Trustworthy Preservation
Planning with Plato

Andreas Rauber
Department of Software Technology and Interactive Systems

Vienna University of Technology
rauber@ifs.tuwien.ac.at
http://www.ifs.tuwien.ac.at/~andi
Outline

- Introduction
- Why do we need preservation planning?
- Preservation planning and Plato
- Bringing it all together and closing
Why do we need Digital Preservation?
Why do we need Digital Preservation?
Why do we need Digital Preservation?

- Digital Objects require specific environment to be accessible:
  - Files need specific programs
  - Programs need specific operating systems (-versions)
  - Operating systems need specific hardware components

- SW/HW environment is not stable:
  - Files cannot be opened anymore
  - Embedded objects are no longer accessible/linked
  - Programs won’t run
  - Information in digital form is lost (usually total loss, no degradation)

- Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.
Why Preservation Planning?

- Several preservation strategies developed
  - For each strategy: several tools available
    - For each tool: several parameter settings available
- How do you know which one is most suitable?
- What are the needs of your users? Now? In the future?
- Which aspects of an object do you want to preserve?
- What are the requirements?
- How to prove in 10, 20, 50, 100 years, that the decision was correct / acceptable at the time it was made?
Consistent workflow leading to a preservation plan

Analyses, which solution to adopt

Considers
- preservation policies
- legal obligations
- organisational and technical constraints
- user requirements and preservation goals

Describes the
- preservation context
- evaluated preservation strategies
- resulting decision including the reasoning

Repeatable, solid evidence

Trust and audit
Preservation Planning

- Trust and Audit
- Compliance to best practices, standards
- 3 core initiatives, of which 2 prescriptive
  - RLG- National Archives and Records Administration Digital Repository Certification Task Force: Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
  - NESTOR: Catalogue of Criteria of Trusted Digital Repositories
  - DCC/DPE: DRAMBORA: Digital Repository Audit Method Based on Risk Assessment
- Embedding into OAIS model
Preservation Planning

Preservation Planning in Plato

- Define requirements
  - Define basis
  - Choose records
  - Identify requirements
  - Tree templates and fragments

- Evaluate alternatives
  - Go/No-Go
  - Define alternatives
  - Develop experiment
  - Run experiment
  - Evaluate experiment
  - Mapping characteristics to requirements

- 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

Administration
- Proposals
- Recommendations
- Inventory reports
- Performance info
- Consumer comments

Knowledge base
- Develop Preservation Strategies and Standards
  - Technology alerts
  - External data standards
  - Prototype results

Monitor
- Requirement alerts
- Emerging standards

Designated Community
- Reports
- Prototype requests
- Monitor Technology
  - Prototype requests

PRODUCER
- Product technologies
- Surveys

PROFESSORSHIP
Outline

- Introduction
- Why do we need preservation planning?
- Preservation planning and Plato
- Bringing it all together and closing
Plato

- Preservation Planning Tool
- Reference implementation of planning workflow
- Web-based application, 1st release 2.0 Nov. 12 2008, latest version 2.1 released November 2009
- Documents the process and ensures that all steps are considered
- Automation of the planning process via integration of registries and services
- Knowledge base to support planning
- Creates a preservation plan (XML, PDF)
- [http://www.ifs.tuwien.ac.at/dp/plato](http://www.ifs.tuwien.ac.at/dp/plato)
Preservation Planning with Plato

Plato

- Assists in analyzing the collection
  - Profiling, analysis of sample objects via Pronom and other services
- Allows creation of objective tree
  - Within application or via import of mindmaps
- Allows the selection of Preservation action tools
Plato

- Runs experiments and documents results
- Allows definition of transformation rules, weightings
- Performs evaluation, sensitivity analysis,
- Provides recommendation (ranks solutions)
PP Workflow

