



Preservation Planning

Andreas Rauber
Institut für Softwaretechnik und Interaktive
Systeme
TU Wien

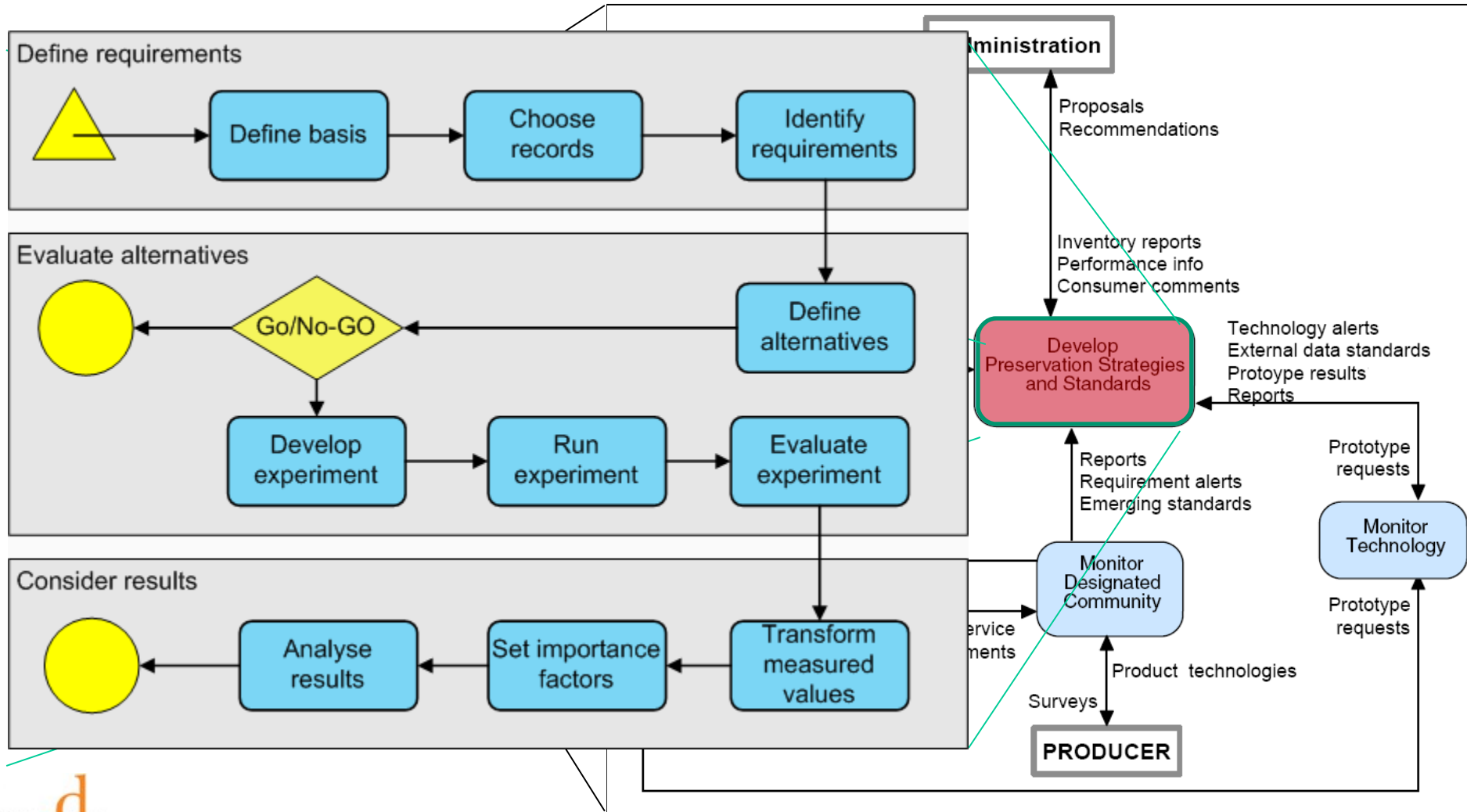
<http://www.ifs.tuwien.ac.at/dp>



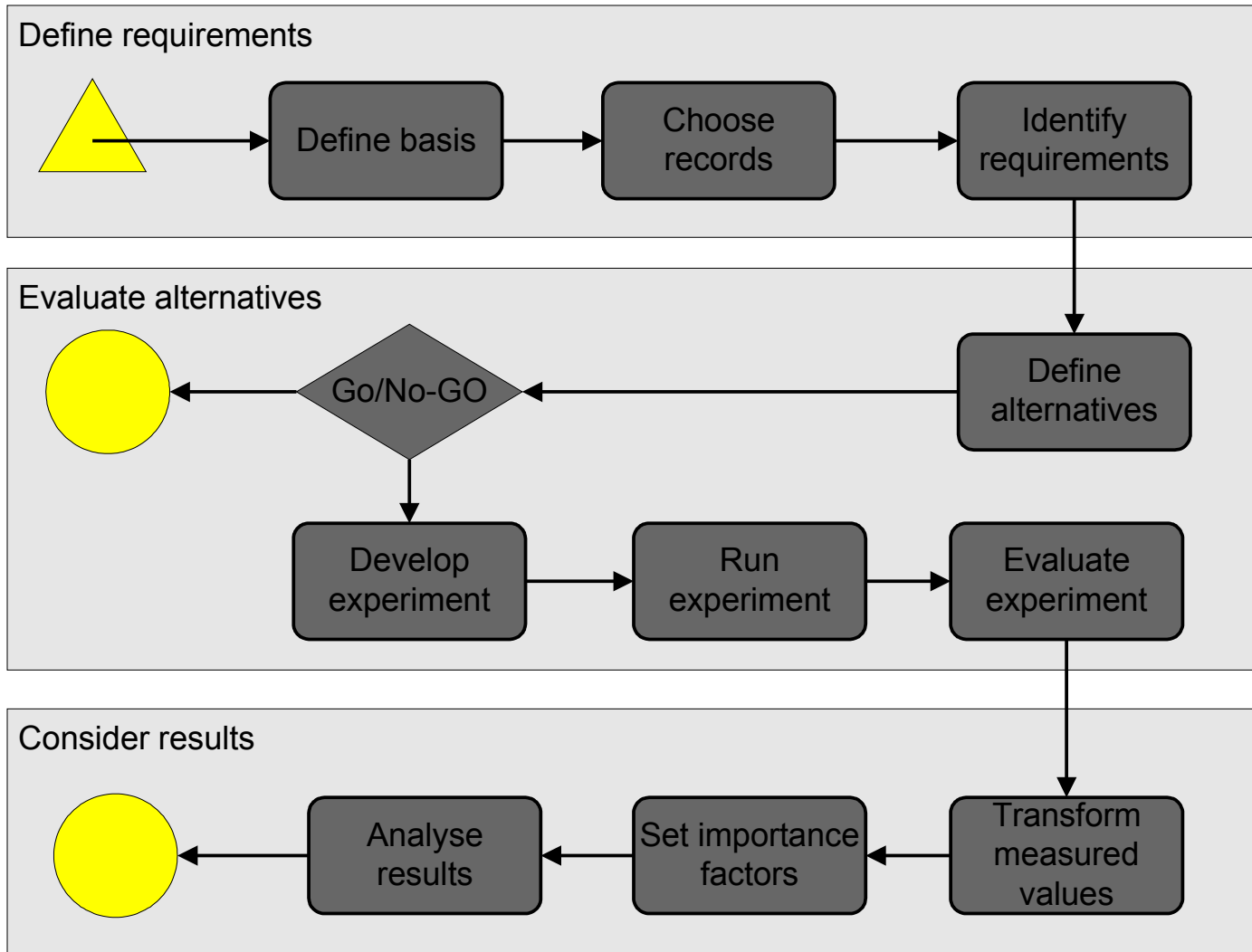
FACULTY OF **INFORMATICS**

- Several preservation strategies developed
- How do you know what is most suitable?
 - Right choice depends on the needs (no clear preferences)
- How to measure and evaluate the results of each preservation strategy?
- What are the requirements?
- How to define a controlled and trusted environment and a procedure for applying or testing preservation strategies?

