Preservation planning 3

What to measure and how
Monitoring and Scalability
Multi-criteria decision making

April 22, 2013
Christoph Becker
Vienna University of Technology
www.ifis.tuwien.ac.at/~becker
How can we select the optimal preservation action for a given setting?

- What are the drivers and constraints on the decision space?
- What are the goals and objectives?
- What are the factors influencing the decision makers’ preferences?
- How can we model multiple competing objectives and requirements?
- How should we evaluate software components?
How can we ensure trustworthy preservation planning?

- What are the requirements on trust that need to be addressed?
- What decision steps and evidence need to be documented?
- What are the aspects that a plan needs to address, and what are the elements needed to cover them?
- How can we ensure reliable evaluation procedures and repeatable evidence?
Preservation Planning: Question 3

- How can we ensure that decision processes scale up?
  - How can we automate decision making?
  - How can we integrate continuous monitoring?
  - Which properties can be measured automatically, and how?
  - How can we create a controlled environment for observing the behaviour of components in a reproducible way?
Preservation Planning: Key concepts

- Repeatable, standardized planning workflow
- A weighted hierarchy of objectives
  - Measurable criteria on the leaf level of the tree
  - Utility functions make criteria comparable
- Controlled experimentation on sample content
  - Evidence-based decision making
- Standardized structure for plan specification
  - Transparency and documentation
  - Comparability across scenarios
  - Integration with repository systems
- Planning tool Plato guides, validates, documents
  - www.ifs.tuwien.ac.at/dp/plato
- Automation: Reduce manual effort
Levels of control

- **Guidance**
  - Strategies
  - Compliance
  - Regulations

- **Control**
  - Specific preservation policies

- **Planning**
  - Preservation Planning
  - Monitoring

- **Operations**
  - Preservation actions
  - Quality Assurance
  - Content analysis
A planning experiment
Database study

Content branch
### Results: Weighted multiplication

Result-Tree with all Alternatives, Aggregation method: Weighted multiplication.

<table>
<thead>
<tr>
<th>Node</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>Archive to XML: 3.88</td>
</tr>
<tr>
<td>Object characteristics</td>
<td>Archive to XML: 2.14</td>
</tr>
<tr>
<td>Content</td>
<td>Archive to XML: 2.38</td>
</tr>
<tr>
<td>appearance</td>
<td>Archive to XML: 1.14</td>
</tr>
<tr>
<td>context</td>
<td>Archive to XML: 1.23</td>
</tr>
<tr>
<td>behaviour</td>
<td>Archive to XML: 1.00</td>
</tr>
<tr>
<td>Format characteristics</td>
<td>Archive to XML: 1.36</td>
</tr>
<tr>
<td>Tool characteristics</td>
<td>Archive to XML: 1.34</td>
</tr>
</tbody>
</table>

### Results: Weighted sum

Result-Tree with all Alternatives, Aggregation method: Weighted sum.
This tree contains only strategies that do not have knock-out evaluation criteria; see above.

<table>
<thead>
<tr>
<th>Node</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>Archive to XML: 4.25</td>
</tr>
<tr>
<td>Object characteristics</td>
<td>Archive to XML: 2.40</td>
</tr>
<tr>
<td>Content</td>
<td>Archive to XML: 2.63</td>
</tr>
<tr>
<td>appearance</td>
<td>Archive to XML: 0.41</td>
</tr>
<tr>
<td>context</td>
<td>Archive to XML: 0.69</td>
</tr>
<tr>
<td>behaviour</td>
<td>Archive to XML: 0.07</td>
</tr>
<tr>
<td>Format characteristics</td>
<td>Archive to XML: 0.95</td>
</tr>
</tbody>
</table>
Scanned books requirements

Technical characteristics
- Browser support
- Ubiquity
- Open standard
- OCR applicable
- Reversible migration
- Compression
- ISO standard
- Relative file size
- Master = access copy
- Stability

Object characteristics
- Image size identical
- Image identical
- Additional meta data
- Color depth identical
- Resolution identical
- Resolution unit identical
- Color profile identical
- Existent OCR still valid

