# Adding quality-awareness to evaluate migration web-services and remote emulation for digital preservation

**Christoph Becker**, Hannes Kulovits, Michael Kraxner, Riccardo Gottardi, Andreas Rauber, Randolph Welte*

Vienna University of Technology, Austria

http://www.ifs.tuwien.ac.at/dp

*University of Freiburg, Germany

http://www.uni-freiburg.de/

ECDL 2009
Κέρκυρα, Greece, September 2009

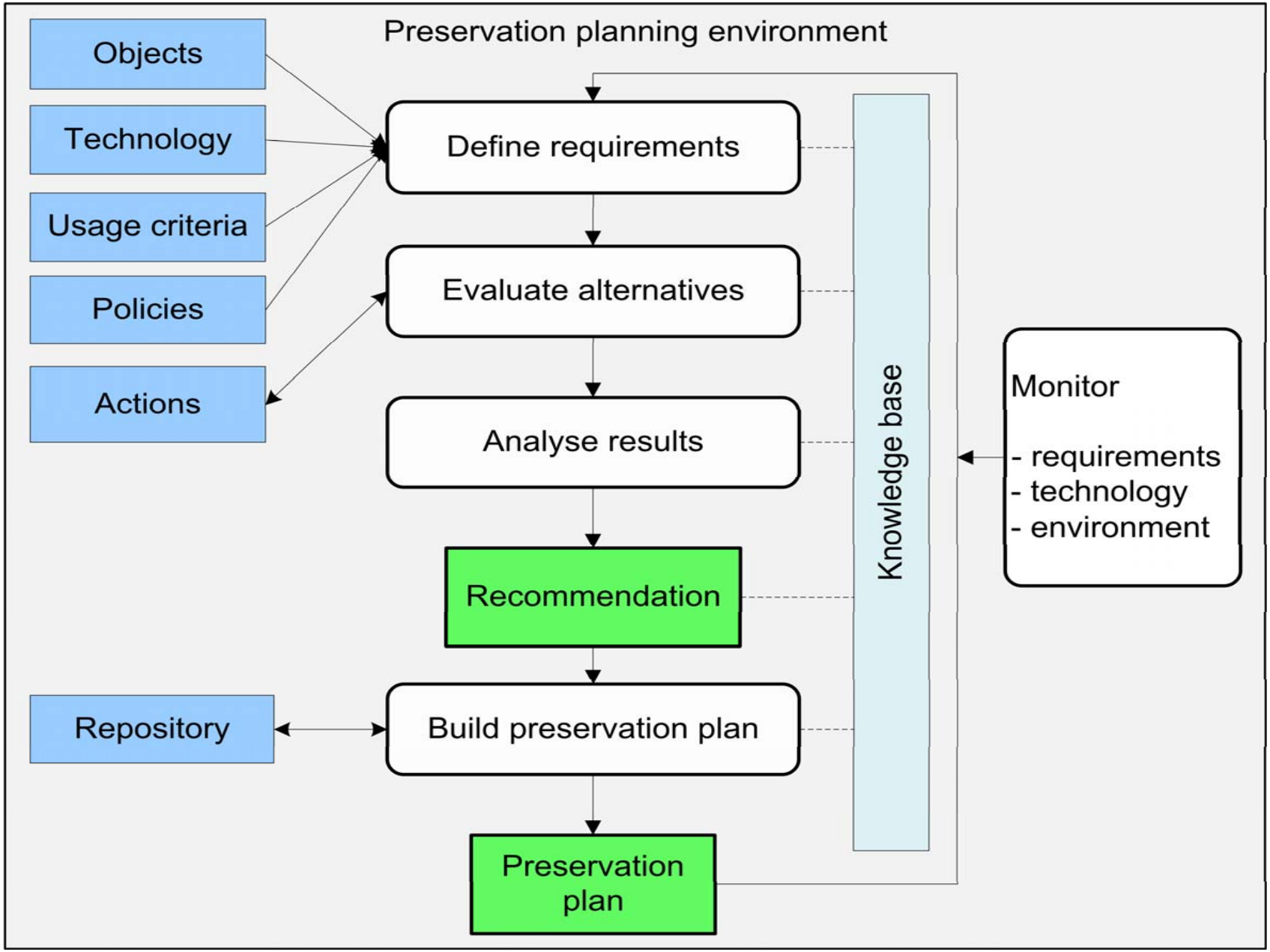
# Outline

- Digital preservation and distributed services
  - Preservation planning
  - Move towards distributed services
  - Shortcomings
- Preservation action monitoring infrastructure
  - Add quality-awareness to migration services
  - Integrate remote online acess to emulation

