Emulation: Bridging the Past to the Future without Altering the Object

*Dirk von Suchodoletz, Randolph Welte
University of Freiburg – Department of Computer Science
Preservation Challenges

• Digital objects require software / hardware environments to be accessed
• Environments change over the time and obsolete most of digital material
• Mainline strategy: Migration
  – Risky to rely on it exclusively
  – Not suitable for all object types
Dynamic Digital Objects

- Dynamic digital objects
  - Applications
  - Operating systems
  - Databases
- Non-linear, user interaction, multiple views
- No real option:
  - Printing of source, adaption to recent environments; even if source code available
  - Video-recording, screenshots of game or application session
Different Approach

• **Emulation** – no changes on the object, but recreation of original environment
  – Emulators around for quite a while, supplemented by virtualization
  – Can operate on different layers of software / hardware stack
  – Number of objects to cover differs significantly; thus hardware layer seems very attractive to focus on
• Help to bridge widening gap of the computers past to the future
Emulation Layers

- Different options (text doc, game examples)
  - MS Word document in OpenOffice
  - Running MS Word 97 in Wine on Linux X86
  - Emulation of X86 machine to run complete Windows 98 environment with MS Word installed

- Depends on the type of object
  - Proprietary formats may prevent proper interpretation
  - OS, application APIs often obfuscated (MS)
Emulator Examples

• Dioscuri X86 emulator recreating an 286, 386 PC of the early 1990th
  – Java programming language, modular approach – components like disk, floppy, VGA, CPU, RAM put together to form the machine
  – Running DOS and Windows 3.0
  – Step by step extension to 486++

• QEMU – using popular C programming language multi architecture emulator for X86, PPC, Sparc, ...
  – Large user community
  – Actively developed

• Both Open Source – no vendor dependencies, adaptable
• Independently of migration or emulation – digital object is to be handled somehow
  – (Re)creation of a certain hardware software environment for access / execution
• Standard workplace environments for migrated objects, but ...

Migration takes place 'within' the archive and is applied to all objects of a certain type
Emulation is used for creation of environments of objects outside of archive (view path)
Requirements

- Emulation not working just on its own – additional software is required
- Emulation approach requires recreation of ancient hardware / software environments for access / execution
  - E.g. spreadsheet document requires the proper spreadsheet application for interpretation and displaying
  - Spreadsheet software is dependent on an operation system
  - Operating system was programmed for a very specific or a range of hardware architectures
  - Additional components like fonts might be needed for range of documents, especially for non-latin typesets
- Object transport into viewing / execution environment to be taken care of
View Path

- View Path – pathway from object to specific environment
- Reference environment – specifically defined software hardware combination for object access, rendering
- Formalization needed – view path as the requirements to be followed to actually access, display the object of interest
- Introduced with Preservation Layer Model (PLM) of IBM/DIAS project
- Concept extended in the ongoing project
  - More flexible layout
  - Introduction of metrics for multiple view path
  - Match to users, organizational needs – significant properties
View Path Characteristics

- Variable length; depending on
  - Type of object and platform (image, document, application ...)
  - Emulator preservation strategy – getting longer with emulation stacking

![Diagram showing path characteristics with icons for documents, pictures, video, audio, executable applications, appropriate operating system, emulation of the required architecture, stacked emulator, and emulator for the hardware req. by the OS.]
• Often more than one view path exists
  – Depending on object more than one renderer available
  – Rendering / execution results may differ (significantly)
  – Less and more simple, expensive view path
• Introduce metrics for decision making in preservation planning processes
• Offer users options to choose depending on their research interest, preferences
Significant Properties

- View paths not fixed for other dimensions too
- Significant properties highly debated term in digital preservation
  - Determine options how to preserve objects
  - Evaluate and compare preservation strategies and outcomes
- Metrics could be related to significant properties
  - Definitely depended on the designated user communities
  - Archivists, librarians, computer museum curators or retro gamers may not share same vision of significant properties of objects
  - E.g. ask for a definition of the term “authenticity” to get a wide range of good answers
- Use metrics to include users experience, feedback to improve results for similar objects (comparable to recommender systems)
Each step in view path might add software, tools requirements
- Applications
- Operating systems
- Helper tools like packer
- Hardware drivers for VGA, audio, network

Format and tools registries needed
- PRONOM for file type detection and application selection
- Open, regularly updated, broad user communities
Additional Components

- Additional information and metadata needed in software archive
  - Application handbooks
  - Howtos and trouble shooting guides
  - Application update packages
  - License keys, access codes
- Depending on object
  - Fonts for documents
  - Codecs for video, audio
  - Software extensions like DirectX, OpenGL libraries
• After object digest out of archive or user data of other sources
  – Transport into emulation environment
  – After or during environment setup
• Means of object transport
  – Virtual optical (ISO) or floppy disks as images
  – Disk container files
  – Network connections like FTP, SMB/CIFS
  – “Shared Folders” (as e.g. found in VMware or VirtualBox)
  – Copy&Paste (e.g. text areas in Dioscuri)
Transport Containers

- Data transport requires formats understood by the target environment, e.g.
  - Floppy disks, ubiquitous in for many platforms for a rather long period
  - Images easy to create and store
  - Optical disks: ISO images well understood by many emus

- More complex
  - Container files of the several emulators
  - Creator tools required
  - Adding objects to disk container files before emulators started
Reference Environment

- Emulation might require quite some steps until object is actually accessible
  - Average archive user is often not trained computer professional
  - Lots of problems to setup emulation environment on average machines
  - Many software components needed are proprietary
- Workstations with defined environment e.g. in library reading rooms
- Offer pre-created environments over the network
Web Access for Emulation

- Global remote access to emulation (services)
  - Access to different emulation environments like Dioscuri, MESS, QEMU, ...
    for Windows 3.11, Windows 98, C64, Atari, ...
  - Up- and download of objects over the net
  - PRONOM detection of object type and view path suggestion
  - Starting the appropriate emulator and software environment for object access
• Client Server application for remote operation
  – Client side: Java application executable in average browsers with JRE 1.5
  – Server side: Standard Linux environment to host the several emulators
  – Open Source
  – Extensible to more emulators, environments
  – Please have a look at our poster explaining the tool a little bit more!
• Additonal archival objects required for view path handling and required software environments
• Storage of view path caches for fast access in GRATE or specifically defined reference workstations
• View path aggregation of often used environments
• With this information – define and use metrics to calculate archive management costs
  – Differentiate view path options
  – Get cost structures to preserve certain object types
  – Evaluate shared, distributed archive approaches to local ones
Emulation in OAIS

- Emulation requires certain archive management activities on:
  - Ingest
  - Operation
  - Digest
- Emulation might require OAIS extension:
  - Suggestions from the list of archival tasks explained:
    - Check view paths on a regular base, especially if reference environments changed
    - Add emulators if required and migrate existing ones if needed, discard obsoleted
    - ...
Thank you!  Questions!

- Randolph Welte  
  rwelte@uni-freiburg.de
- Dirk von Suchodoletz  
  dsuchod@uni-freiburg.de

- www.planets-project.eu
- www.ks.uni-freiburg.de/projekte/fla