Emulation as a Service

*Scalable Curation and Access to Complex Objects*

Isgandar Valizada, Klaus Rechert

(Albert-Ludwigs-University of Freiburg)
bwFLA – Emulation-as-a-Service

- Emulation as an alternative preservation strategy
  - in discussion for years (cf. Rothenberg, 1995)
  - however
    - emulation has often been seen as too complex for non-experts (i.e. to be used in memory institutions)
    - did not scale well, not cost-effective
    - preservation planning unclear
Observation: there is only a limited number of computer platforms

→ Follow the idea of division of labor
  → centralize technical management and maintenance
  → share expertise and costs
  → create a scalable and modular service
→ Emulation as a Service (EaaS)
EaaS - Emulation Components

- **Unified access** to emulation:
  - Encapsulation of different emulators and technology to common component
  - Attachment of user-media
    - dynamically (e.g. Floppy, CD-ROM)
    - permanent (e.g. HDD)
  - Interactive access to emulated environments (e.g. HTML5 viewer)
  - Technical interaction with the environment (IP, specialized protocols)
  - Main building block for complex environments
    - Client/Server etc.
  - API exposed as Web Service (WS)
    - Interoperability to other systems
bwFLA‘s Emulation-as-a-Service

- On-demand Resources
  - EaaS components require almost no statically allocated resources
  - Allocation of computing resources „on-demand“
  - On demand cloud / cluster deployment
    - Amazon EC2 (e.g. € ~0.05 / h per user session (single node))
    - OpenStack
Example WF: Rendering a Digital Object

- In order to render digital objects (data), an object has to be associated to a rendering environment

- Assumptions & Preconditions
  - the dig. object (data) is already in an repository
    - i.e. automatically accessible / retrievable
  - the user has knowledge on desired rendering environment
  - the user has expectations on dig. objects utility
Preparation: Step 1 - Example

- Basic (but incomplete) meta-data is available from the Transmediale Archive
- Only a limited choice of prepared emulated environments is provided
  - Expert-mode with ability to modify / change / build environments is under development
Preparation: Step 2 - Example

- Object chosen in step 1 is available as CD-ROM (ISO)
- In this example the object is a set of web pages and flash applications
- Object has been rendered in the MS Internet Explorer
Preparation: Step 2 - Example

- The user is able to describe the environment's technical properties
- The user is able to take a screenshot which is displayed for access
- The user is able to describe specific object performance issues
Preparation: Step 3 - Example

1. Basic Artifact Properties

<table>
<thead>
<tr>
<th>Information Displayed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title: Map 50</td>
</tr>
<tr>
<td>Author:</td>
</tr>
<tr>
<td>Year: 2001</td>
</tr>
</tbody>
</table>

2. Test and Evaluation

3. Complete Ingest

Update Album Art

[Image of Map 50 album art]

Upload Image
Handle It!

- Result of the workflows is a set of meta-data describing the technical environment + data.
- Create a **Handle** (HDL) to cite your environment + data
  - HTML5 rendering in the browser, interactively usable
  - [http://hdl.handle.net/11270/71e34632-155a-4a05-b2e3-c57ba99067b2](http://hdl.handle.net/11270/71e34632-155a-4a05-b2e3-c57ba99067b2)
All the data belongs to you!

- **User control**
  - Users‘ want to or are required to stay in control of their digital assets
  → Metadata, dig. objects and customized environments are to be stored at the owner‘s site

→ **Requirements**
  → Decentralized data management
  → Service orchestration
  → Efficient data transport
  → Efficient data management
Technical Architecture

![Diagram of Technical Architecture with labels such as Image Archive, User Environment, Meta Data, and connections like connect read-only and connect read-write.]

- Image Archive
  - User-modifications
  - Base Image
  - Meta Data
  - Reference to Base Image HDL/DOI

- User Environment
  - User-modifications
  - Meta Data
  - Connect read-only

- Emulation Component
  - IP
  - ODBC
  - Data

- WS-API
  - VNC
  - RDP
  - HTML5
Derivative Environments

- EaaS provides base images
- Generate „delta-files“ of changes (blockwise)
  - simple blockbased metadata → can be consolidated manually
- Unaltered blocks are linked to the base image

<table>
<thead>
<tr>
<th>Offset</th>
<th>Length</th>
<th>Mapped to File</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0x10000</td>
<td>0x270000 derived.qcow2</td>
</tr>
<tr>
<td>0x10000</td>
<td>0x10000</td>
<td>0x60000 base.qcow2</td>
</tr>
<tr>
<td>0x10000000</td>
<td>0x10000</td>
<td>0xab0000 base.qcow2</td>
</tr>
<tr>
<td>0x20000000</td>
<td>0x210000</td>
<td>0x50000 derived.qcow2</td>
</tr>
<tr>
<td>0x20210000</td>
<td>0x800000</td>
<td>0x2b0000 base.qcow2</td>
</tr>
<tr>
<td>0x30000000</td>
<td>0x10000</td>
<td>0xac0000 base.qcow2</td>
</tr>
<tr>
<td>0x3ffe0000</td>
<td>0x20000</td>
<td>0x80000 base.qcow2</td>
</tr>
</tbody>
</table>
Orchestration