Content
- Retain filename
- Pixel compare possible
- Open Source
- Batch support
- Duration
- Stability
- Log output
- Compatibility with server environment

Process characteristics
- External expertise necessary
- Migration process
- Adoption of archival system after migration
- Costs
### Results: Weighted multiplication

**Node**

#### Scans

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>4.50</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>3.71</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:0.00</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>3.68</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>3.65</td>
</tr>
</tbody>
</table>

#### Object characteristics

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>1.59</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>1.38</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:0.00</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>1.38</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>1.38</td>
</tr>
</tbody>
</table>

#### Content

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>2.24</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>1.63</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:0.00</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>1.63</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>1.63</td>
</tr>
</tbody>
</table>

#### image size identical

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>1.19</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>1.19</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:0.00</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>1.19</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>1.19</td>
</tr>
</tbody>
</table>

#### image identical

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>1.19</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>1.19</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:0.00</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>1.19</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>1.19</td>
</tr>
</tbody>
</table>

#### additional meta data

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>1.19</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>1.08</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:1.08</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>1.08</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>1.08</td>
</tr>
</tbody>
</table>

#### color depth identical

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep status quo:</td>
<td>1.19</td>
</tr>
<tr>
<td>ImageMagick - TIFF to JP2:</td>
<td>1.19</td>
</tr>
<tr>
<td>GraphicsMagick - TIFF to JP2:1.19</td>
<td>&lt;1.00</td>
</tr>
<tr>
<td>Kakadu - TIFF to JP2:</td>
<td>1.19</td>
</tr>
<tr>
<td>GeoJasper - TIFF to JP2:</td>
<td>1.19</td>
</tr>
</tbody>
</table>
Four cases, three solutions: Scanned images

- Bavarian State Library, 72TB TIFF6: Leave and monitor
- British Library, 80TB TIFF5: Migrate to JP2 (ImageMagick)
- Royal Library of Denmark, ~10,000 aerial photographs in TIFF6: Leave and monitor
- State and University Library Denmark, scanned yearbooks in GIF: Migrate to TIFF 6

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Chosen action</th>
<th>Main reasons</th>
</tr>
</thead>
<tbody>
<tr>
<td>72 TB scanned book pages in TIFF6</td>
<td>Leave unchanged and monitor</td>
<td>Color profile complications, lack of JP2 browser support, Process costs</td>
</tr>
<tr>
<td>80 TB scanned newspapers in TIFF5</td>
<td>Migrate to JP2</td>
<td>Storage costs, Standardization</td>
</tr>
<tr>
<td>Aerial photographs in TIFF6</td>
<td>Leave unchanged and monitor</td>
<td>Lack of JP2 browser support, Process costs</td>
</tr>
</tbody>
</table>
Software quality in preservation scenarios

- ISO 25010 SQUARE: Software quality model
  - *(Business factors not part of SQUARE)*
- Need to define indicators and metrics for specific domain
Action Properties

- Functional Correctness
- Functional Appropriateness

Generic

Functional Completeness

- Specific
- Content

Functional Suitability

Maintainability

- Modularity
- Reusability
- Analysability
- Modifiability
- Testability

Maturity

Stability Indicators

Reliability

- Availability
- Error Tolerance
- Fault Tolerance
- Recoverability

Operability

- Portability
- Installability
- Replaceability

Compatibility

- Interoperability
- Interfaces

Resource Utilization

- Performance
- Efficiency

Capacity

Time Behaviour

- Time per MB
- Time per Sample

Throughput

Memory Utilization

- Memory per Sample
- Resource Utilization

Batch Processing Maximum Files

Co-existence

- Directory Watch
- Interfaces

API

- .Net API
- Java API
- Other API

Supported Platforms

- Ease of Integration in Current Environment
- Installability
- Portability

Replaceability

Appropriateness Recognizability

- Availability
- Quality
- Documentation
- Learnability
- Ease of Use
- Ease of Administration
- Configuration Method
- Object Configuration Method