Preservation Planning in Plato

Define requirements
- Define basis
- Choose records
- Identify requirements

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

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

Preservation Action Recommendation

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

Preservation Plan

Tree templates and fragments

Mapping characteristics to requirements

Knowledge base

FACULTY OF INFORMATICS
Define Basis

- Basic preservation plan properties
- Describe the context
  - Institutional settings
  - Legal obligations
  - User groups, target community
  - Organisational constraints
- 5 triggers
  - New Collection Alert (NCA)
  - Changed Collection Profile Alert (CPA)
  - Changed Environment Alert (CEA)
  - Changed Objective Alert (COA)
  - Periodic Review Alert (PRA)
Choose Sample Objects

- Identify consistent (sub-)collections
  - Homogeneous type of objects (format, use)
  - To be handled with a specific (set of) tools

- Describe the collection
  - What types of objects?
  - How many?
  - Which format(s)?

- Selection
  - Representative for the objects in the collection
  - Right choice of sample is essential
  - They should cover all essential features and characteristics of the collection in question
  - As few as possible, as many as needed
  - Often between 3 – 10
Choose Sample Objects

- Stratification – all essential groups of digital objects should be chosen according to their relevance

- Possible stratification strategies
  - File type
  - Size
  - Content (e.g. document with lots of images, including macros)
  - Time (objects from different periods of times)

- File Format Identification
  - DROID
  - PRONOM
PP Workflow

Preservation Planning in Plato

Define requirements
- Define basis
- Choose records
- Identify requirements
  - Tree templates and fragments

Evaluate alternatives
- Go/No-Go
- Develop experiment
- Run experiment
  - Evaluate experiment
    - Mapping characteristics to requirements

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

Preservation Action Recommendation

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

Knowledge base
Identify Requirements

- Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain
- Put the requirements in relation to each other → Tree structure
- Top-down or bottom-up
  - Start from high-level goals and break down to specific criteria
  - Collect criteria and organize in tree structure
Input needed from a wide range of persons, depending on the institutional context and the collection.
Identify requirements

- Core step in the process
- Define all relevant goals and characteristics (high-level, detail) with respect to given application domain
- Usually four major groups
  - Object characteristics (content, metadata,…)
  - Record characteristics (context, relations,…)
  - Process characteristics (scalability, error-detection,…)
  - Costs (set-up, per object, HW/SW; personnel,…)

![Diagram](image-url)
Identify requirements

- Appearance
- Structure
- Behaviour
- Authenticity
- Stability
- Scalability
- Usability
- Technical costs
- Personnel costs

... or analogue...

... or digital
Identify requirements

- Creation within PLATO with Tree-Editor
Identify requirements

- Assign measurable unit to each leaf criterion
  - As far as possible automatically measurable
    - seconds / Euro per object
    - colour depth in bits
    - ...
  - Subjective measurement units where necessary
    - diffusion of file format
    - amount of expected support
    - ...
  - No limitations on the type of scale used
Identify requirements

- Creation within PLATO with Tree-Editor

**Planets Preservation Planning Tool (Plato)**

Identify Requirements

- Objective Tree
  - Descriptive Information

Focus | Node | Single | Scale | Restriction | Unit
---|---|---|---|---|---
Website | Record characteristics | | | | |
Website | Technical characteristics | | | | |
Website | Ubiquity | | | | |
Website | Support | | | | |
Website | Documentation | | | | |
Website | Stability | | | | |
Website | Ease of identification | | | | |
Website | Ease of validation | | | | |
Website | Lossiness | | | | |
Website | IPR | | | | |
Website | Complexity | | | | |

Release 1.1 - Institute of Software Technology and Interactive Systems: «office bears»
Identify Requirements: Example

Ubiquitous/Widespread/Specialised/Obsolete → Ubiquity

Primary/Secondary → Quality

Full/Partial/None → Disclosure

Standard/Open/Proprietary → Openness

Public/Limited/None → Availability

<1 year/1-2 years/3-5 years/>5 years → Speed of change

None/Previous version only/Some previous versions/All previous versions → Backwards compatibility