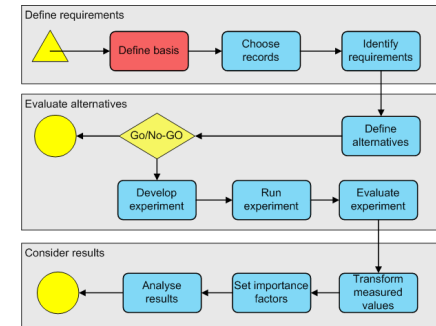
Preservation Planning



PP Workflow

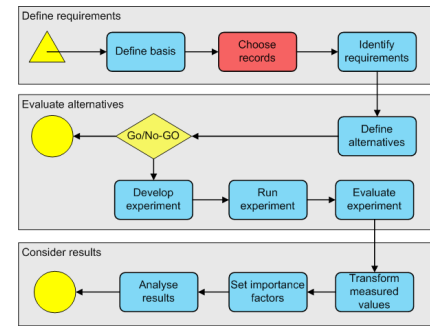


Define basis



- What are the objects?
- What are the essential characteristics?
 - Content, context, structure, form and behaviour
- What are the requirements?
 - Authenticity, reliability, integrity, useability
 - Metadata (for different purposes)
- What preservation strategies will be applied and evaluated?

Choose objects/records



➤ Different object types

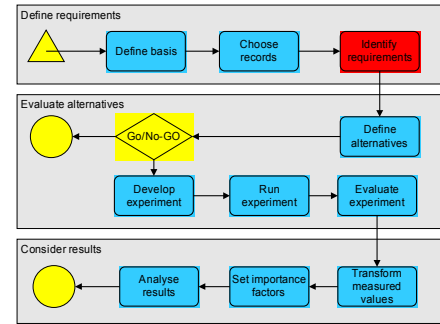
- Text documents, audio, video, e-mail, multimedia, databases, data sets, ...

➤ Distinction between

- Physical (technical) object = computer file, and
- The intellectual object (e.g. what is shown on the screen)

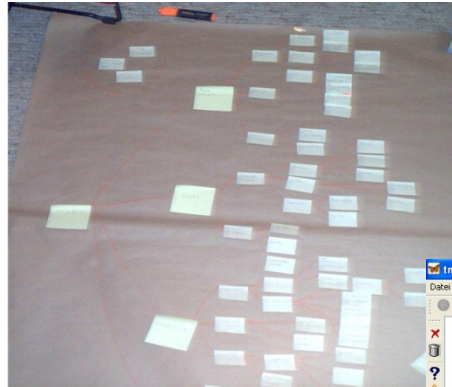
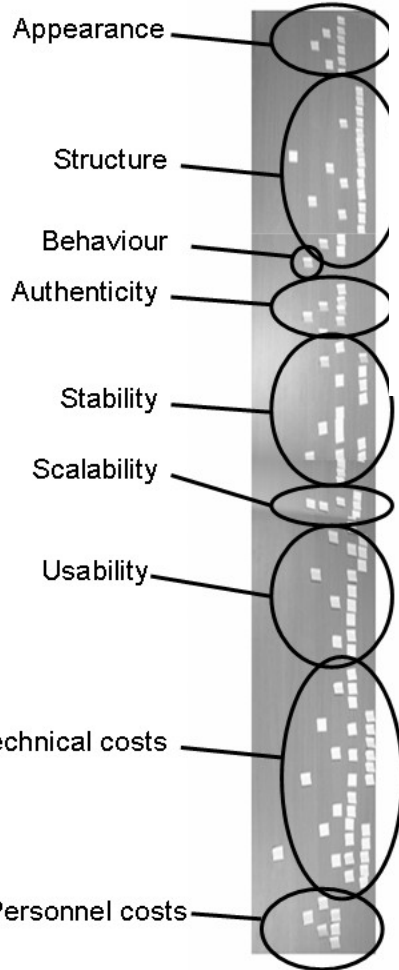
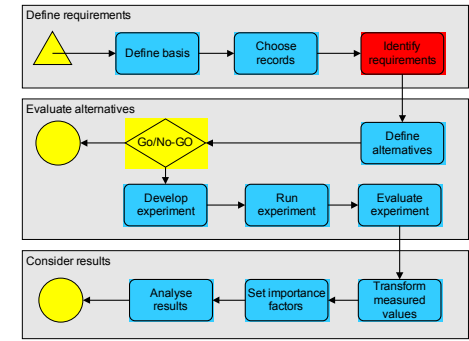
➤ Choice of objects affects the evaluation