User Error Protection

- User Interface Aesthetics
- Accessibility
- Licensing Schema

Setup

Execution

Costs

Business

Modular Design

Plug-in System

Customization

Plug-in System

Configuration Method

Outcome Object

Specific

Content

Support Available

- Under Development
- Release Cycle
- Active Community
- Manufacturer Status

Number of Input-Formats Supported

- Number of Output-Formats Supported
- Macro Support
- Input Object Validation
- Output File Format Verification
- Retain Filename
- Activity Traceability
- Activity Logging Format
- Activity Logging Amount
- Error Reporting
- Understandability of Reported Errors
- Reported Errors Contain Handling Advice
- Number of Objects Supported
- Fraction of Objects Supported
- Keyboard Support
- Mouse Support
- Sound Support
- CD-ROM Support

Refer to Category
Scanned books requirements

**Technical characteristics**
- Browser support
- Ubiquity
- Open standard
- OCR applicable
- Reversible migration
- Compression
- ISO standard
- Relative file size
- Master = access copy
- Stability

**Process characteristics**
- HW
- SW
- External expertise necessary
- Migration process
- Adoption of archival system after migration

**Object characteristics**
- Content
  - Image size identical
  - Image identical
  - Additional meta data
  - Color depth identical
  - Resolution identical
  - Resolution unit identical
  - Color profile identical
  - Existent OCR still valid

**Context**
- Retain filename
- Pixel compare possible
- Open source
- Batch support
- Duration
- Stability
- Log output
- Compatibility with server environment
Decision criteria: What to measure?

- Each criterion concerns either the action or its outcome
  - **Outcome**
    - **Object** (authenticity, editability, …)
    - **Format** (licensing, standardisation, complexity…)
    - **Effect** (Costs…)
  - **Action**
    - **Runtime** properties (performance, stability, logging…)
    - **Static** (price, license…)
    - **Judgement** (configuration interface usability…)
How to measure?

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Data collection and measurement</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Object → Outcome → Action

Format → Effect → Runtime

Effect → Action → Static

Action → Outcome → Judgement

Criterion
How to measure?

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Data collection and measurement</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>Image pixelwise identical Footnotes preserved</td>
<td>Measurements of output and input, comparison</td>
<td>FITS, JHove, image comparison...</td>
</tr>
<tr>
<td>Object</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### How to measure?

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Data collection and measurement</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome</td>
<td>Image pixelwise identical Footnotes preserved</td>
<td>Measurements of output and input, comparison</td>
<td>FITS, JHove, ImageMagick...</td>
</tr>
<tr>
<td>Object</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Format</td>
<td>Format is ISO standardised</td>
<td>Measurements of the output, Trusted external data sources</td>
<td>DROID, PRONOM, UDFR, P2</td>
</tr>
<tr>
<td>Effect</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Some file format requirements

- Specifications available
  - Is an XML schema enough?
  - Syntacs and semantics needed
- Standardized (ISO, ANSI, ITEF, ...)
- Accepted and widely used (indicators?)
- Not covered by patent
- Free of compression
- Free of any cryptographical techniques
- Flexible and extensible?
Data sources

- PRONOM
  - Sparse data
- www.digitalpreservation.gov/formats
  - Incomplete
- Wikipedia
  - reliable?
- The web
  - unstructured
- P2: Combination of PRONOM with dbpedia
  - Linked Data
  - ~45,000 statements
  - Still far from complete
<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Data collection and measurement</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Object</td>
<td>Image pixelwise identical</td>
<td>Measurements of output and input, comparison</td>
<td>FITS, JHove, ImageMagick...</td>
</tr>
<tr>
<td></td>
<td>Footnotes preserved</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome Format</td>
<td>Format is ISO standardised</td>
<td>Measurements of the output, Trusted external data sources</td>
<td>DROID, PRONOM, UDFR, P2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome effect</td>
<td>Annual bitstream preservation costs (€)</td>
<td>Measurements of the output, external data sources, models (LIFE)...</td>
<td>LIFE model</td>
</tr>
</tbody>
</table>

**How to measure?**

- **Criterion**: The focus is on measuring the impact of different tools and methods on the preservation of digital assets.
- **Object**: The focus is on the preservation of digital objects, including their format and content.
- **Format**: The measurement focuses on the standardization of digital formats.
- **Effect**: The measurement focuses on the impact of preservation actions on the cost and effectiveness of preservation.
- **Action**: The measurement includes tools and methods used for preservation.
- **Runtime**: The measurement includes the runtime of preservation actions.
- **Static**: The measurement includes static data sources.
- **Judgement**: The measurement includes the final judgement on the effectiveness of preservation actions.