Automatic/Manual/No → Ease of identification

Automatic/Manual/No → Ease of validation

Lossy/Lossless → Lossiness

Y/N → IPR

High/Medium/Low → Complexity

None/Detectable/Recoverable → Error tolerance

Larger/Same/Smaller → Comparative size
Visitor counter and similar functionalities can be

- Frozen at harvesting time
- Omitted
- Remain operational, i.e. the counter will be increased upon archival calls
  (is this desired? count? demonstrate functionality?)
Define Alternatives

PLANETS Preservation Planning Tool (Plato)

Define the alternatives of the Project

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>195615</td>
<td>TIFF (tool A)</td>
<td>Convert to TIFF using the well-tested and expensive tool 'A'</td>
</tr>
<tr>
<td>195613</td>
<td>TIFF (tool B)</td>
<td>Convert to TIFF/4 using this new tool named 'B'</td>
</tr>
<tr>
<td>195614</td>
<td>GIF (tool C)</td>
<td>Convert to GIF using the well-tested tool 'C'</td>
</tr>
<tr>
<td>195615</td>
<td>PNG (tool D)</td>
<td>Convert to PNG using the well-tested tool 'D'</td>
</tr>
</tbody>
</table>

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:

Create alternatives for selected services

Release 1.2 - Institute of Software Technology and Interactive Systems: «office bears»

FACULTY OF INFORMATICS
PP Workflow

Preservation Planning in Plato

Define requirements
- Define basis
- Choose records
- Identify requirements
- Tree templates and fragments

Evaluate alternatives
- Develop experiment
- Run experiment
- Go/No-Go
- Define alternatives
- Evaluate experiment
- Mapping characteristics to requirements

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

Preservation Action Recommendation

Build preservation plan
- Create executable preservation plan
- Define preservation plan
- Validate preservation plan
- Preservation Plan
Develop and Run Experiment

- Call the migration tools / emulators
- Convert the sample objects / open in emulators
- Take a look at the results. each of the sample object in various versions
- Measure time / memory needed to migrate/open
- Measure program output, error messages, …
PP Workflow

Preservation Planning in Plato

Define requirements
- Define basis
- Choose records
- Identify requirements
- Tree templates and fragments

Evaluate alternatives
- Go/No-Go
- Define alternatives
- Develop experiment
- Run experiment
- Evaluate experiment
- Mapping characteristics to requirements

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

Preservation Action Recommendation
- Build preservation plan
  - Create executable preservation plan
  - Define preservation plan
  - Validate preservation plan
- Preservation Plan
Evaluate Experiment

PLANEES Preservation Planning Tool (Plato)

Evaluate Experiment
Expand All | Collapse All
Polar bear image preservation

<table>
<thead>
<tr>
<th>Focus</th>
<th>Node</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Polar bear image preservation</td>
</tr>
<tr>
<td></td>
<td>Process</td>
</tr>
<tr>
<td></td>
<td>Complexity</td>
</tr>
<tr>
<td></td>
<td>Cost</td>
</tr>
<tr>
<td></td>
<td>Image properties</td>
</tr>
<tr>
<td></td>
<td>Bits of colour depth</td>
</tr>
<tr>
<td></td>
<td>Technical characteristics</td>
</tr>
<tr>
<td></td>
<td>Official standard</td>
</tr>
<tr>
<td></td>
<td>Filesize (in Relation to Original)</td>
</tr>
</tbody>
</table>

Comments:

