

Towards a preservation strategy evaluation workflow

Presentation for the ERPANET Workshop
By Carl Rauch
13th – 15 th of October 2004

Department for Software Technology & Interactive Systems

Motivation

Problem

- Several different preservation strategies
- Different collections with diverse preservation requirements
- Workflows defined only for the process of preservation

Requirements

- A workflow to compare strategies analytically
- A strategy to identify the different requirements
- Measures to be equally applicable to new preservation strategies

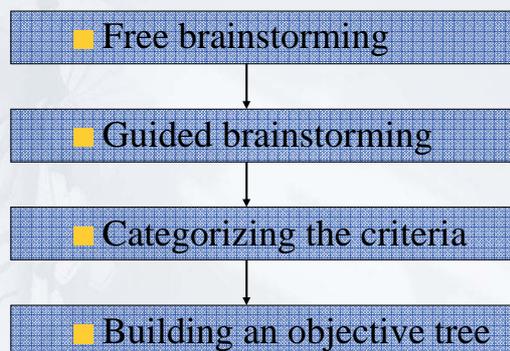
Solutions

- A workflow to find the preservation requirements
- A generic framework, which can be applied to different preservation requirements
- A decision support system to clearly rank preservation solutions

Outline

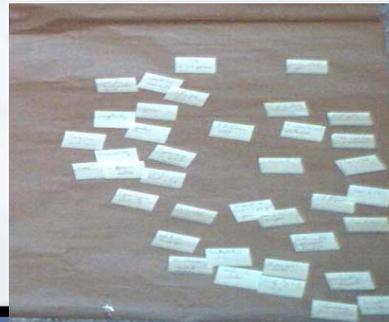
- Defining the preservation requirements
- Evaluating and comparing of preservation solutions

Defining preservation requirements



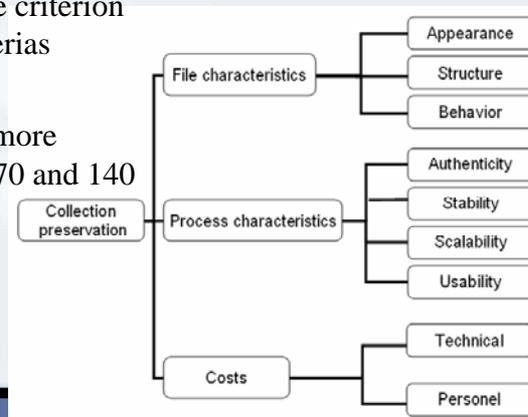
Free brainstorming

- First: Clearly define collection and included files
- Duration: ½ - ¾ hour
- Discussion of every criterion within the group
- Little input from administrator
- Usually around 20 criteria identified



Guided brainstorming

- Introducing the generic objective tree
- Discussion of every single criterion and adding of further criterias
- Duration: 1-3 hours
- Usually far more and far more detailed criteria (around 70 and 140 in two tests)



Categorizing the criteria

- Adjusting criteria to the related topic of the objective tree
- Eliminating duplicates
- Further discussion
- Duration: 1-2 hours



Building an objective tree

- Structuring all criteria, finding dependencies
- Eliminating duplicates and entering new criteria
- Duration: 1-2 hour

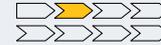


Top level	Level 2	Level 3	Level 4	Level 5	Level 6	Top level	Level 2	Level 3	Level 4	Level 5	Level 6
File Char.	Appearance	Quality	Audio resolution Sample rate Compression rate Water marking				Usability	Ingest	Complexity	Metadata ingest speed Need for technical assistance Time to understand the system	
		Functionality	Sound Replay (e.g. Stereo) Replay Velocity Content					Maintenance	Speed of ingest Format profiles Speed of data transfer to archive Complexity	Time to understand the system Need for technical assistance Amount of automation	
	Structure	Metadata		Marker Source History Time Code Information Technical Equipment Title Integrator Copyright				Reopening	Frequency of migration cycles Time to use	Reading speed of tape Reading speed of index Opening speed	Online Nearline Offline
			Kind of connection	Link to external metadata-database Metadata included in the file Metadata incorporated with the file Import of metadata from the original file Automatic metadata extraction Manual metadata insertion					Complexity	Preparation of often mentioned files Automation Tool support Technical assistance Multiple data access Partial decompression possible	
		Reqs. Storage	Storage size Uncompressed storage					Technical			
	Behavior										
Top level	Level 2	Level 3	Level 4	Level 5	Level 6	Top level	Level 2	Level 3	Level 4	Level 5	Level 6
Process Char.	Authenticity	Irreversibility of changes	Watermarking Storage safety Registration of changes Signatures			Cost	Technical	Initial	Software-Costs	HSM Software Back-up Software Standard software	
	Stability	File-format verification Storage	Multiple storage media Durability	Duration of generation Estimated duration Industry support Backwards compatibility					Hardware	Ingest	Device for write/read a tape Station for ingest Tape cartridge Storage sub systems Server Userstation
		Format	Hardware independence External control and access Portability Software	Tape support Storage structure				Running	Software-Costs	Reopening HSM Software Standard Software	
	Scalability	Autom. error recognition Storage capacity	Migration complexity Open format Spread of the format Robustness (Error Tolerance)						Hardware	Ingest Storage	Device for write/read a tape Station for ingest Tape cartridge Storage subsystems Server Userstation
		Format range scalability	Workload of computer Limit of scalability	Technical limit Commercial limit			Personnel	Enrolment Running costs		Reopening	
									Maintenance Reopen		