**Tools**

- FITS, JHove, ImageMagick...
- DROID, PRONOM, UDFR, P2
- LIFE model
## How to measure?

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Data collection and measurement</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Object</td>
<td>Image pixelwise identical Footnotes preserved</td>
<td>Measurements of output and input, comparison</td>
<td>FITS, JHove, ImageMagick...</td>
</tr>
<tr>
<td>Outcome Format</td>
<td>Format is ISO standardised</td>
<td>Measurements of the output, Trusted external data sources</td>
<td>DROID, PRONOM, UDFR, P2</td>
</tr>
<tr>
<td>Outcome effect</td>
<td>Annual bitstream preservation costs (€)</td>
<td>Measurements of the output, external data sources, models (LIFE)...</td>
<td>LIFE model</td>
</tr>
<tr>
<td>Action runtime</td>
<td>Throughput (MB per millisecond), Memory usage</td>
<td>Measurements taken in controlled experimentation</td>
<td>MiniMEE</td>
</tr>
</tbody>
</table>
Profiling memory usage of Java tools
Profiling timing of Java tools
Comparing tool performance

- GIMP
- JavaImageIO
- Linear (GIMP)
- Linear (JavaImageIO)

$R^2 = 0.1609$

$R^2 = 0.289$
### How to measure?

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Data collection and measurement</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Object</td>
<td>Image pixelwise identical Footnotes preserved</td>
<td>Measurements of output and input, comparison</td>
<td>FITS, JHove, ImageMagick...</td>
</tr>
<tr>
<td>Outcome Format</td>
<td>Format is ISO standardised</td>
<td>Measurements of the output, Trusted external data sources</td>
<td>DROID, PRONOM, UDFR, P2</td>
</tr>
<tr>
<td>Outcome effect</td>
<td>Annual bitstream preservation costs (€)</td>
<td>Measurements of the output, external data sources, models (LIFE)...</td>
<td>LIFE model</td>
</tr>
<tr>
<td>Action runtime</td>
<td>Throughput (MB per millisecond), Memory usage</td>
<td>Measurements taken in controlled experimentation</td>
<td>MiniMEE</td>
</tr>
<tr>
<td>Action static</td>
<td>License costs per CPU (€), Open Source License</td>
<td>Trusted external data sources, manual evaluation, sharing</td>
<td>UDFR, Pronom, P2, manual</td>
</tr>
<tr>
<td>Category</td>
<td>Example</td>
<td>Data collection and measurement</td>
<td>Tools</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Outcome Object</td>
<td>Image pixelwise identical Footnotes preserved</td>
<td>Measurements of output and input, comparison</td>
<td>FITS, JHove, ImageMagick...</td>
</tr>
<tr>
<td>Outcome Format</td>
<td>Format is ISO standardised</td>
<td>Measurements of the output, Trusted external data sources</td>
<td>DROID, PRONOM, LoC format site, UDFR, P2</td>
</tr>
<tr>
<td>Outcome effect</td>
<td>Annual bitstream preservation costs (€)</td>
<td>Measurements of the output, external data sources, models (LIFE)...</td>
<td>LIFE model</td>
</tr>
<tr>
<td>Action runtime</td>
<td>Throughput (MB per millisecond), Memory usage</td>
<td>Measurements taken in controlled experimentation</td>
<td>MiniMEE</td>
</tr>
<tr>
<td>Action static</td>
<td>License costs per CPU (€), Open Source License</td>
<td>Trusted external data sources, manual evaluation, sharing</td>
<td>UDFR, P2, manual</td>
</tr>
<tr>
<td>Action judgement</td>
<td>Technical interoperability, configuration flexibility</td>
<td>Manual judgement, sharing</td>
<td></td>
</tr>
</tbody>
</table>
Core requirement: Keep object intact

- Essential object characteristics
  - Content
  - Appearance
  - Structure
  - Behaviour
  - Context
Validating a migrated image

- We run a TIFF file against a JPEG 2000 conversion tool and check:
  - Yes, it’s in JPEG 2000 format
  - Yes, it’s well-formed
  - Yes, it’s valid
  - Yes, it still has the same dimensions
- .... But is it still the same image?
Is manual evaluation feasible?

