Plato: A Service Oriented Decision Support System for Preservation Planning

Christoph Becker, Hannes Kulovits, Andreas Rauber, Hans Hofman
ACM/IEEE Joint Conference on Digital Libraries (JCDL 2008)

Pittsburgh, PA, USA. June 16-20, 2008
Outline

- Digital Preservation and Preservation Planning
  - Evaluation of potential actions
- Plato: The Planets Preservation Planning Tool
  - Underlying methodology and workflow
  - Service discovery and integration
    - Characterisation
    - Preservation action
  - State of development and roadmap
- Current and future work
The Longevity of Digital Objects

- We create, shape, and exchange information digitally.
- Digital objects need technical environment to “function”.
- Heterogeneity and complexity of formats and environments and the speed of technological change make long-term access a challenge.

- Dominant types of preservation actions:
  - Migration
  - Emulation
Evaluating preservation strategies

- Variety of solutions and tools exist
- Each strategy has unique strengths and weaknesses
- Requirements vary across settings
- Decision on which solution to adopt is complex
- Documentation and accountability is essential

- Preservation planning assists in decision making
- Evaluating preservation strategies on representative samples according to specific requirements and criteria
Preservation Planning

Preservation Planning in Plato

Define requirements
- Define basis
- Choose records
- Identify requirements

Evaluate alternatives
- Go/No-Go
- Develop experiment
- Run experiment
- Evaluate experiment

Consider results
- Analyse results
- Set importance factors
- Transform measured values

Preservation Action Recommendation
- Create executable preservation plan
- Define preservation plan
- Validate preservation plan

Preservation Plan

Tree templates and fragments

Knowledge base

Administration
- Proposals
- Recommendations
- Inventory reports
- Performance info
- Consumer comments
- Technology alerts
- External data standards
- Prototype results
- Reports

Develop Preservation Strategies and Standards

Monitor Designated Community
- Monitoring
- Reports
- Requirement alerts
- Emerging standards

Prototype requests

Monitor Technology
- Prototype requests

PRODUCER

FACULTY OF INFORMATICS
Planets Preservation Planning Workflow

- Define requirements
  - Basis
  - Sample objects
  - Requirements
- Evaluate potential actions
- Analyse results
- Build a preservation plan
Preservation Planning in Plato

- Web based planning tool implementing the Planets preservation planning workflow
- Publicly available
- Ongoing development
- Integration of registries and services for
  - File format identification
  - Preservation action
  - Characterisation and comparison
- Frontend to a distributed architecture of preservation services
## Sample Records

| Full name: | sample 1 | |
| Short name: | eins.png | |
| Has data: | [download] | |
| Original technical environment: | | |
| Description: | | |

| Full name: | sample number two | |
| Short name: | zwei.png | |
| Has data: | [download] | |
| Original technical environment: | | |
| Description: | | |

PUID: mt/11
Name: Portable Network Graphics
Version: 1.0
Mime-type: image/png

Identify format

---

Preservation Planning in Plato

- Define requirements
  - Define basis
  - Choose tools

- Evaluate alternatives
  - Define alternatives
  - Define experiment

- Consider results
  - Collect results

- Analyze results
  - Sort importance factors

- Transform measured values

Multi-preservation plan

- Create executable preservation plan

---

**Red Circle:**
- Focus on the highlighted sample records and their attributes.
- The PUID for the first sample record is mt/11.
- The second sample record also has a PUID mt/11.
- The first sample record has a short name of eins.png and a file type of image/png.
- The second sample record has a short name of zwei.png and a file type of image/png.

---

**Identify format button:**
- Click to identify the format of the sample records.

---

**DROID logo:**
- Click for additional information or tools related to DROID.

---

**Preservation Plan:**
- Click for more details on the preservation plan process.
Requirements and Influence Factors

- Technology
- Standards
- User requirements
- Characteristics of digital objects

Technical characteristics
- Infrastructure characteristics
- Process characteristics

Requirements for preserving a collection of digital objects

Object characteristics
- Content
  - Appearance
  - Structure
  - Behaviour
  - Context

- Legal constraints
- Policies
- Organisational requirements
- Business needs, Budget constraints
## The Objective Tree in Plato

