Choosing an Optimal
Digital Preservation Strategy

Andreas Rauber
Department of Software Technology
and Interactive Systems
Vienna University of Technology
http://www.ifs.tuwien.ac.at/~andi

FACULTY OF !NFORMATICS

Motivation

- need for preservation action has been recognized
- preservation planning requires a series of decisions
  - what are my overall goals? what is my collection?
  - which strategy to follow? (e.g. migration)
  - which format to migrate to? (e.g. pdf)
  - which version of PDF? (e.g. 2.4, PDF/A,...)
  - which tool to use? under which OS?
  - which parameters?
  - what do I loose? acceptable loss? performance? costs?
- is my preservation plan (accountably) good?

Outline

- goals and objectives
- utility analysis
  - objectives definition
  - alternatives evaluation
  - result analysis
- digital preservation utility analysis tool
- case studies
- benefits and beneficiaries
- current activities and outlook

Goals and Objectives

- motivate and allow stakeholders to precisely specify their goals

- provide structured model to describe and document these

- create defined setting to evaluate approaches

- document outcome of evaluations to allow informed, accountable decision

- while being generic and simple, applicable for a wide range of institutions

Utility Analysis

- cost-benefit analysis model
- used in the infrastructure sector
- adapted for digital preservation needs
- 6 steps grouped into 3 phases

(1) objectives definition   (2) alternatives evaluation
(3) result analysis
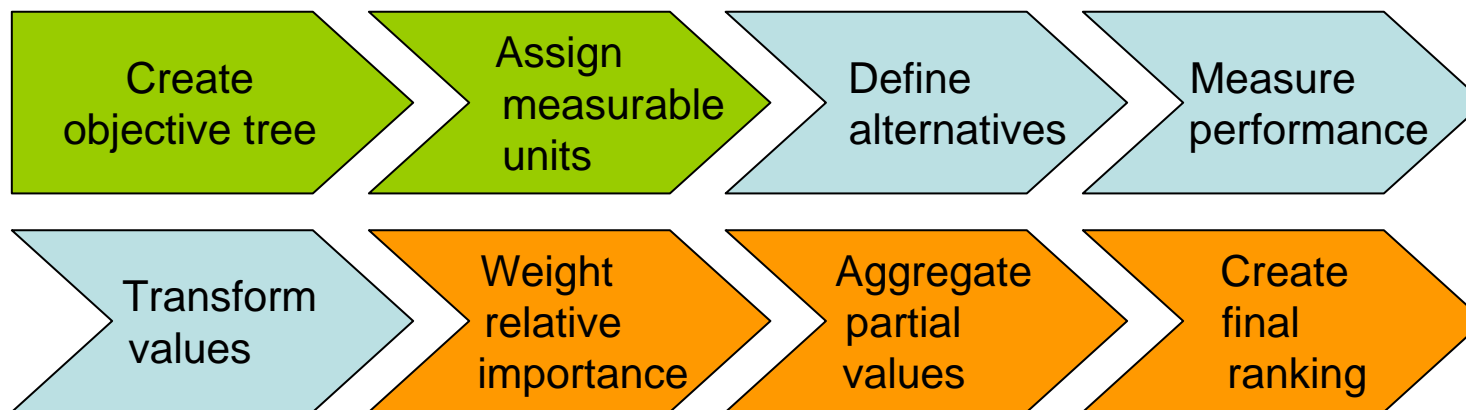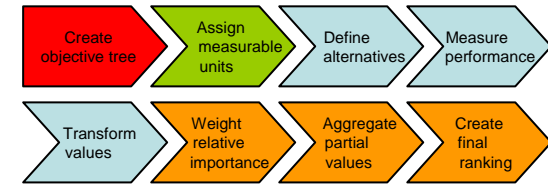
(1) create objective tree:

- list all requirements in tree structure
- start from high-level institutional goals
- break down to fine-granular, specific criteria
- usually 4 top-level branches:
  - object characteristics *(color depth, macros, ...)*
  - record characteristics *(context, relations, ...)*
  - process characteristics *(scalability, error detection, ...)*
 - costs *(set-up, per object, HW/SW, personnel, ...)*
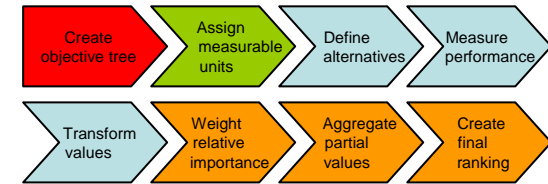
(1) create objective tree (cont.):

- objective tree with several hundred leaves

- usually created in workshops, brainstorming sessions

- re-using branches from similar institutions, collection holdings, ...