# Digital preservation and preservation planning

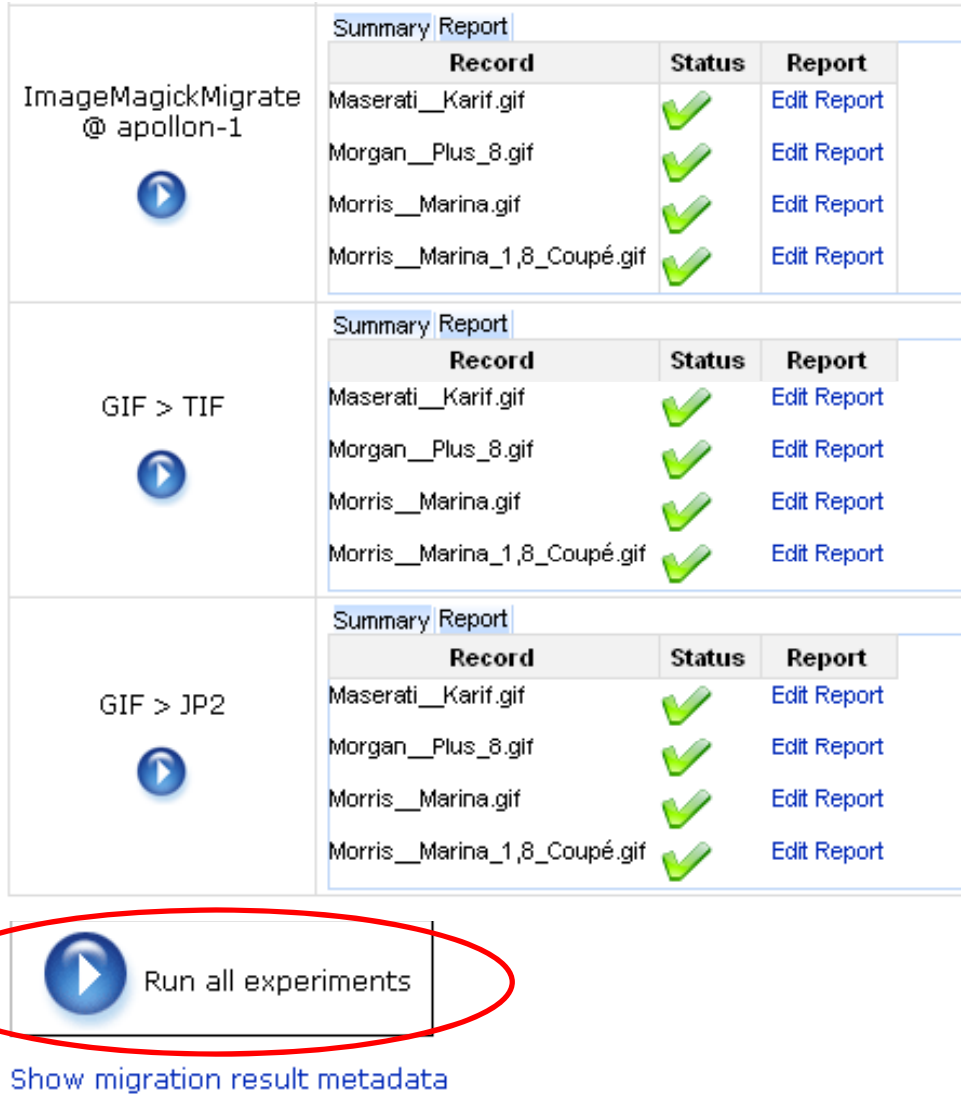
- Digital objects need an environment to function
- Environments change rapidly
  - Migration: change object
  - Emulation: recreate environment
- Many tools exist, and requirements vary
  - Trustworthy decision making procedures needed
  - Need for repeatability, transparency, automation
- Preservation planning
  - Evidence-based evaluation and selection
  - Controlled experimentation and automated measurements
  - Supported by distributed service-oriented environment
  - Planning tool PLATO: www.ifs.tuwien.ac.at/dp/plato

# Preservation planning



# Distributed preservation services

- Evaluation of preservation actions in PLATO is supported by distributed services
  - Identification
  - Migration
  - Characterisation



# Distributed preservation services: Gaps

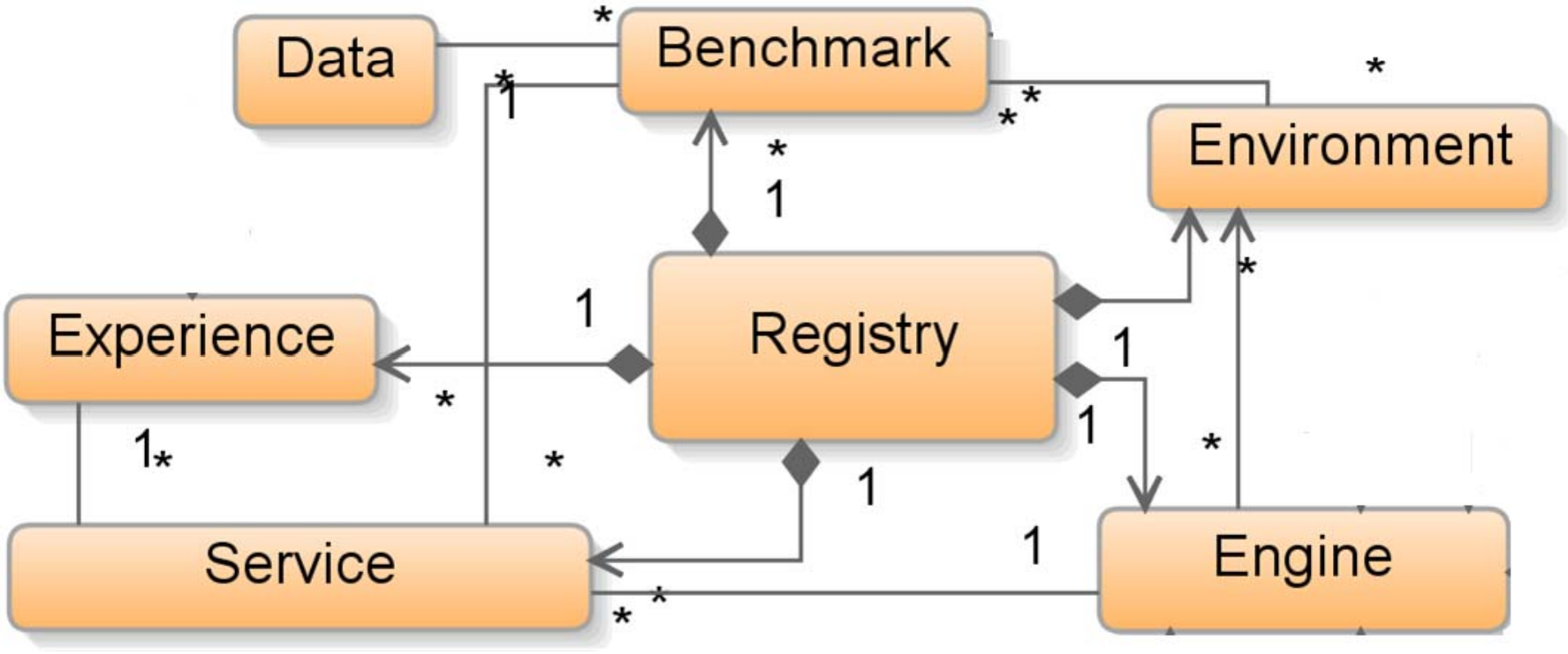
- Loose coupling, late binding, flexible integration...
  - Unknown quality and runtime characteristics
  - We need quality-aware services