- 10 seconds per object, 1 million objects
- ~2800 hours ~ 350 full working days ~ 70 weeks of work
- 5 minutes per object: ~53,000 full working days...

We need evaluation:

- At decision time
- In operation (QoS)

Manual SLA checks?

The answer is: NO

(Diagram by Natasa Milic-Frayling, MSRC)
Approaches to Quality Assurance

- **Interaction complexity**
  - Structural complexity
  - Structural analysis of representation
  - (Algorithmic) analysis on perceptual level
  - Observation of behaviour
  - Interactive art
  - Electronic documents
  - Digital photographs
  - Scanned images
Validating a migrated image

- Dimensions, metadata... easy: extract and compare
- Content... Not always easy
- ImageMagick compare: good for simple cases

<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Metric</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Absolute Error</td>
<td>The number of different pixels (0 means identical images). This value can be thresholded to only count pixels that have a difference larger than a specified threshold.</td>
</tr>
<tr>
<td>PAE</td>
<td>Peak Absolute Error</td>
<td>The highest difference of any single pixel.</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak Signal to Noise Ratio</td>
<td>The ratio of mean square difference to the maximum mean square that can exist between any two images, expressed as a decibel value. The higher the PSNR, the closer the images are, with a maximum difference occurring at 1.</td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
<td>Average over all pixels</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Squared Error</td>
<td>Averaged squared error distance</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root mean squared error</td>
<td>Identical to $\sqrt{\text{MSE}}$.</td>
</tr>
</tbody>
</table>
Evaluating conversion processes

- Approaches to automated Quality Assurance
  - Characterization: Static analysis of representations
    - Semantics not fully contained in files
    - No homomorphic property mapping between (lots of) formats
    - Lack of ground truth
  - Perceptual-level analysis of standardised renderings
Distance metrics: How meaningful?

- AE
- PAE
- RMSE
- ... SSIM

Anything but "0" is a problematic result
A real migration workflow... What’s wrong?

Load TIFF6 from SAN → identify → Characterise TIFF

→ Compare images → > Identify + characterise → Migrate TIFF to JP2 >

→ > CHECK JP2 ok? → > Store piped result to JP2 file → Load next >
Criteria and solutions

- Problem space vs. solution space
- Criteria for objects, representation instances and formats
  - **Representation instance properties**: Aspects of the representation such as encoding, encryption, bytestream length, compression, validity, profile conformance. See [http://publik.tuwien.ac.at/files/PubDat_203384.pdf](http://publik.tuwien.ac.at/files/PubDat_203384.pdf)
  - **Format properties**: Properties of the formats themselves, not of the objects (status of standardisation, support by viewing environments, license, tool support, features, versions,…). [www.digitalpreservation.gov/formats](http://www.digitalpreservation.gov/formats)
Conclusions: Preservation planning

- From strategy and policies to operations
- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation for informed and accountable decisions
- Generic workflow that can be integrated in different institutional settings
- **Plato:** Tool support to perform solid, well-documented analysis
- Provides basic preservation plan

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

- De-facto standard preservation planning tool
- ~1000 user accounts
- Case studies
- Opportunities to leverage scale
- Analyse data set from diverse communities and build improvements

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

• Creating a plan is effort-intensive

• Sharing experience is difficult

• Monitoring changes is manual

• Integrating context, strategies and operations is difficult
Planning Challenges

• Creating a plan is effort-intensive
  ➢ Increase efficiency of planning

• Sharing experience is difficult
  ➢ Increase standardisation and reusability

• Monitoring changes is manual
  ➢ Introduce automation

• Integrating context, strategies and operations is difficult
  ➢ Systematic governance structures
  ➢ Manage policies
  ➢ Integrate systems
Preservation planning environment