Identify requirements



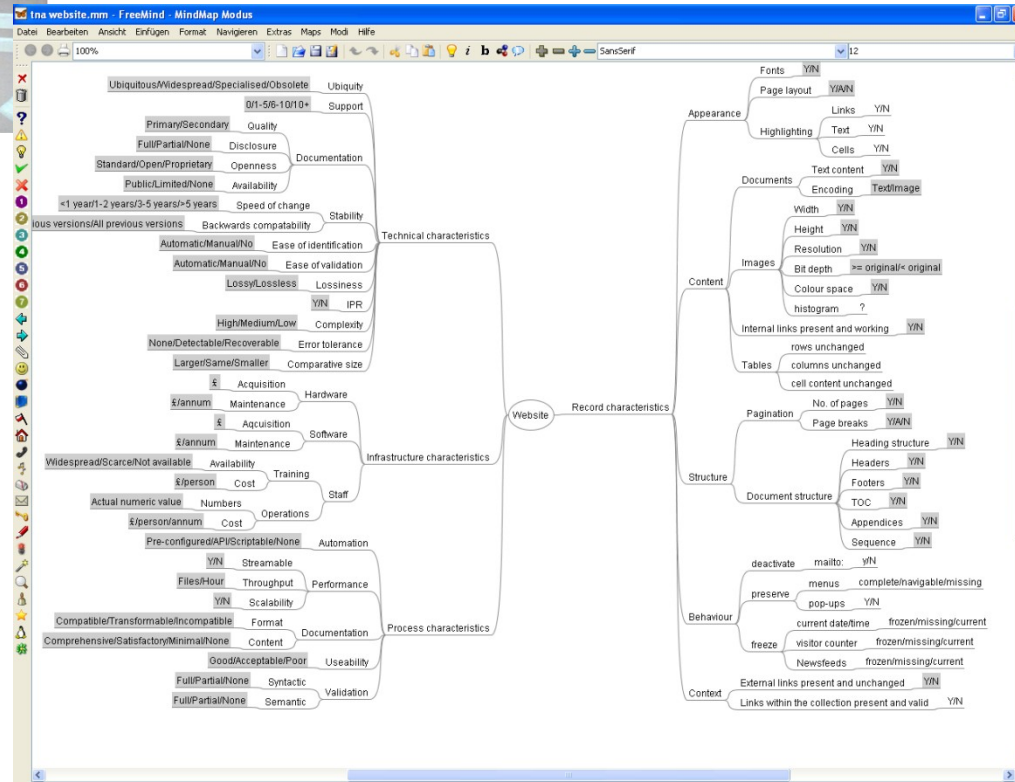
- Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain
- Usually four major groups:
 - object characteristics (content, metadata ...)
 - record characteristics (context, relations, ...)
 - process characteristics (scalability, error detection, ...)
 - costs (set-up, per object, HW/SW, personnel, ...)
- Put the objects in relation to each other (hierarchical)
- Objective tree approaches:
 - bottom-up
 - top-down

Identify requirements

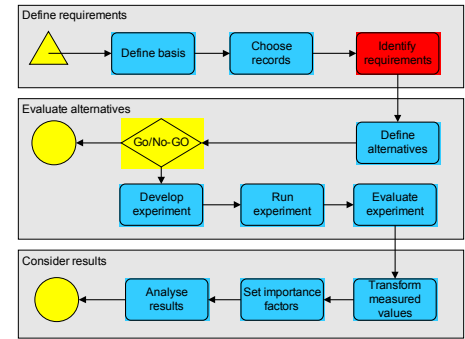


Analog...