- Lack of support for emulation
  - Cannot be executed automatically yet
  - Tedious setup procedures inhibit usage

1. Quality-aware migration services
2. Remote access to emulation

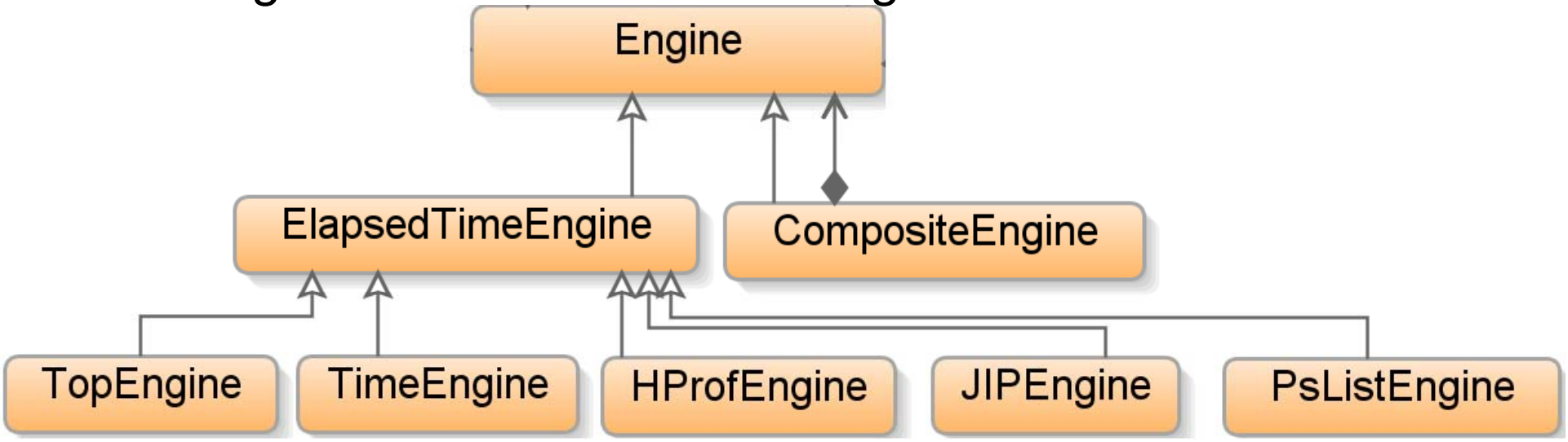
# Core elements of the monitoring framework

- Engines make services quality-aware
- Environments have associated benchmark scores
- Registry accumulates experience



# Types of engines

- Measurement techniques
- Non-invasive provider-side service instrumentation
- Measuring CPU time and memory usage:
  - Elapsed time
  - Linux, Unix: TOP, time
  - Windows: PsList
  - Java: JIP, HPROF
- Plugin structure: Additional engines can be added



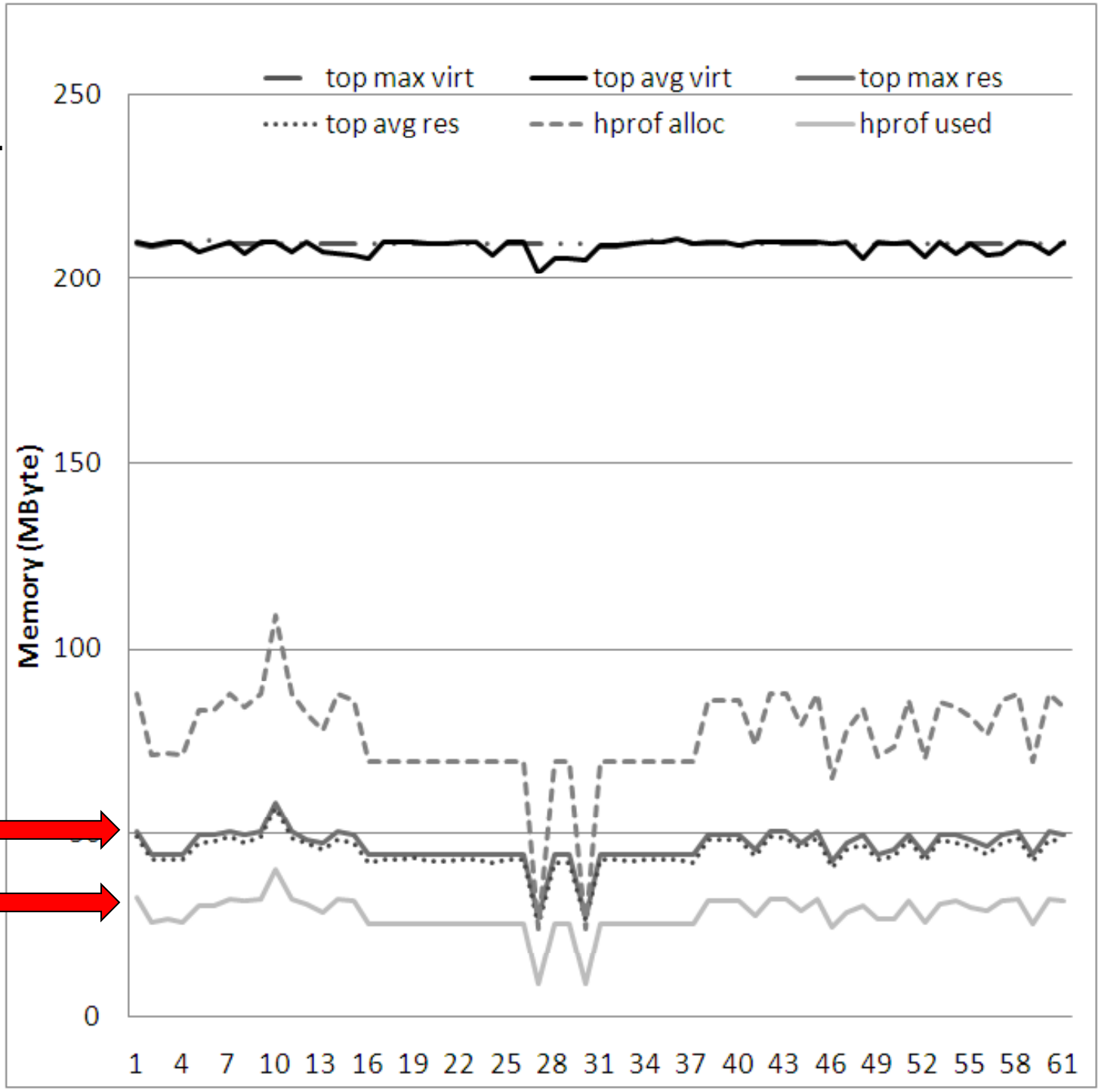
# Experiments

- Series of experiments on migration tools
- 50 to 300 files, 500-2800 MB input data volume
- Native applications on Linux and Windows
- Java programs
- Goals
  - Compare profiling tools
  - Select and verify metrics