- best-practice models, templates

- clear definition of the goals of a preservation endeavor

Objectives Definition
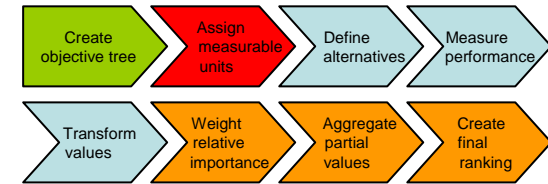
(2) assign measurable units:

- ensure that leaf criteria are objectively (and automatically) measurable
  - seconds/pounds per object
  - bits color depth
  - ...
- subjective scales where necessary
  - diffusion of file format
  - amount of (expected) support
  - ...

# Alternatives Evaluation
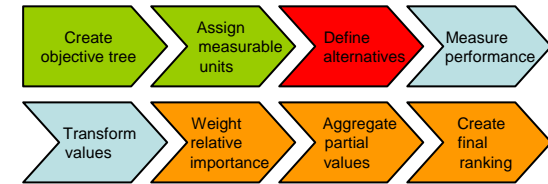
(3) define alternatives:

- list and formally describe the preservation action possibilities to be evaluated
  - tool, version
  - operating system
  - parameters

Alternatives Evaluation
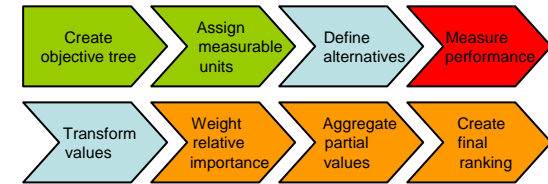
(4) measure alternatives' performance:

- select typical objects from
  - a test-bed repository
  - from own collection
- subject them to the different alternatives identified
- measure performance with respect to leaf criteria in the objective tree
- document the results

Alternatives Evaluation
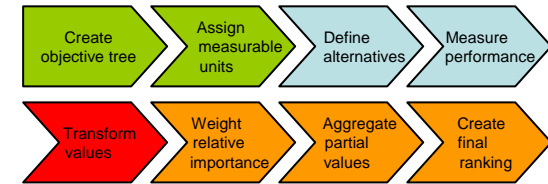
(5) transform measured values:

- measures come in seconds, pounds, 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"
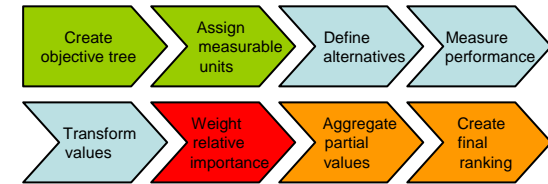
(6) weight relative importance:

- set importance factors
- not all leaf criteria are equally important
- set relative importance of all siblings in a branch
- weights are propagated down the tree to the leaves
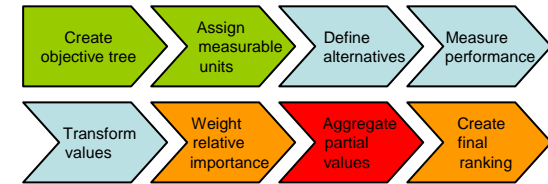
(7) aggregate partial values:

- 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

- creates performance values for each alternative on each of the sub-criteria identified

- measures conformance of each solution with the goals specified
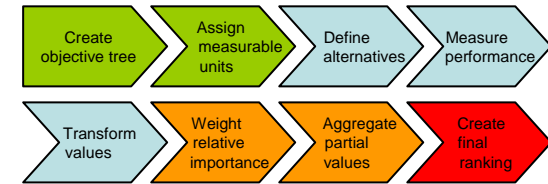
(8) create final ranking:

- rank alternatives according to overall utility value at root
- performance of each alternative
  - overall
  - for each sub-criterion (branch)
- allows performance measurement of combinations of strategies
- final sensitivity analysis against minor fluctuations in
  - measured values
  - importance factors

Outline

- goals and objectives
- utility analysis
  - objectives definition
  - alternatives evaluation
  - result analysis
- **digital preservation utility analysis tool**
- **case studies**
- **benefits and beneficiaries**
- **current activities and outlook**