... or born-digital



Identify requirements



e.g. Color-proof, Frame rate,, ..

e.g. Original compression, ..

e.g. Subtitles, ...

e.g. File format verification

e.g. Durability

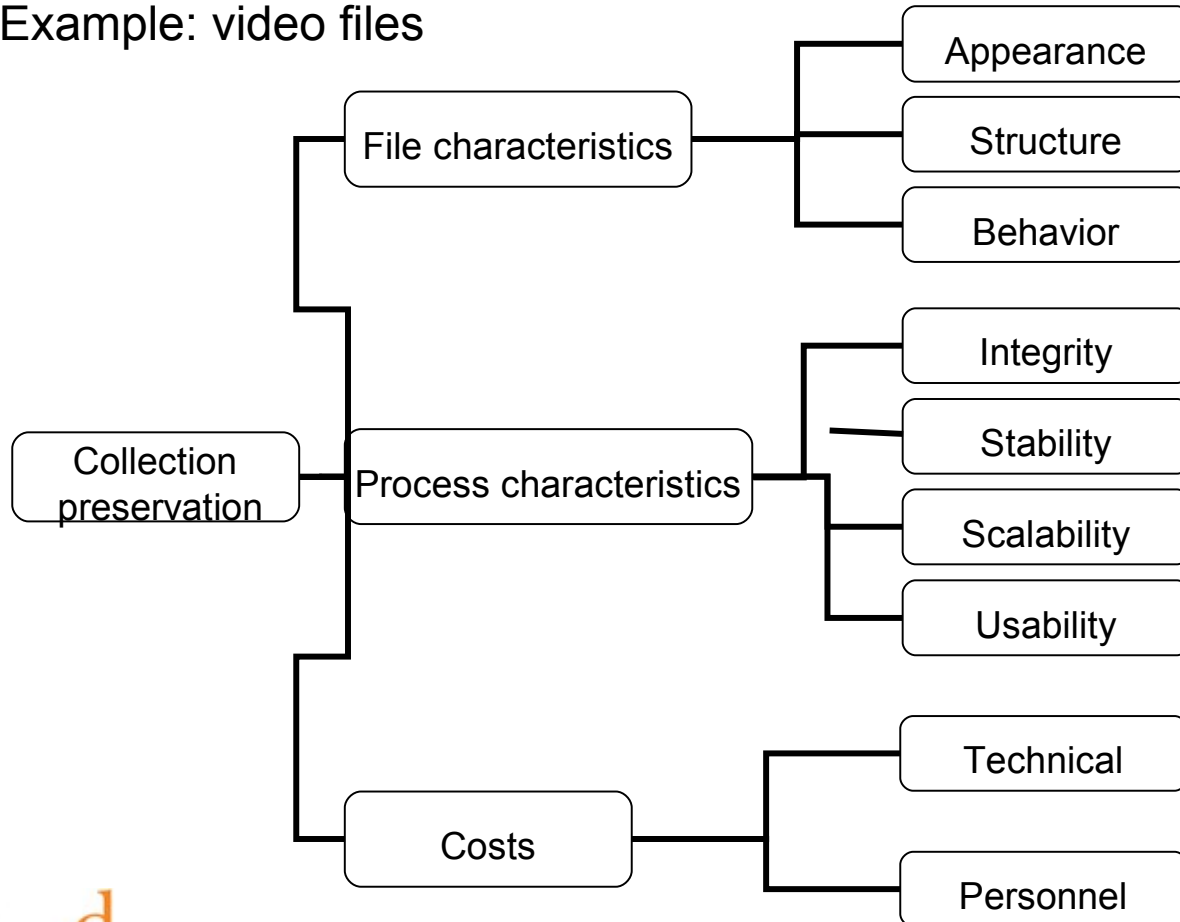
e.g. Format scalability

e.g. Complexity, Functionality

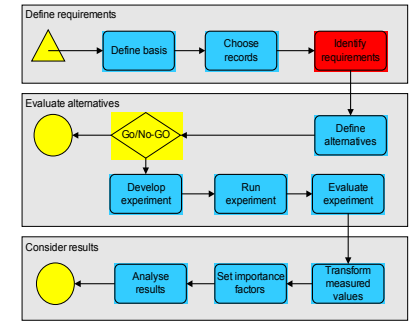
e.g. Hardware, Software

e.g. Enrolment, Maintenance

Example: video files



Assign measurable units



- Assign measurable effect to each leaf
 - Ensure that leaf criteria are objectively (and automatically) measurable
 - seconds/Euro per object
 - Bits of color depth
 - ...
 - Subjective scales where necessary
 - diffusion of file format
 - amount of (expected) support
 - ...
- No limitations on the use of scale

Identify requirements



PLANETS Preservation Planning Tool (PLATO)

PlanningTool > Home

Home
Load Project
New Project

Define Requirements

Define Basis
Define Sample Records
Identify Requirements

Evaluate Requirements

Define Alternatives
Go/No-Go
Develop Experiment
Run Experiment
Evaluate Experiment

Consider Results

Transform Measured Values
Set Importance Factors
Analyse Results
Sum
Multiplication
Sum of Priority
Austin Slight

Select All | Select None | Expand All | Collapse All
X Website > **Technical characteristics**

SelectFocus	Node	Scale	Restriction	Unit
<input type="checkbox"/>	▼ Technical characteristics			
Select	Previous			
<input type="checkbox"/> X	▼ Ubiquity	Ordinal	Ubiquitous/Widespread/Specialised/Obs	