  - Comparing performance
  - Accumulating experience

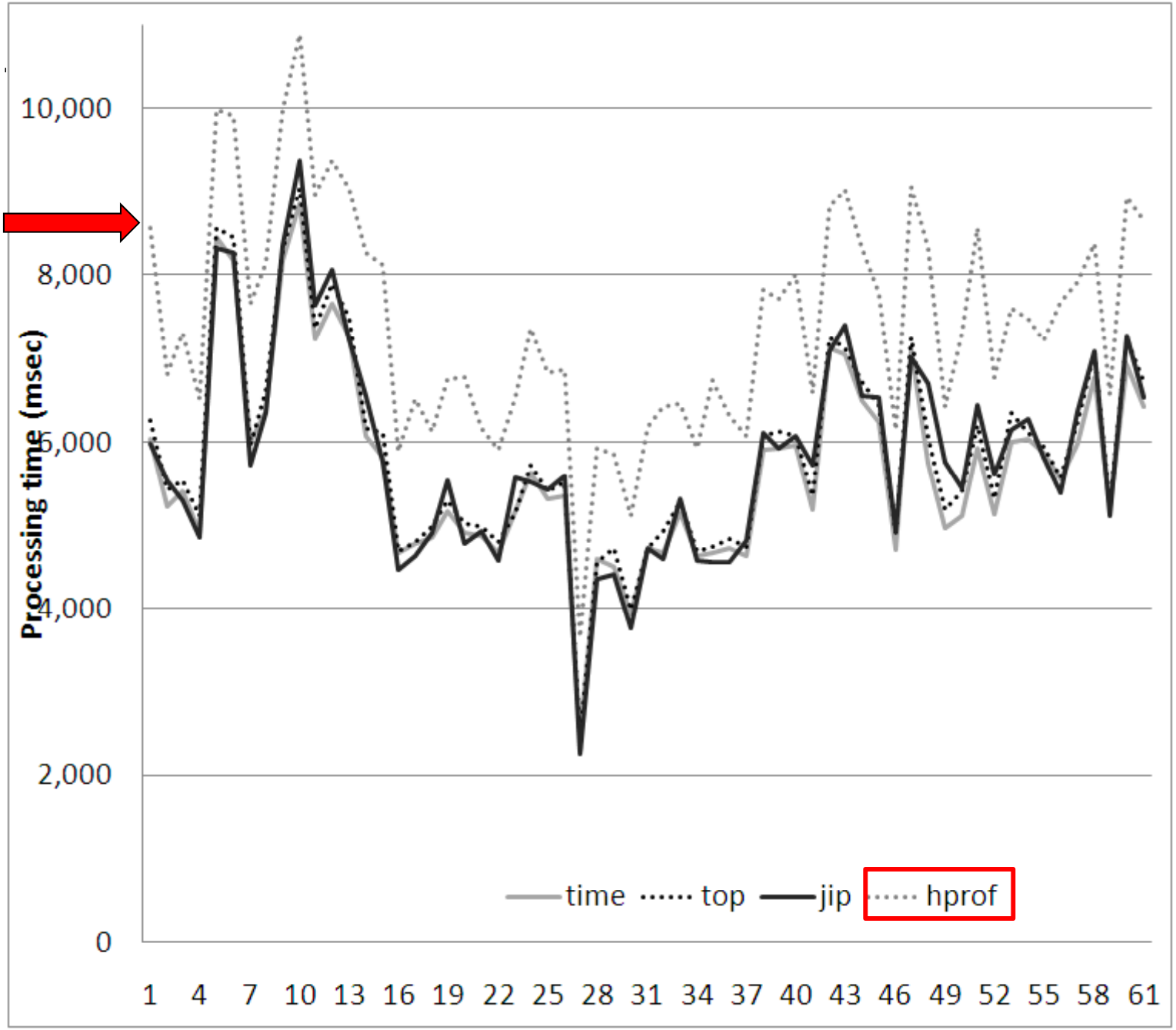
Profiling tools:

Memory usage of Java tools



Profiling tools:

Timing of Java tools



# Composite performance and quality

- „Heisenberg principle“ in profiling
- Composite engine forks execution and collects results