Abstract description for on-demand re-enactment

<emuEnvironment xmlns="EmuEnvironment">
  <uuid>2016</uuid>
  <title>IBM OS/2 2.11</title>
  <arch>i386</arch>
  [...]  

  <drive>
    ...
  </drive>

  <binding id="...">
    ...
  </binding>

</emuEnvironment>
Orchestration: Bindings

- Bindings - description and management of data sources
  - PI to data source (e.g. HDL)
  - Access method
    - copy-on-write if supported by transport protocol
    - copy (fall-back)
  - Access credentials (optional)
Orchestration: Drives

- Drives: mapping data to media
  - reference to a binding definition
    - fs-like relative file / container selection
  - definition of a media type (e.g. floppy / HD / CD-ROM)
  - definition how the medium is connected to the emulated system

```xml
<drive>
  <url>binding://floppy_data_binding/disk1.img</url>
  <iface>fda</iface>
  <bus>0</bus>
  <unit>0</unit>
  <type>floppy</type>
  <boot>false</boot>
  <plugged>true</plugged>
</drive>
```
Network Transport

- Any networked transport protocol should work
  - HTTP(S), FTP, NFS, CIFS, etc.
  - Connector implementation for individual repositories are required (e.g. export / import file)

  → full copy required
    → requires bandwidth, reduces user experience

- Block oriented protocol („emulates“ RA disks)
  - Network Block Device (NBD), simple and lean protocol
  - Bandwidth efficient → „instant“ emulator start
EaaS by Example

- Preservation of and access to inherited personal digital assets
  - citation support (HDL)
  - simple browser-based access
  - embeddable and shareable like Youtube videos

The Digital Heritage of Víštem Flusser

The Víštem Flusser Archive owns a personal computer the philosopher used himself. It stores mostly digital artifacts that represent Flusser’s textual work, but also a software called “Hypertext” which is dependent on the obsolete authoring system HyperCard.

Apple Macintosh Performa 630 computer
Vilem Flusser Archive, Berlin

http://hdl.handle.net/11270/2b87de90-37dc-4d66-a9e6-546a80b0b261
EaaS by Example

- Preservation of and access to (complex) e-books
  - Content protection: presentation only, object data is not accessible to viewers (also important for dig. art)
  - Access restriction: Shibboleth support

Vilem Flusser, *Die Schrift. Hat Schreiben Zukunft, 1987*  
Vilem Flusser Archive, Berlin  
http://hdl.handle.net/11270/767f2c0b-cce6-4623-8caf-f5a890afcb75
EaaS by Example

- Preparation of custom environments
  - Centralized and effective preservation planning
  - All user images are derived from base images
  - Single migration strategy for base images required, to run on next-gen emulator, then applied to all derived images
EaaS by Example

- Tailor environment to a specific object
  - e.g. install a „viewer“
  - e.g. autostart a specific object
- Support persistent user sessions
EaaS by Example

- Connectivity
  - Authentic client access
    - e.g. to archived web sites (esp. MIDI sound)
  - Complex Server Setups
    - e.g. cost-effective preservation of a CMS-based Web-site (server)
    - re-enact server on-demand (click HDL)
    - access with old / current browser
    - demo at iPres14

Geocities Archive
One Terabyte of the Kilobyte Age

http://hdl.handle.net/11270/21332f4a-c224-4877-abe4-ad44084716f0
Issues & Limitations

- Intellectual Property Rights
- External Dependencies
  - External (Big) Data Sources
  - External Functional Dependencies
- DRM, Dongles
- Hardware Dependencies
  - e.g. sensors attached
Summary and Outlook

- bwFLA provides user-centric workflows and tools to be applied on DO today
  - Incorporate user knowledge
  - Incorporate user feedback
- The proposed workflow requires significant manual user interaction and seems costly and time consuming at first sight
  - The basic rendering environment is quite stable concerning software and hardware dependencies
  - The main differences are mostly on the top layer of the VP description
    - only a few additional steps are required if the software archive already contains suitable VP descriptions of today's common digital objects
  - Workflow could be further accelerated by caching strategies by automation of installation tasks
(Semi-)Public Demo

https://demo.bw-fla.uni-freiburg.de

Username: bwfla
Password: demo