Select				
<input type="checkbox"/> X	▼ Support	Ordinal	0/1-5/6-10/10+	
Select				
<input type="checkbox"/> X	▼ Documentation			
Select	▼ Quality	Ordinal	Primary/Secondary	
<input type="checkbox"/> X	▼ Disclosure	Ordinal	Full/Partial/None	
Select	▼ Openness	Ordinal	Standard/Open/Proprietary	
<input type="checkbox"/> X	▼ Availability	Ordinal	Public/Limited/None	
Select				
<input type="checkbox"/> X	▼ Stability			
Select	▼ Speed of change	Ordinal	<1 year/1-2 years/3-5 years/>5 years	
<input type="checkbox"/> X	▼ Backwards compatability	Ordinal	None/Previous version only/Some previo	
Select				
<input type="checkbox"/> X	▼ Ease of identification	Ordinal	Automatic/Manual/No	
Select	▼ Ease of validation	Ordinal	Automatic/Manual/No	
<input type="checkbox"/> X	▼ Lossiness	Ordinal	Lossy/Lossless	
Select				
<input type="checkbox"/> X	▼ IPR	Boolean		
Select				

Next 9 - 11 of 11

Add a node to current selection

Add a leaf to current selection

Remove selection

Identify requirements



PLANETS Preservation Planning Tool (PLATO)

PlanningTool > Home

Home

- Load Project
- New Project

Define Requirements

- Define Basis
- Define Sample Records
- Identify Requirements

Evaluate Requirements

- Define Alternatives
- Go/No-Go
- Develop Experiment
- Run Experiment
- Evaluate Experiment

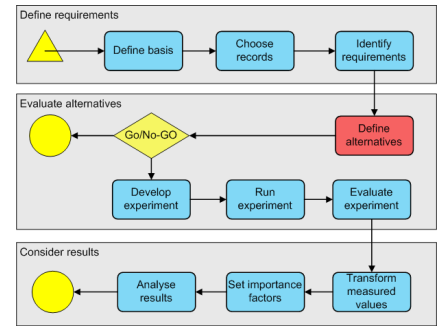
Consider Results

- Transform Measured Values
- Set Importance Factors
- Analyse Results
- Sum
- Multiplication
- Sum of Priority
- Austin Slight