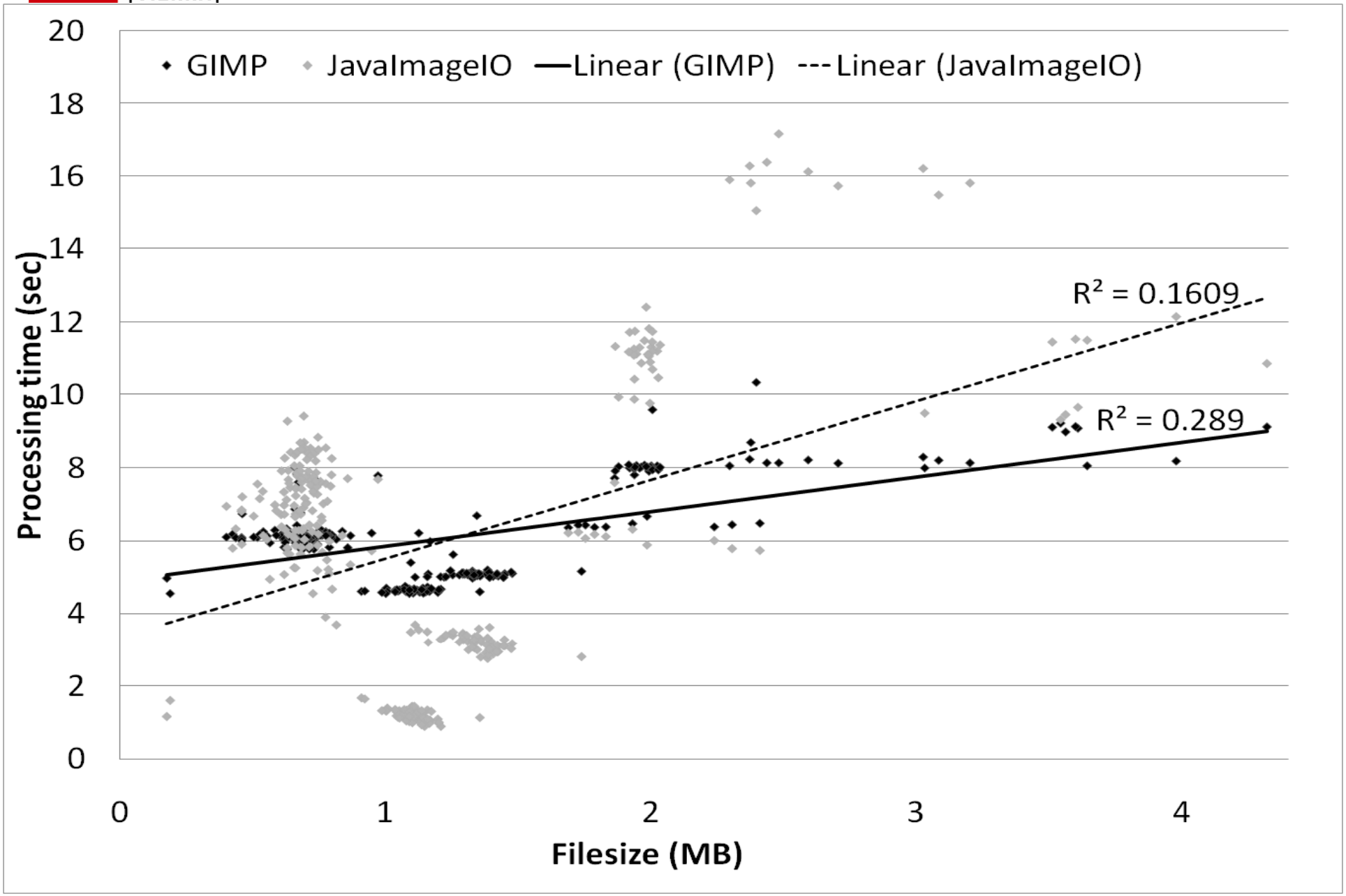
- Evaluators measure quality
  - XCL – eXtensible Characterisation Languages
    - Represent informational content of objects in XML
    - Compare original and transformed content
  - Other QA tools for digital preservation
    - ImageMagick
    - Jhove
    - ...

# Experiments

- Series of experiments on data conversion tools
- 50 to 300 files, 500-2800 MB input data volume
- Native applications on Linux and Windows
- Java programs
- Goals
  - Compare profiling tools
  - Select and verify metrics

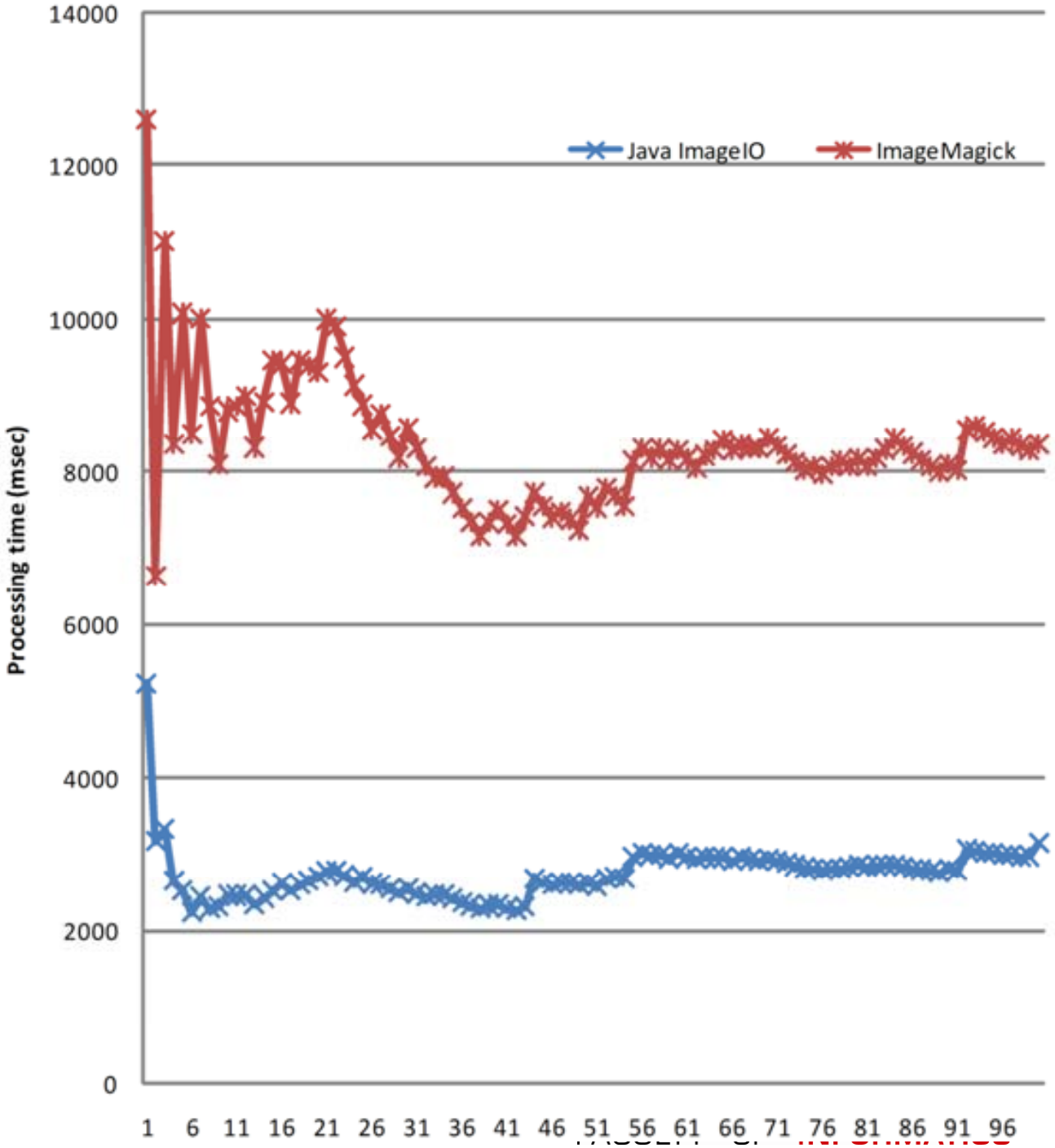
  - **Comparing performance**
  - **Accumulating experience**

