The role of objective trees in preservation planning

Hans Hofman
DCC/DPE/DRIVER/Nestor Joint workshop
Berlin, 28 November 2007
Evaluating preservation actions

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

- Evaluation of preservation actions on representative sample content according to specific requirements
- Part of decision making in preservation planning
Decision support for preservation planning

- Systematic procedure for evaluating preservation actions/strategies
  - By conducting experiments on sample content
  - Based on the Dutch Testbed and subsequently applied in DELOS

- Case studies
  - Electronic documents, interactive art, web archives...
  - Identify essential characteristics and requirements for preservation strategies
  - Validate methodology and workflow

- Development of software tool
  - Plato – Planning Tool
  - Web application supporting the workflow
Phase 1: Define requirements

1. Define basis
   - Describe Collection (profile)
   - Institutional settings

2. Choose sample objects/records
   - Representative for the type of objects that requires action
   - Right choice of samples is essential

3. Define requirements
   - “Objective tree”
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
Stakeholders

- Input from a wide range of persons, depending on the institutional context and the objects

Administration
- IT Staff
- Managers
- Lawyers

Domain experts
- Technical characteristics
- Infrastructure characteristics
- Process characteristics
- Requirements for preserving a collection of digital objects
- Object characteristics

Curators
- Content
- Appearance
- Structure
- Behaviour
- Context

Producers
- Others

Technical experts
- Consumers

Others
Phase 2: Evaluate Alternatives

4. Define Alternatives
5. Go/No-Go decision
6. Develop experiment
7. Run experiment
8. Evaluate experiment
Phase 3: Consider Results

9. Transform measured values to a unified scale to make them comparable

10. Set importance factors to model the relative importance of siblings in each branch

11. Analyse results
Transform measured values

- Measures come in seconds, euro, bits, goodness values,…
- Need to make them comparable
- Transform measured values to uniform scale
- Transformation tables for each leaf criterion
- Scale 0-5 (0 is *unacceptable*)

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Target value</th>
</tr>
</thead>
<tbody>
<tr>
<td>256.0</td>
<td>px -&gt; 1</td>
</tr>
<tr>
<td>512.0</td>
<td>px -&gt; 2</td>
</tr>
<tr>
<td>1024.0</td>
<td>px -&gt; 3</td>
</tr>
<tr>
<td>2048.0</td>
<td>px -&gt; 4</td>
</tr>
<tr>
<td>4096.0</td>
<td>px -&gt; 5</td>
</tr>
</tbody>
</table>

[Diagram of the process flow]
Transformation

PLANETS Preservation Planning Tool (Plato)

Transform Measured Values

Expand All | Collapse All
Minimalist root node

<table>
<thead>
<tr>
<th>Focus</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>▼ Minimalist root node</td>
</tr>
<tr>
<td>X</td>
<td>▼ Image properties</td>
</tr>
<tr>
<td>X</td>
<td>▼ Karma</td>
</tr>
<tr>
<td>X</td>
<td>▼ Filesize (in Relation to Original)</td>
</tr>
<tr>
<td>X</td>
<td>▼ A Single-Leaf</td>
</tr>
<tr>
<td>X</td>
<td>▼ IntRange 0-10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image properties</th>
<th>Amount of Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>Target value</td>
</tr>
<tr>
<td>256.0</td>
<td>px -&gt; 1</td>
</tr>
<tr>
<td>512.0</td>
<td>px -&gt; 2</td>
</tr>
<tr>
<td>1024.0</td>
<td>px -&gt; 3</td>
</tr>
<tr>
<td>2048.0</td>
<td>px -&gt; 4</td>
</tr>
<tr>
<td>4096.0</td>
<td>px -&gt; 5</td>
</tr>
</tbody>
</table>

Threshold stepping:
-  Steps
-  Linear

Aggregation mode:
-  Worst result
-  Arithmetic mean

Minimalist root node > Karma

<table>
<thead>
<tr>
<th>Ordinal Value</th>
<th>Target Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>4.0</td>
</tr>
<tr>
<td>Bad</td>
<td>2.0</td>
</tr>
<tr>
<td>Evil</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Aggregation mode:
-  Worst result
-  Arithmetic mean

Results: 1 2
PDF/A ToolA 1024 2048
PDF/A ToolB 2048 2048

Minimalist root node > Filesize (in Relation to Original)
Analyse Results

• Aggregate values
  – Multiply the transformed measured values in the leaf nodes with the leaf weights
  – Sum up the transformed weighted values over all branches of the tree
• Rank alternatives according to overall performance value at root
• Performance of each alternative
  – overall
  – for each sub-criterion (branch)
• Comparison of different alternatives
### Analyse Results

**Sum**
- PDF/A (Tool A)
- PDF/A (Tool B)

**Minimalist root node**

<table>
<thead>
<tr>
<th>Focus</th>
<th>Name</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>▼</td>
<td>Minimalist root node</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▼ Image properties</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▼ Karma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▼ Filesize (in Relation to Original)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▼ A Single-Leaf</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▼ IntRange 0-10</td>
<td></td>
</tr>
</tbody>
</table>
Integrating Planets concepts and services
Summary

• Systematic approach for identifying all criteria that will influence preservation planning

• Workflow for evaluating and choosing preservation actions

• Tool support: Plato
  • 1st version end of November 2007 (project internal)
  • 2nd version publicly available, second half of 2008

• Planets: developing one integrated environment for preservation planning
Thank you very much for your attention.

hans.hofman@nationaalarchief.nl

www.planets-project.eu