Select All | Select None | Expand All | Collapse All
X Website

SelectFocus	Node	Scale	Restriction	Unit
<input type="checkbox"/>	Website			
<input type="checkbox"/>	Record characteristics			
<input type="checkbox"/>	Appearance			
<input type="checkbox"/>	Fonts	Boolean		
<input type="checkbox"/>	Page layout	Ordinal	Y/A/N	
<input type="checkbox"/>	Highlighting			
<input type="checkbox"/>	Links	Boolean		
<input type="checkbox"/>	Text	Boolean		
<input type="checkbox"/>	Cells	Boolean		
<input type="checkbox"/>	Content			
<input type="checkbox"/>	Documents			
<input type="checkbox"/>	Text content identical	Boolean		
<input type="checkbox"/>	Encoding	Ordinal	Text/Image	
<input type="checkbox"/>	Images			
<input type="checkbox"/>	Width	Integer		pixel
<input type="checkbox"/>	Height	Integer		pixel
<input type="checkbox"/>	Resolution	Integer		dpi
<input type="checkbox"/>	Bit depth	Ordinal	>= original/< original	
<input type="checkbox"/>	Colour space identical	Boolean		

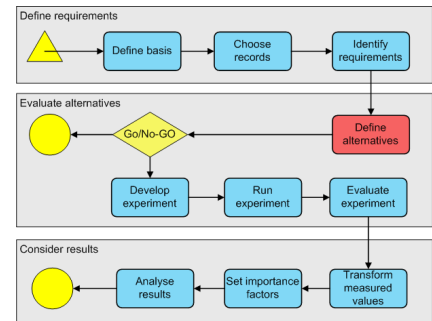
Define alternatives



- Given the type of objects and requirements, what strategies would be best suitable/are possible?
 - Migration
 - Emulation
 - Both
 - Other?

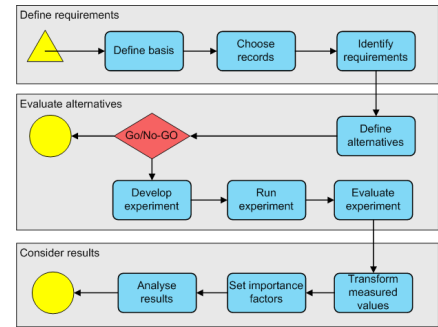
- For each alternative precise definition of
 - Which tool (OS, version,...)
 - Which functions of the tool in which order
 - Which parameters

Specify resources



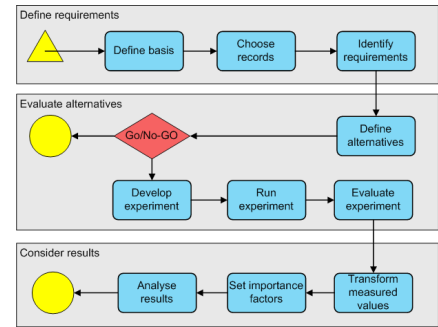
- Detailed design and overview of the resources for each alternative
 - human resources (qualification, roles, responsibility, ...)
 - technical requirements (hardware and software components)
 - time (time to set-up, run experiment,...)
 - cost (costs of the experiments,...)

Go/No-Go



- Deliberate step for taking a decision whether it will be useful and cost-effective to continue the procedure, given
 - The resources to be spent (people, money)
 - The availability of tools and solutions,
 - The expected result(s).
- Review of the experiment/ evaluation process design so far
 - Is the design complete, correct and optimal?
- Need to document the decision
- If insufficient: can it be redressed or not?

