Carnegie Mellon University

Olive: A Digital Archive for Executable Content

Gloriana St. Clair, Dan Ryan, Mahadev Satanarayanan(Satya), Vasanth Bala, Erika Linke

Coalition for Networked Information, December 12-13, 2011

The *Olive* Project

Thesis:

"Today, an increasing fraction of the world's intellectual output is in the form of executable content. These include simulation models, tutoring systems, expert systems, data visualization tools, and so on. Even content that could be static (such as a company's product information Web page) is often dynamically generated through execution of code that customizes the content and appearance at runtime... [Olive seeks] to freeze and precisely reproduce the execution state that produced this dynamic content."

Satyanarayanan, M. et. al., Collaborating With Executable Content Across Space and Time, Proceedings of the 7th International Conference on Collaborative Computing: Networking, Applications and Worksharing, October 2011

History

2008

- Satya and Vas conceive of Olive, IBM reviewer suggests potential library application

2009

- Satya contemplates long term preservation and opens communication with Gloriana St. Clair to discuss past efforts and current preservation responsibilities of the library

2010

- Dave McQueeney at IBM hosts invitational workshop arranged by Satya, Vas, Gloriana, Erika and Joel Smith (CIO at Carnegie Mellon)

IBM Workshop at Watson Labs (2011)

Attendees:

Gloriana St. Clair (CMU), Mahadev Satyanarayanan (CMU), Erika Linke (CMU), Vasanth Bala (IBM), Dianne Fodell (IBM), Giovanni Pacifici (IBM), Clifford Lynch (CNI), Chris Shillum (Elsevier), Mike Handy (Library of Congress), Claire Cocco (OCLC), Mike Lesk (ACM), Patrick Dreher (RenCI), Timothy Dilauro (Johns Hopkins), Kevin Guthrie (Ithaka), Paul Jones (ibiblio), Vijay Kumar (MIT), Naomi Nelson (Duke), Peng Ning (NCSU), Mladen Vouk (NCSU), Joan Smith (Emory), Michael Witt (Purdue)

Feedback

- Group excited to see demonstration of old software actually running in Olive
- Softwares should run as described on original packaging
- Many Intellectual Property issues must be considered moving forward
- Extensibility concerns with Many-to-one relationships from executables to parcels
- Metadata tracking and formatting questions
- Sustainability as a major challenge

Conclusions

Light text body copy

- Satya and Vas to pursue bootstrapping funding
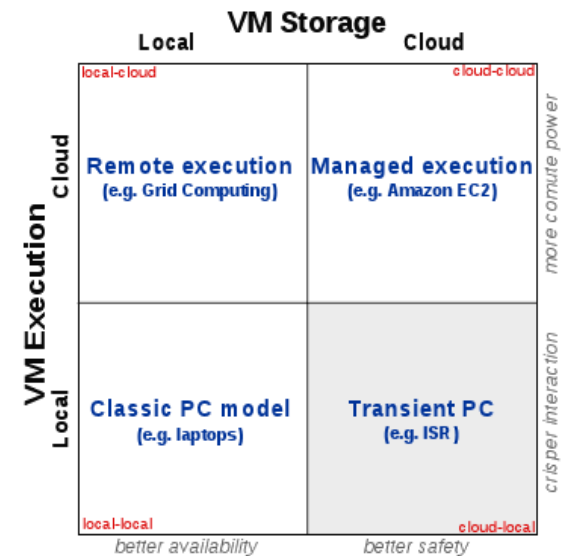- Gloriana and Erika to pursue funding for the mounting use cases

???

The *Olive* Engine

Internet Suspend/Resume (ISR)

Software developed by Satya and Vas which enables users to stream virtual machines instances remotely

- Facilitates a new model of execution

- End users can modify individual VM instances without impact on others
- VM is divorced from any specific software package for portable preservation

Why ISR?