# Comparing tool performance



# Accumulating experience

- Average processing time per Mbyte
- Monitoring and outlier detection



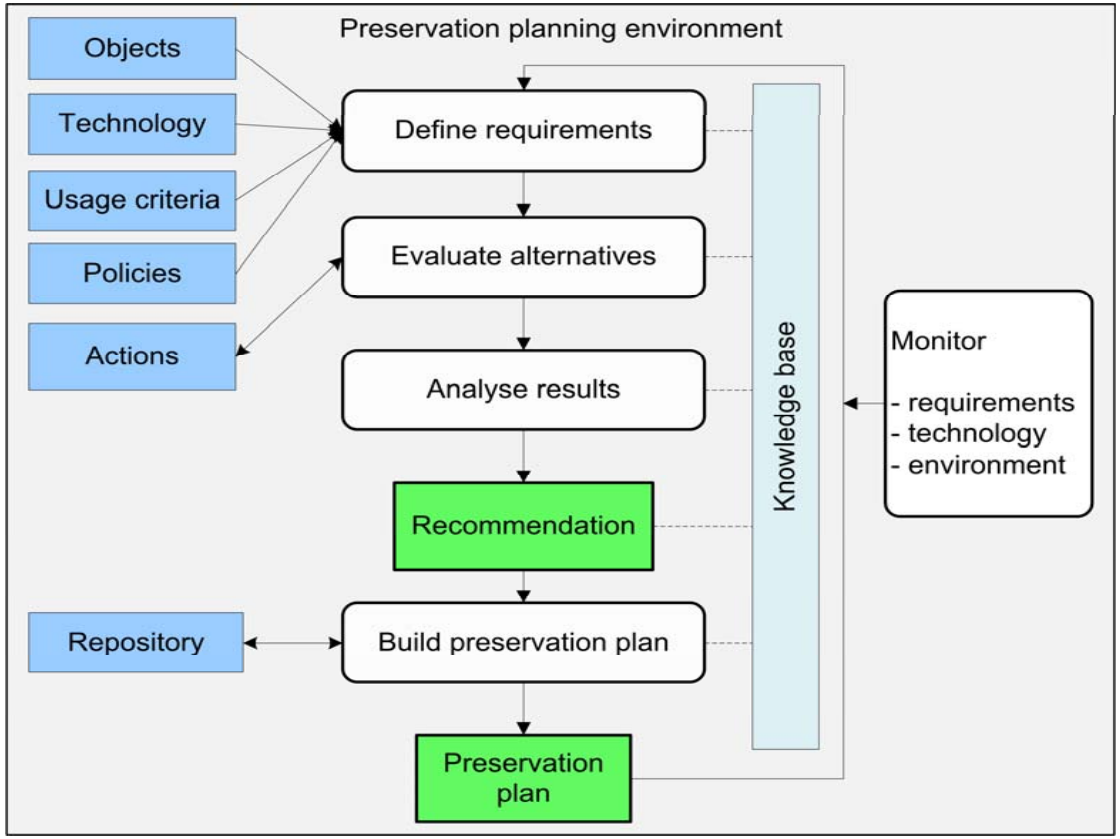
# Client-side measurements

- Provider-side instrumentation does not cover network latency, (un)marshalling, protocol layers etc.
- Server- and client-side measurements complementary
- Additional client-side measurements
  - Allow feedback and accumulation of measurements
  - Allow additional quality feedback
  - Prevent manipulation

- Each service response contains a generated key for adding client-side measurements

# Automated evaluation

- Measurements returned as metadata
- Mapping to defined quality criteria
- Visualisation supports analysis and comparison
- Select optimal tool
- Preservation plan defines monitoring criteria
- Deployed components continually monitored



# Visual analysis

- Planning tool **Plato:** www.ifs.tuwien.ac.at/dp/plato
- Upcoming release integrates quality-aware migration