<table>
<thead>
<tr>
<th>Process &gt; Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative</td>
</tr>
<tr>
<td>TIFF (tool A)</td>
</tr>
<tr>
<td>TIFF (tool B)</td>
</tr>
<tr>
<td>GIF (tool C)</td>
</tr>
<tr>
<td>PNG (tool D)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Process &gt; Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative</td>
</tr>
<tr>
<td>TIFF (tool A)</td>
</tr>
<tr>
<td>TIFF (tool B)</td>
</tr>
<tr>
<td>GIF (tool C)</td>
</tr>
<tr>
<td>PNG (tool D)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image properties &gt; Bits of colour depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative</td>
</tr>
<tr>
<td>TIFF (tool A)</td>
</tr>
<tr>
<td>TIFF (tool B)</td>
</tr>
<tr>
<td>GIF (tool C)</td>
</tr>
<tr>
<td>PNG (tool D)</td>
</tr>
</tbody>
</table>
PP Workflow

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

Preservation Action Recommendation

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

Preservation Plan

Tree templates and fragments
Mapping characteristics to requirements

Knowledge base
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
- Linear transformation, logarithmic, special scale
- Scale 1-5 plus "not-acceptable"
PP Workflow

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
  - Mapping characteristics to requirements

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

Preservation Action Recommendation

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

Tree templates and fragments

Knowledge base

FACULTY OF INFORMATICS
Set Importance Factors

- Not all leaf criteria are equally important
- By default, weights are distributed equally
- Adjust relative importance of all siblings in a branch
- Weights are propagated down the tree to the leaves
## Analyse Results

**Planets Preservation Planning Tool (Plato)**

### Analyze Results

**Aggregation method:** Multiplication

**How do the aggregation mechanisms work?**

<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

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

### Example: Electronic documents

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Total Score Weighted Sum</th>
<th>Total Score Weighted Multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDF/A (Adobe Acrobat 7 prof.)</td>
<td>4.52</td>
<td>4.31</td>
</tr>
<tr>
<td>PDF (unchanged)</td>
<td>4.53</td>
<td>0.00</td>
</tr>
<tr>
<td>TIFF (Document Converter 4.1)</td>
<td>4.26</td>
<td>3.93</td>
</tr>
<tr>
<td>EPS (Adobe Acrobat 7 prof.)</td>
<td>4.22</td>
<td>3.99</td>
</tr>
<tr>
<td>JPEG 2000 (Adobe Acrobat 7 prof.)</td>
<td>4.17</td>
<td>3.77</td>
</tr>
<tr>
<td>RTF (Adobe Acrobat 7 prof.)</td>
<td>3.43</td>
<td>0.00</td>
</tr>
<tr>
<td>RTF (ConvertDoc 4.1)</td>
<td>3.38</td>
<td>0.00</td>
</tr>
<tr>
<td>TXT (Adobe Acrobat 7 prof.)</td>
<td>3.28</td>
<td>0.00</td>
</tr>
</tbody>
</table>

- Deactivation of scripting and security are knock-out criterion (PDF)
- RTF is weak in *Appearance* and *Structure*
- Plain text doesn’t satisfy several minimum requirements
PP Workflow
Outline

- Introduction
- Why do we need preservation planning?
- Preservation planning and Plato
- Bringing it all together and closing
Digital Preservation

What is a preservation plan?

- 10 Sections
  - Identification
  - Status
  - Description of Institutional Setting
  - Description of Collection
  - Requirements for Preservation
  - Evidence for Preservation Strategy
  - Cost
  - Trigger for Re-evaluation
  - Roles and Responsibilities
  - Preservation Action Plan

Preservation Plan Template
What we have now:

- Basic Preservation Plan:
  - PDF: Preservation Plan.pdf
  - XML: Preservation Plan.xml

- That was developed in a solid, repeatable and documented process
- That is optimal for the needs of a given institution and for the data at hand
Conclusions

- Physical preservation ensures longevity of resources
- Simple risk analysis reporting
- Preservation Planning to ensure “optimal” preservation
- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation
- Basis for well-informed, accountable decisions
- Follows recommendations of TRAC and nestor
- **Plato:**
  - Tool support to perform solid, well-documented analyses
  - Creates core preservation plan
Thank you!

http://www ifs tuwien ac at/dp