Objects

Technology

Usage criteria

Policies

Actions

Define requirements

Evaluate alternatives

Analyse results

Recommendation

Build preservation plan

Preservation plan

Knowledge base

Monitor
- requirements
- technology
- environment

Repository
An Automated Monitoring service

- Knowledge base
  - Entities and their properties
  - Measures of properties over time
  - Triggers define conditions and events
- Flexible and extensible
  - A well-defined, flexible data model
  - Adaptors for different information sources
- Monitoring Capabilities
  - Internal Monitoring
  - External Monitoring
  - Monitor compliance, risks and opportunities
### Compliance, risk and opportunities

<table>
<thead>
<tr>
<th>PLAN</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated?</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Alternative 1</td>
<td>✔</td>
<td>✔</td>
<td>✗</td>
<td>✔</td>
</tr>
<tr>
<td>Alternative 2</td>
<td>✗</td>
<td>✗</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>Alternative 3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td>Alternative 4</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

- Opportunities for operations (new action tool)
- Risks to operations (errors uncovered in QA tool)
- Opportunities for operations (new QA tool)

Planning needs to generate Service Level Agreements (SLAs) and monitoring conditions.
Information sources to be monitored

Format registries
Content profiles
Components
Workflows
Experiments
Policies
Planning
Operations

Knowledge base

Plato
Web Client

Watch Frontend
A preservation decision space

- Define goals, scope, mandate
- Preservation intent
- Significant properties
- Format profiles
- Risk profiles
- Select actions
- Evaluate actions
- Actionable plans
- Service Level Agreements
- Preservation actions
- Quality assurance

Scale requires efficiency and automation

Strategy

Tactic

Preservation Context and Management
- Monitoring

Preservation Planning

Preservation Operations

Business

Technology
Goal: Efficiency and effectiveness

- Challenges
  - Effort for specifying criteria
  - Effort to perform evaluation

- We need effective and efficient near-optimal decisions

- Idea: Leverage scale
  - Standardize criteria
  - Develop quantitative metrics
  - Cross-reference and analyse
  - Provide visual insight

- Opportunity: Standardize, cross-reference, analyze
Collect

- 14 decision cases: Format conversion components
- 631 decision criteria (textually described)

Align

- Decision criteria models
- Software quality attributes (SQUARE: ISO 250x0)

Map

- Specify 368 uniquely identified criteria
- Categorise decision criteria of decision cases

Analyze

- Develop visual analysis tools
- Define and implement impact factors

Interpret

- Impact factors for criteria
- Impact factors for sets of criteria
Analyse criteria

Utility Functions

Scores

thresholds (linear) 1
thresholds (linear) 2
thresholds (linear) 3

0 1 2 3 4 5 6 7 8 9 10

48%
10%
29%

Pie chart

Bar chart
Analyze: Impact Factors

• Factors need to reflect
  – Usage frequency
  – Average weight
  – Sensitivity: How much do the utility scores change when the facts change?
    • Objective evidence
    • Utility function
  – Utility of 0 filters unacceptable candidates

• Using Goal-Question-Metric for selecting factors
Analyse a criterion (set) C

Goal

Understand key decision factors

Question

How often does C occur in scenario S?

How important is C?

How frequently does C exclude candidates?

Metric

Coverage

Maximum Impact

Range

Selectivity

% [0,5] %
Automation for controlled experiments

Goal: Identify automated tests to implement

Question:
- How important is criterion c in scenario S?
- How much does c vary?
- How difficult is automated measurement of c?