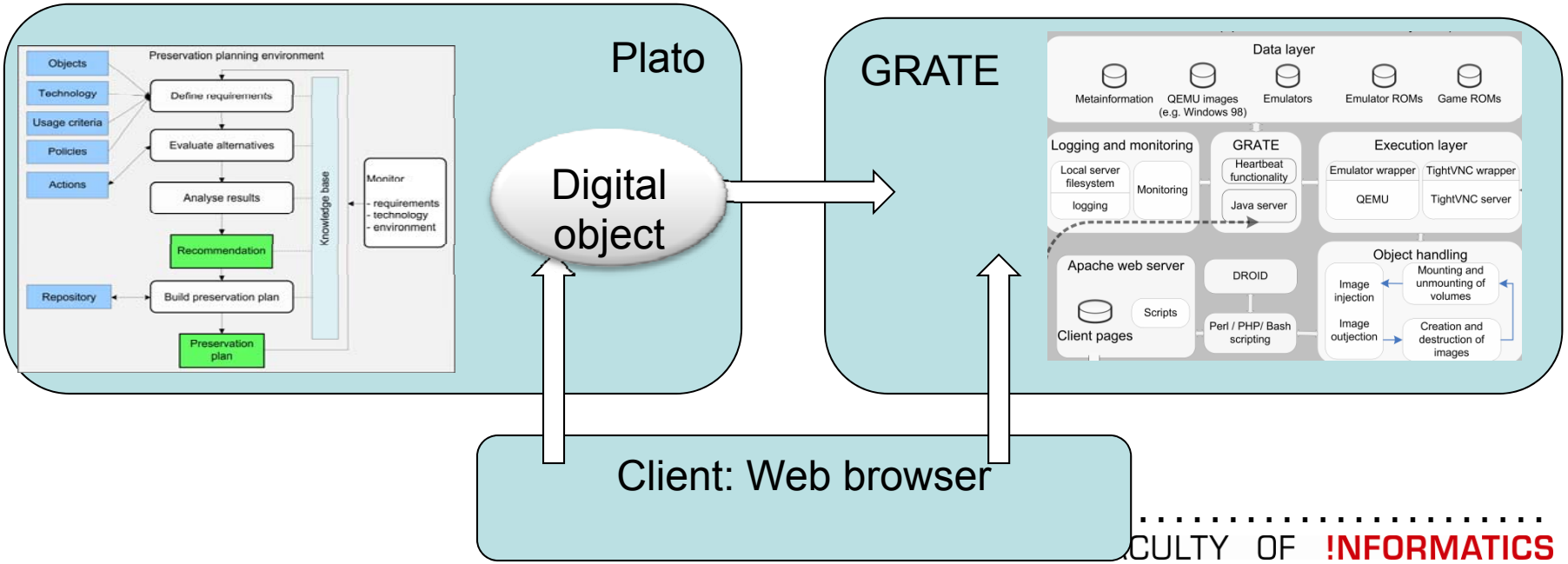
**Results: Weighted multiplication**

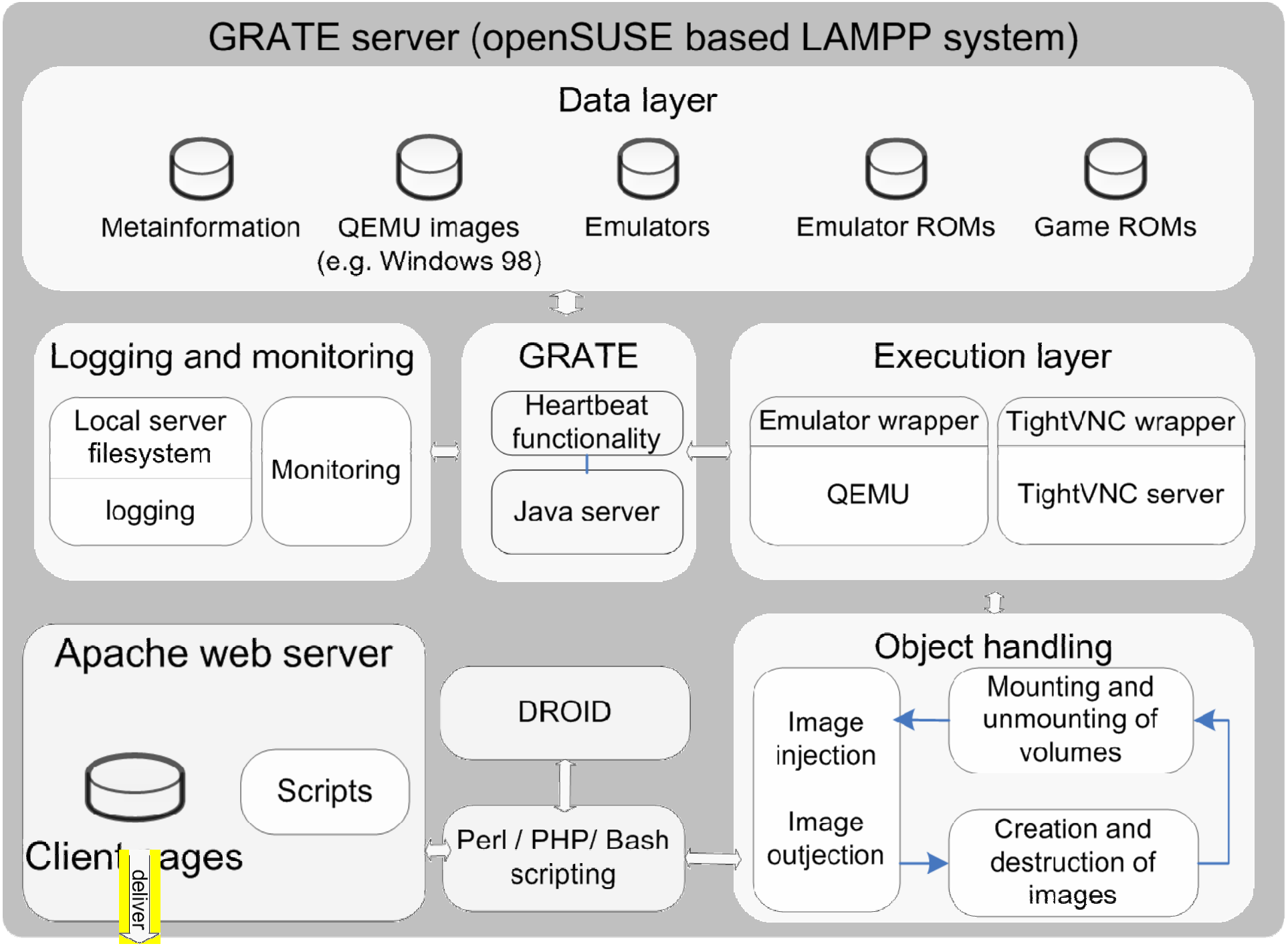
Result-Tree with all Alternatives, Aggregation method: Weighted multiplication
Expand All | Collapse All
**National Library Publications**