Digital Pres. Utility Analysis Tool
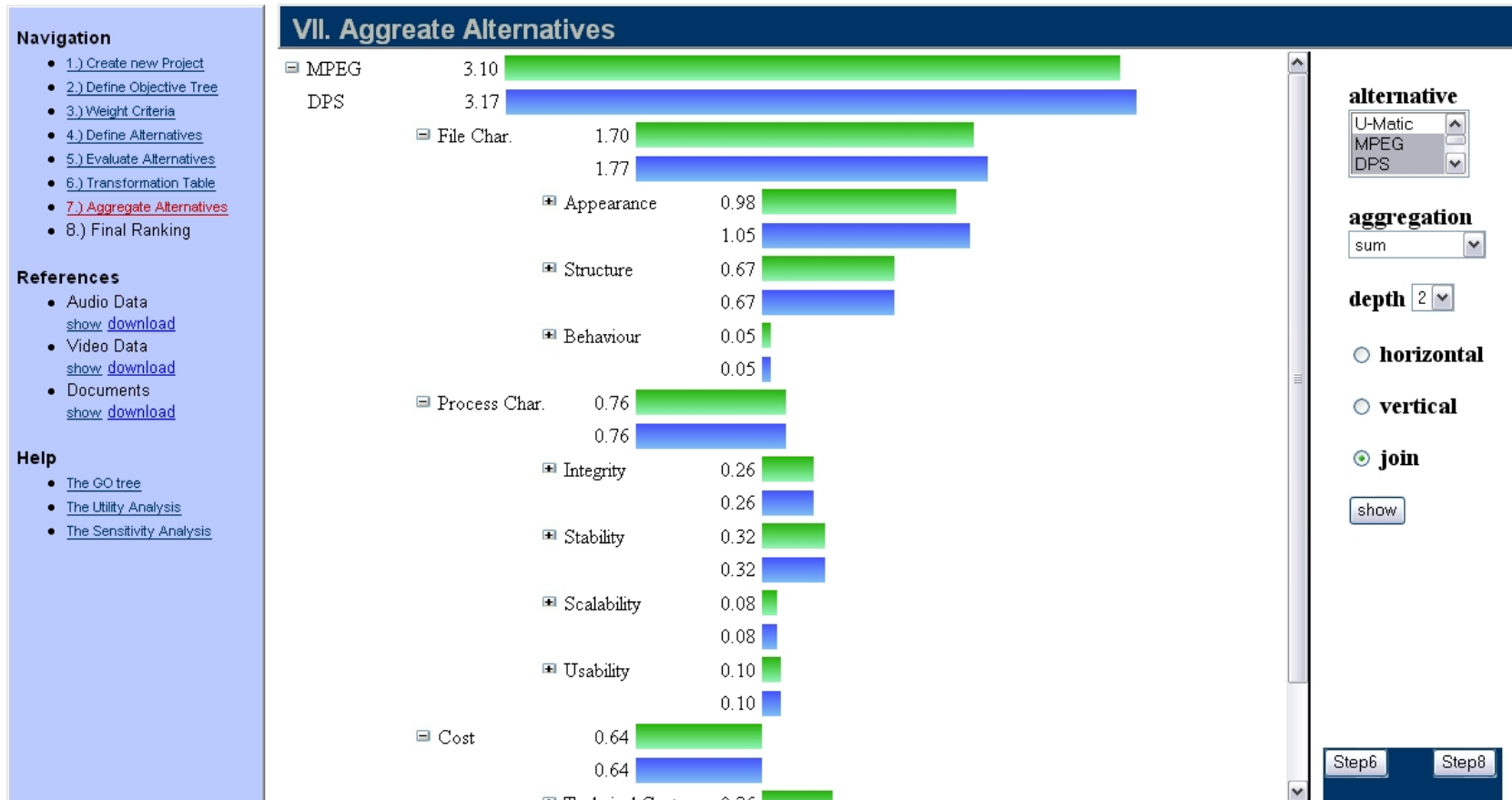
Case Studies

- TUWIEN not a DP-institution
- performed series of case studies with different institutions
  - video records of the *Austrian Phonogram Archives*
  - audio records of the *Austrian Phonogram Archives*
  - document records of the *Dutch National Archives*
  - thesis publications with the *Austrian National Library* (in progress)
- presented at a range of international forums and training sessions for discussion

Benefits

- a simple, methodologically sound model to specify and document DP requirements

- structured setting for evaluating performance of alternative preservation actions

- documented evaluation for informed and accountable decisions

- set of templates to assist institutions

- generic workflow that can easily be integrated in different institutional settings

- provides means for vendor tool assessment

Beneficiaries

- **cultural heritage institutions**
  - libraries, archives, museums
- **suitable for institutions of different sizes**
  - small, specialized to large, heterogeneous collections
- **suitable for institutions with different levels of expertise**
  - support for special digital preservation department
  - decision support and assistance for small institutions based on best-practice model and guided workflow
- **eventually for tool providers to test their tools**

Current Work and Outlook

- continue work on case studies to create sound templates for a range of settings
  ➡ "Generic Objective Tree"

- evolve the tool set to provide comprehensive decision support based on template plans

- automate experiment evaluation stage by incorporating
  - file format information (e.g. PRONOM)
  - object characterization tools (e.g. JHOVE)

Thanks to

- the team at TUWIEN
  - Carl Rauch
  - Stephan Strodl
  - Christoph Bartenstein
- members of the DELOS Preservation Cluster
- case study participants
- Digital Preservation Award Judging Panel