Outline

- Defining the preservation requirements
- Evaluating and comparing of preservation solutions

Assign effects to objectives



- **Measurable effects:** for example in mm, EURO per year, seconds for file ingestion,

- **Categorizing evaluation:** Valued with subjective impression, necessary, where no measureable evaluation found, for example paragraph formation or numbering of chapters. An extreme form is a simple yes/no decision.

Source: VUT

Definition of alternatives



- Migration & Standardisation →
 - Migrate documents to ADOBE PDF
 - Migrate documents to OpenOffice.org
 - Migrate documents to PostScript
 - Migrate documents to a newer version of the same software

- Emulation & Encapsulation →
 - Encapsulate digital objects

- Computer Museum →
 - Try to preserve the hardware environment

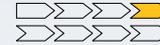
- Digital Tablet →
 - Try to construct a digital tablet

- No change to the strategy →
 - Do not adapt the strategy

- No preservation effort →
 - Do not take care of preservation

Source: VUT

Alternatives' evaluation



- Measure of the alternatives' performance, using either:
 - Original files
 - Files out of a test bed

	Newer MS Word version	OpenOffice.org Writer	PDF 5.0	No changes at all
Page borders	0 mm	+ 3 mm	0 mm	0 mm
Ingestion: sec. per file	10 sec	10 sec	15 sec	0 sec
Software costs per year	50 €	0 €	0 €	0 €
Numbering of chapters	3	N.A.	5	5
Paragraph Formation	3	2	5	5

Source: VUT

Transform measured values



- Define the transformation table:

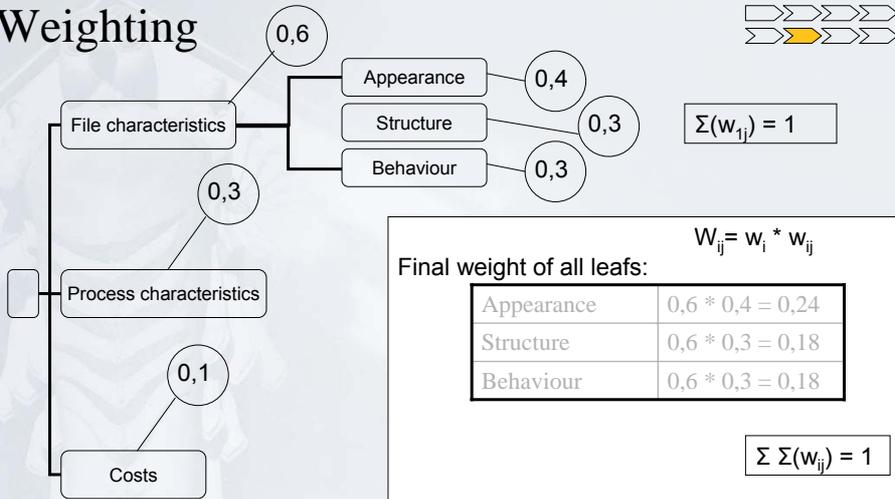
	5	4	3	2	1	N.A
Page borders	+/- 0 mm	+/- 1 mm	+/- 2 mm	+/- 3 mm	+/- 4 mm	> 4mm
Ingest: sec. per file	0 -5 sec	5-10 sec	10-15 sec	15-25 sec	25-40 sec	>.40 sec
Software costs per year	0 €	1-30 €	31-50 €	51-70 €	71-100 €	> 100 €
Numbering of chapters	1	2	3	4	5	N.A.
Paragraph formation	1	2	3	4	5	N.A.

- Transform the results according to make them comparable

	Newer MS Word version	OpenOffice.org Writer	PDF 5.0	No changes at all
Page borders	5	2	5	5
Ingestion: sec. per file	3	3	2	5
Software costs per year	3	5	5	5
Numbering of chapters	3	N.A.	5	5
Paragraph Formation	3	2	5	5

Source: VUT

Weighting



Source: VUT

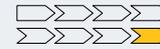
Aggregating part values

- **Part values per objective:**
 - Final Weights times the Transformed Values

- **Total value per alternative**
 - Sum of all part values of a strategy
 - Includes also „not acceptable“ alternatives

Source: VUT

Final Ranking



- Ranking of the alternatives according to their total values, not acceptable alternatives are ranked worst
- Final sensitivity analysis
- Concerning non measurable influences on the decision, such as good knowledge of an alternative, good relation to a supplier,

Source: VUT

Summary

The Utility Analysis workflow provides:

- A structured approach to identify objectives
- A process to categorize these goals
- A powerful methodology to compare various preservation solutions