| Focus | Name | Result |
|---|---|---|
| | ⊟ National Library Publications | Adobe Acrobat->PDFA: 0.00<br>PdfMagiConversion: 3.44<br>Adobe Acrobat->HTML: 3.18 |
| X | ⊞ Object characteristics | Adobe Acrobat->PDFA: 1.55<br>PdfMagiConversion: 1.63<br>Adobe Acrobat->HTML: 1.52 |
| X | ⊞ Technical characteristics | Adobe Acrobat->PDFA: 1.14<br>PdfMagiConversion: 1.14<br>Adobe Acrobat->HTML: 1.16 |
| X | ⊟ Process Characteristics | Adobe Acrobat->PDFA: 0.00<br>PdfMagiConversion: 1.14<br>Adobe Acrobat->HTML: 1.08 |
| | Duration | Adobe Acrobat->PDFA: 0.00<br>PdfMagiConversion: 1.23<br>Adobe Acrobat->HTML: 1.06 |
| | Automation of the process | Adobe Acrobat->PDFA: 1.55<br>PdfMagiConversion: 1.90<br>Adobe Acrobat->HTML: 1.55 |
| X | ⊞ Integrity | Adobe Acrobat->PDFA: 1.00<br>PdfMagiConversion: 1.00<br>Adobe Acrobat->HTML: 1.00 |
| X | ⊞ Costs | Adobe Acrobat->PDFA: 1.67<br>PdfMagiConversion: 1.63<br>Adobe Acrobat->HTML: 1.67 |

# Remote access to emulation: GRATE

- Global Remote Access To Emulation
- University of Freiburg, Germany
- Enables dynamic loading of emulators on a dedicated server and immediate access via a browser interface
- Integrated into the Planets environment




GRATE server (openSUSE based LAMPP system)
Data layer
Metainformation
QEMU images
(e.g. Windows 98)
Emulators
Emulator ROMs
Game ROMs
Logging and monitoring
Local server filesystem
logging
Monitoring
GRATE
Heartbeat functionality
Java server
Execution layer
Emulator wrapper
QEMU
TightVNC wrapper
TightVNC server
Apache web server
Scripts
Client images
deliver
DROID
Perl / PHP/ Bash scripting
Object handling
Image injection
Image outjection
Mounting and unmounting of volumes
Creation and destruction of images

# Client: Web application

deliver

Inject keystrokes into OS

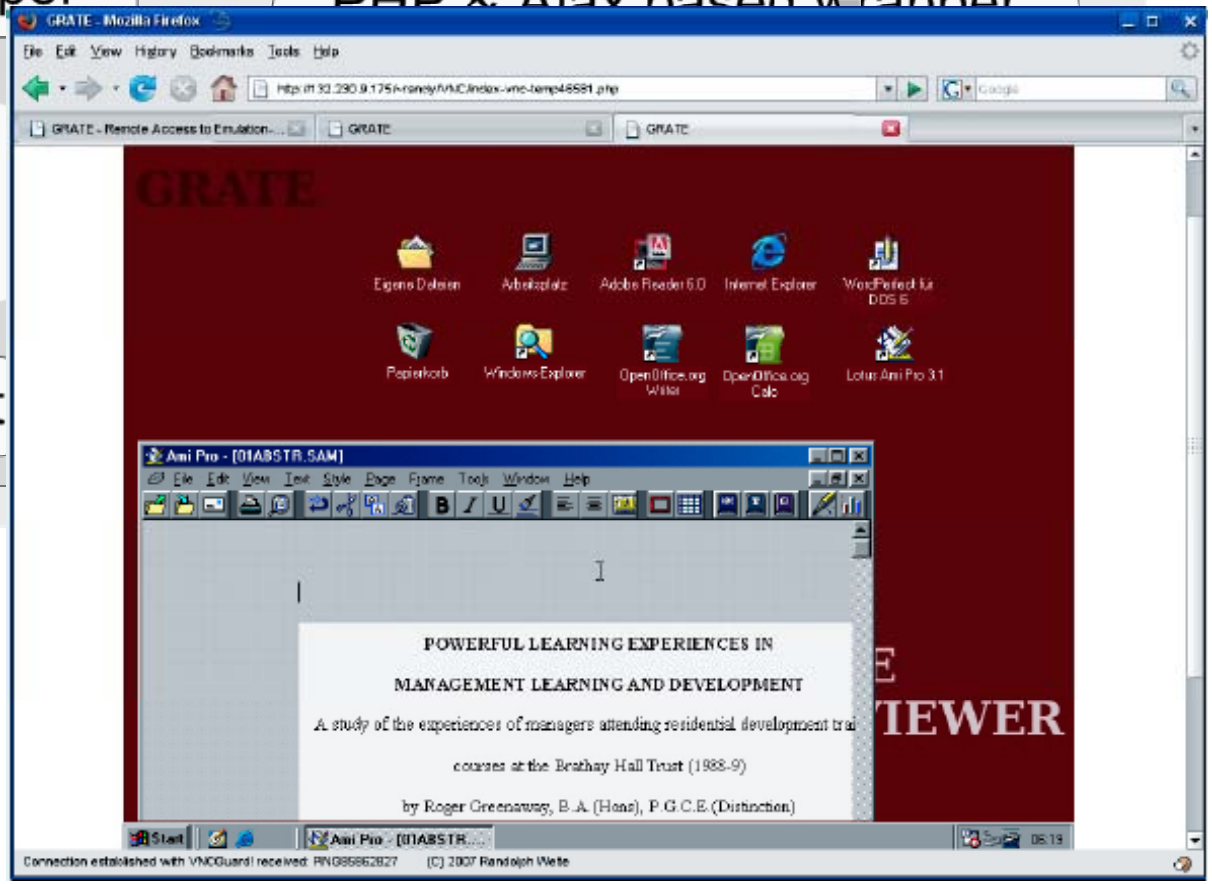
## Java-enabled web browser

PHP & Ajax based wrapper

Java (applet) based client

Heartbeat client (applet

PHP & Ajax based wrapper



# Summary

- Gaps in distributed services for preservation actions
  - Migration services lack quality and runtime information
  - Emulation support is low
- Closing the gaps
  - Direct invocation of emulators online
  - Support for the evaluation and monitoring of migration tools
- Planning tool PLATO
- www.ifs.tuwien.ac.at/dp/plato

# Questions?

becker@ifs.tuwien.ac.at
www.ifs.tuwien.ac.at/~becker