**PLANETS Preservation Planning Tool (Plato)**
Institute of Software Technology and Interactive Systems

**Identify Requirements**

<table>
<thead>
<tr>
<th>Focus</th>
<th>Node</th>
<th>Single</th>
<th>Scale</th>
<th>Restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Website</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Record characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appearance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Behaviour</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>deactivate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mailto</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>preserve</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>menus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pop-ups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>freeze</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>current date/time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>visitor counter</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Newsfeeds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ubiquity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Documentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ease of identification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ease of validation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Preservation Planning in Plato

Define requirements
- Define basis
- Choose records
- Identify requirements

Evaluate alternatives
- Go/No-Go
- Define alternatives
- Develop experiment
- Run experiment
- Evaluate experiment

Consider results
- Analyse results
- Set importance factors
- Transform measured values

Build preservation plan
- Create executable preservation plan
- Define preservation plan
- Validate preservation plan

Preservation Plan

Knowledge base

Tree templates and fragments
Mapping characteristics to requirements
Service discovery and invocation

Create alternatives from applicable services

Sample record #1 has format JPEG File Interchange Format, 1.01. You can look up services that are able to handle this object type in the following registries:

<table>
<thead>
<tr>
<th>Planets Preservation Action Tool registry</th>
<th>Preservation Action</th>
<th>Target Format</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planets Preservation Action Tool registry</td>
<td>JPG &gt; BMP</td>
<td>Windows Bitmap, version 3.0</td>
<td>JPG&gt;BMP</td>
</tr>
<tr>
<td>Planets Preservation Action Tool registry</td>
<td>JPG &gt; TIF</td>
<td>Tagged Image File Format, version 3</td>
<td>JPG&gt;BMP&gt;TIF</td>
</tr>
<tr>
<td>Planets Preservation Action Tool registry</td>
<td>JPG &gt; TIF #2</td>
<td>Tagged Image File Format, version 3</td>
<td>JPG&gt;TIF</td>
</tr>
<tr>
<td>Planets Preservation Action Tool registry</td>
<td>JPG &gt; TIF_2</td>
<td>Tagged Image File Format, version 3</td>
<td>JPG&gt;TIF_2</td>
</tr>
<tr>
<td>Planets Preservation Action Tool registry</td>
<td>JPG &gt; PNG</td>
<td>Portable Network Graphics, version 1.0</td>
<td>JPG&gt;PNG</td>
</tr>
</tbody>
</table>

CRiB Service Registry

Create alternatives for selected services
Evaluation of actions

- Apply actions on sample content
- Evaluate outcomes

- Significant properties of objects
  - Content
  - Appearance
  - Structure
  - Behaviour
  - Context
- Image width in Pixel, colour depth in bit...
- Technical characteristics extracted from objects
Comparing migrated documents

ODF

MigratorA

PDF/A

XCEL for ODF

Extractor

XCEL for PDF/A

Comparator

Content in XCDL

Content in XCDL
### Analyse Results

**Aggregation method:** Sum

<table>
<thead>
<tr>
<th>Select</th>
<th>Alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>PDF/A ToolA</td>
</tr>
<tr>
<td>✓</td>
<td>PDF/A ToolB</td>
</tr>
</tbody>
</table>

### Minimalist root node

#### Fokus

<table>
<thead>
<tr>
<th>Name</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimalist root node</td>
<td></td>
</tr>
<tr>
<td>Image properties</td>
<td></td>
</tr>
<tr>
<td>Amount of Pixel</td>
<td></td>
</tr>
<tr>
<td>Karma</td>
<td></td>
</tr>
<tr>
<td>Filesize (in Relation to Original)</td>
<td></td>
</tr>
<tr>
<td>A Single-Leaf</td>
<td></td>
</tr>
<tr>
<td>Intrange 0-10</td>
<td></td>
</tr>
</tbody>
</table>

### Recommendation

**Recommendation:**

**Reasoning:**
Summary

- Preservation planning in Plato
- Workflow
- Characterisation
  - Format identification
  - Collection profiling
  - Risk assessment
  - XCL and comparison
- Discovering applicable actions
- Building a preservation plan
Future work

- Pluggable infrastructure for the automated evaluation of preservation actions
- Integrated knowledge base and recommender systems
- Case studies on electronic documents, image archives, electronic art, computer games...

- Plato 1.3 in July
- Plato 2.0 scheduled for October 2008
  - Service integration
  - Preservation plan
- Plato 3.0 early 2010
Questions?

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

www.ifs.tuwien.ac.at/dp/plato

www.planets-project.eu