Graph showing data distribution across different values.
Interpret

- Sets of criteria ranked by coverage

### Criteria Sets Summary

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Coverage</th>
<th>Max Impact</th>
<th>Range</th>
<th>Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ Functional Correctness (outcome object)</td>
<td>346</td>
<td>100%</td>
<td>3.457</td>
<td>0.95</td>
<td>13.79%</td>
</tr>
<tr>
<td>SQ Functional Correctness - TIP</td>
<td>195</td>
<td>100%</td>
<td>3.457</td>
<td>0.786</td>
<td>10.34%</td>
</tr>
<tr>
<td>Business</td>
<td>18</td>
<td>85.71%</td>
<td>0.33</td>
<td>0.155</td>
<td>2%</td>
</tr>
<tr>
<td>SQ Functional Completeness</td>
<td>44</td>
<td>78.57%</td>
<td>1.022</td>
<td>0.133</td>
<td>4.08%</td>
</tr>
<tr>
<td>SQ Portability</td>
<td>5</td>
<td>78.57%</td>
<td>0.5</td>
<td>0.054</td>
<td>0%</td>
</tr>
<tr>
<td>Format</td>
<td>31</td>
<td>71.43%</td>
<td>0.822</td>
<td>0.198</td>
<td>9.3%</td>
</tr>
<tr>
<td>SQ Performance Efficiency</td>
<td>7</td>
<td>64.29%</td>
<td>0.5</td>
<td>0.103</td>
<td>2.7%</td>
</tr>
<tr>
<td>SQ Resource Utilization</td>
<td>3</td>
<td>64.29%</td>
<td>0.5</td>
<td>0.103</td>
<td>2.7%</td>
</tr>
<tr>
<td>SQ Usability</td>
<td>6</td>
<td>64.29%</td>
<td>0.16</td>
<td>0.035</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Functional Correctness - RIP</td>
<td>15</td>
<td>57.14%</td>
<td>0.156</td>
<td>0.03</td>
<td>2.7%</td>
</tr>
<tr>
<td>SQ Compatibility</td>
<td>5</td>
<td>57.14%</td>
<td>0.085</td>
<td>0.01</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Time Behaviour</td>
<td>3</td>
<td>57.14%</td>
<td>0.5</td>
<td>0.094</td>
<td>3.33%</td>
</tr>
<tr>
<td>SQ Reliability</td>
<td>8</td>
<td>42.86%</td>
<td>0.154</td>
<td>0.016</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Functional Correctness - Image Similarity</td>
<td>12</td>
<td>35.71%</td>
<td>0.401</td>
<td>0.056</td>
<td>8.33%</td>
</tr>
<tr>
<td>SQ Functional Correctness - IP</td>
<td>136</td>
<td>35.71%</td>
<td>0.625</td>
<td>0.135</td>
<td>12.5%</td>
</tr>
<tr>
<td>SQ Maintainability</td>
<td>3</td>
<td>28.57%</td>
<td>0.08</td>
<td>0.007</td>
<td>0%</td>
</tr>
<tr>
<td>Outcome Effects</td>
<td>4</td>
<td>21.43%</td>
<td>1.54</td>
<td>0.137</td>
<td>26.67%</td>
</tr>
<tr>
<td>SQ Capacity</td>
<td>1</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>
Interpret

