported to the client over VNC
- No software installation is required for client
- Supports public (e.g. Amazon EC2) and private clouds
- Execution from mobile or thin clients is possible



# Improving Native Client Support

Goal: To improve execution fidelity of virtual machines on native clients.

## Task

- Integrate support for native client caching
- Leverage a single virtual machine manager (KVM) to make maintenance manageable
- Provide enhanced native hardware-accelerated VM execution

## Function

- Expanded support for a wider range of content
- Exposes an interface to local hardware, like local GPU's or integrated cameras
- Provides advanced VM state prefetching for low bandwidth connections

# Separating Curation from Access

Goal: To ensure that curation tasks can be maintained separately from access and execution; To ensure modularity of operations.

## Task

- Ensure independent implementations of VM access/launch and upload/curation
- Integrate rights clearance, certification, etc.
- Optimize VM access and launch operations
- Track all operations and user agreements

## Function

- Allows for independent development and implementation of operations
- Ensures independent workflows and operations
- Generates detailed records for auditing and legal requirements

# Finding an Organizational Model

Goal: To determine the best organizational structure to maintain and support Olive in the long term.

## **Task**

- Put out an RFI for a whitepaper by an economist or policy expert
- Choose among the respondents and determine deliverables

## **Function**

- Offers case-specific insight
- Provides foundation for establishing necessary supporting organizations and partnerships

# Possible Base Models



## Red Hat Model

- Open source software, available to all users
- Subscription based support and training for enterprise clients



## Apache Model

- Decentralized contributors, elected board of representatives
- Corporate and private sponsorship to support maintenance/stability



## JSTOR Model

- Usage rights licensed by clients
- Negotiate content inclusion with suppliers

## MetaArchive Model

- Resources are distributed among participating institutions
- Maintenance costs recovered in annual membership fees



# Convening a Summit

Goal: To bring a coalition of interested parties in the public, private, and academic fields together to come to a consensus about the future of Olive.

## **Task**

- Identify key representatives in public, private, and academic positions
- Supply economic whitepaper to all representatives for review
- Organize and convene summit

## **Function**

- Ensures that key stakeholders will have input into the structure of Olive
- Offers a forum for detailed discussion about options in light of the whitepaper
- Builds consensus and cooperation for future research in this area

# Test Cases

We arranged to ingest several different types of content as use cases:

## **Great American History Machine**

- Built by Prof. David Miller at CMU in the 1980's, provides GIS for early census data

## **ChemCollective**

- Award winning educational software built by Prof. David Yaron

## **Mystery House**

- Apple II game for which experimental metadata was generated under IMLS-funded *Preserving Virtual Worlds 2*

## **Doom**

- Popular PC game with heavy graphics released in 1993

## **Information Systems**

- Special issue of the Elsevier journal which will include executable content

# Challenges & Looking Ahead

Olive faces other significant challenges which are outside of the scope of current funding levels. Some things we see as necessary to the development of Olive include:

- Extensible support for a wide range of licensing options
- Support for rich, evolving metadata to accompany content parcels
- Development of content-based search across the VM archive, using VM state introspection
- Handling copyright/IP for content



# Descant & Acknowledgements

“Knowing the truth about the vanishing did not lessen their opinion of Bilbo at all; for they saw that he had some wits, as well as luck and a magic ring—and all three are very useful possessions.”

*The Hobbit, p. 163*

This challenging work is only possible thanks to the wit, luck, and magic supplied by our colleagues and partners. Special thanks to:

- Benjamin Gilbert
- Jan Harkes
- Dave Yaron
- David Miller
- Vas Bala and the IBM research team
- Carnegie Mellon University, IMLS, and the Sloan Foundation
- Many more contributors

# Learn More

Find us online:

<http://olivearchive.org>

Partners:

Institute for Museum and Library Services

<http://www.ims.gov>

Alfred P. Sloan Foundation

<http://www.sloan.org>

Reed Elsevier Inc.

<http://www.reedelsevier.com>

# Questions?