Software

- Not restricted to one platform or package
- Users can interact with, manipulate, and change the VM without disturbing the integrity of the original source
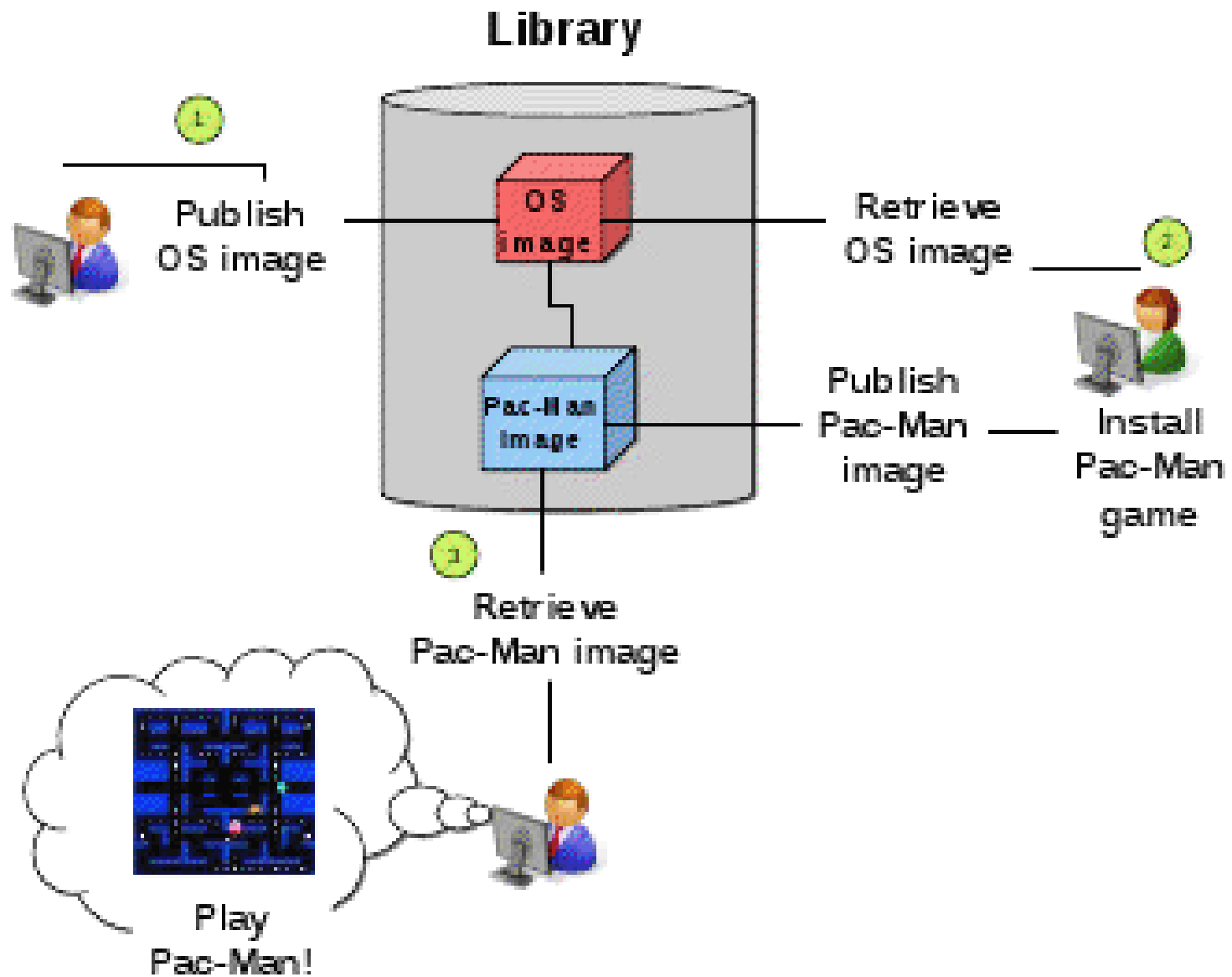- Eliminates the need for porting old code for preservation

Hardware

- Allows for execution state preservation without the preservation of hardware
- Maintains the historical integrity of the original hardware more and more accurately as VM technology improves

Users

- Users are not restricted from manipulating and executing the data

???

Library

Publish OS image · OS image · Retrieve OS image

Publish Pac-Man image · Install Pac-Man game

Retrieve Pac-Man image

Play Pac-Man!

Why the Library?

The Role of the Library

- Preserve information of all subjects and media
- Organize and mediate access to relevant information
- Curate collections using expert knowledge about both archiving and the content of the collection itself

The logical step for libraries is to apply the traditional librarian skillset to the digital era by acquiring the skills necessary to preserve, archive, and curate digital collections of executable content

Looking forward

Challenges

- Ease of use & difficulty of operation
- Implementation and testing
- Intellectual Property Questions
- Scalability
- Sustainability
- Unknown Unknowns

???

Evolving IMLS Grant Proposal

Overview

Partnership grant with IBM, Elsevier, Dave Yaron, and the authors of the *Preserving Virtual Worlds* whitepaper on best practices in executable content preservation seeking to implement several use cases in executable content preservation

Content

- *ChemLab* educational chemistry lab software developed by Dave Yaron (CMU)

- Executable content from a special issue of a journal published by Elsevier

- Interactive game with selection and implementation advice from the folks involved in *PVW*

The New Librarian

Adapting to the Digital Age

Librarians have a responsibility to preserve and curate information from all fields, but are falling short in the area of digital preservation. In order to fulfill our obligation to the scholarly community and the general public, we must build strategies and practices for digital preservation

Meeting the Demand

- Selection & Curation
- Preservation
- Dissemination

Acknowledgements

- Benjamin Gilbert (ISR)
- Others?
- More others?

Questions?