- Sets of criteria ranked by selectivity

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Coverage</th>
<th>Max Impact</th>
<th>Range</th>
<th>Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Effects</td>
<td>4</td>
<td>21.43%</td>
<td>1.54</td>
<td>0.137</td>
<td>26.67%</td>
</tr>
<tr>
<td>SQ Functional Correctness (outcome object)</td>
<td>346</td>
<td>100%</td>
<td>3.457</td>
<td>0.95</td>
<td>13.79%</td>
</tr>
<tr>
<td>SQ Functional Correctness - IP</td>
<td>136</td>
<td>35.71%</td>
<td>0.625</td>
<td>0.135</td>
<td>12.5%</td>
</tr>
<tr>
<td>SQ Functional Correctness - TIP</td>
<td>195</td>
<td>100%</td>
<td>3.457</td>
<td>0.786</td>
<td>10.34%</td>
</tr>
<tr>
<td>Format</td>
<td>31</td>
<td>71.43%</td>
<td>0.822</td>
<td>0.198</td>
<td>9.3%</td>
</tr>
<tr>
<td>SQ Functional Correctness - Image Similarity</td>
<td>12</td>
<td>35.71%</td>
<td>0.401</td>
<td>0.056</td>
<td>9.33%</td>
</tr>
<tr>
<td>SQ Functional Completeness</td>
<td>44</td>
<td>78.57%</td>
<td>1.022</td>
<td>0.133</td>
<td>4.08%</td>
</tr>
<tr>
<td>SQ Time Behaviour</td>
<td>3</td>
<td>57.14%</td>
<td>0.5</td>
<td>0.094</td>
<td>3.33%</td>
</tr>
<tr>
<td>SQ Functional Correctness - RIP</td>
<td>15</td>
<td>57.14%</td>
<td>0.156</td>
<td>0.03</td>
<td>2.7%</td>
</tr>
<tr>
<td>SQ Performance Efficiency</td>
<td>7</td>
<td>64.29%</td>
<td>0.5</td>
<td>0.103</td>
<td>2.7%</td>
</tr>
<tr>
<td>SQ Resource Utilization</td>
<td>3</td>
<td>64.29%</td>
<td>0.5</td>
<td>0.103</td>
<td>2.7%</td>
</tr>
<tr>
<td>Business</td>
<td>18</td>
<td>85.71%</td>
<td>0.33</td>
<td>0.155</td>
<td>2%</td>
</tr>
<tr>
<td>SQ Capacity</td>
<td>1</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Compatibility</td>
<td>5</td>
<td>57.14%</td>
<td>0.085</td>
<td>0.01</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Maintainability</td>
<td>3</td>
<td>28.57%</td>
<td>0.08</td>
<td>0.007</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Portability</td>
<td>5</td>
<td>78.57%</td>
<td>0.5</td>
<td>0.054</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Reliability</td>
<td>8</td>
<td>42.86%</td>
<td>0.154</td>
<td>0.016</td>
<td>0%</td>
</tr>
<tr>
<td>SQ Usability</td>
<td>6</td>
<td>64.29%</td>
<td>0.16</td>
<td>0.035</td>
<td>0%</td>
</tr>
</tbody>
</table>
Maximum to average impact of criteria sets

- Outcome Effects: 0.383 to 1.540
- SQ Functional Correctness (outcome object): 0.504 to 1.406
- SQ Functional Correctness - IP: 0.183 to 0.625
- SQ Functional Correctness - TIP: 0.261 to 0.625
- SQ Performance Efficiency: 0.159 to 0.500
- SQ Portability: 0.112 to 0.500
- SQ Resource Utilization: 0.159 to 0.500
- SQ Time Behaviour: 0.159 to 0.500
- Format: 0.277 to 0.469
- SQ Functional Correctness - Image Similarity: 0.076 to 0.313
- Business: 0.138 to 0.250
- SQ Functional Completeness: 0.111 to 0.244
- SQ Functional Correctness - RIP: 0.134 to 0.231
- SQ Usability: 0.032 to 0.160
- SQ Compatibility: 0.017 to 0.085
- SQ Reliability: 0.040 to 0.008
- SQ Maintainability: 0.020 to 0.004
- SQ Capacity: 0.000 to 0.000
Interpret

• In the homogeneous set of 6 cases, 219 criteria:
  – Some widely used criteria have negligible impact and/or variation
  – Some “expensive” criteria are not selective, i.e. not necessarily critical
  – 31 of the 105 mapped criteria are dominated
  – None of the low-impact sets is fully dominated

• Some of the criteria that are expensive to measure are in fact insignificant and/or have zero variation

• Summary factors and rankings provide valuable insight
Conclusions

• Formalized semantics increase criterion transparency
  – increases quality and trust

• Set of metrics is helpful for assessing decision criteria in different ways
  – Decision maker: Understand impact of criterion on decisions
  – Researcher: Understand key decision factors in certain scenarios, organisation contexts, etc.
  – Developer: Understand critical success factors of candidates

• Drives improvement of DS environment
DSS improvements

- Filter and shortlist candidates
- Prioritize development of automation
- Rank criteria to measure and cut off data collection
- Approximate non-critical measures
- Remove non-critical criteria
- Recommend criteria based on selectivity and co-occurrence
- Automated suggestion of objectives and criteria based on organisational context
- ...
DPUE

• Today is the deadline for choosing topics
• Next week is the **hard** deadline for submitting a concept
  – This concept is part of the grading scheme
  – Send it upfront to your contact point if you want feedback!
• Make sure to get in touch with your contact point
  – Discussion time will be helpful in all cases
• Any questions? Get in touch!
Thank you for your attention!

Questions?

DPUE task 2 questions?

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