Go/No-Go



PLANETS Preservation Planning Tool (PLATO)

PlanningTool > Home

- Home
- Load Project
- New Project

Define Requirements

- Define Basis
- Define Sample Records
- Identify Requirements

Evaluate Requirements

- Define Alternatives
- Go/No-Go
- Develop Experiment
- Run Experiment
- Evaluate Experiment

Consider Results

- Transform Measured Values
- Set Importance Factors
- Analyse Results
 - Sum
 - Multiplication
 - Sum of Priority
 - Austin Slight

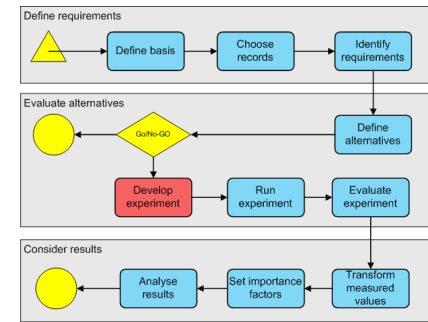
List All Available Alternatives:

Decision

Reason for Decision

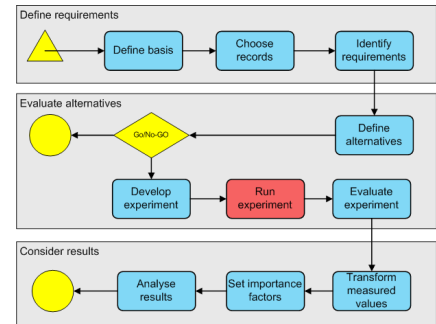
Action Needed

Develop experiment



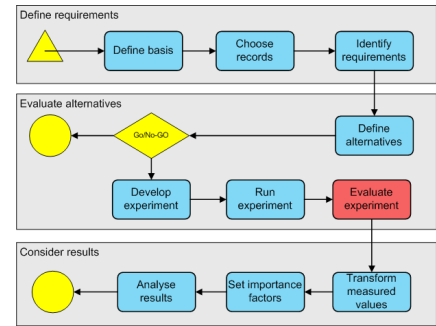
- Formulate for each evaluation or experiment or preservation process detailed
 - Development plan
 - steps to build and test software components
 - procedures and preparation
 - parameter settings for integrating preservation services
 - Test plan (mechanisms how to)
 - Evaluation/experiment plan (workflow/sequence of activities)

Run experiment



- Before conducting an evaluation or running an experiment, the experiment process as designed has to be tested
 - It may lead to re-design or even termination of the evaluation/ experiment process
- The results will be evaluated in the next stage
- The whole process needs to be documented

Evaluate experiment



- Evaluate the outcome of each alternative for each leaf of the objective tree
- The evaluation will identify
 - Need for repeating the process
 - Unexpected (or undesired) results
- Includes both technical and intellectual aspects
- Evaluation may include comparing the results of more than one experiment/evaluation.

DELOS DP TESTBED

Navigation

- [Create new Project](#)
- [1.\) Define Basis](#)
- [2.\) Choose Record](#)
- [3.\) Define Objective Tree](#)
- [Assign measurable units](#)
- [4.\) Define Alternatives](#)
- [5.\) Specify Ressources](#)
- [6.\) Go/No-Go](#)
- [7.\) Develop Experiment](#)
- [8.\) Run Experiment](#)
- [9.\) Evaluate Alternatives](#)
- [10.\) Transformation Table](#)
- [11.\) Weight Criteria](#)
- [12.\) Aggregate Alternatives](#)
- [13.\) Final Ranking/Sensitivity Analysis](#)

Output

- [XML Output](#)

Contact

- BugReport [Stephan Strodl](#)
- Website [DELOS DP Testbed](#)
- Last Update: 20.1.2006

IX. Evaluate Alternatives

- Signal representation [evaluate](#) ✓
- (e.g. RGB, ..)
- Interlace pictures [evaluate](#) ✓
- Progressive pictures [evaluate](#) ✓
- Picture-audio synchronisation [evaluate](#) ✓
- [-] Structure
 - [-] Metadata
 - [-] File history
 - Original compression [evaluate](#) ✓
 - Conversion steps [evaluate](#) ✓
 - Modifications [evaluate](#) ✓
 - File history
 - Storage
 - Temperature [evaluate](#) ✓
 - Organisation [evaluate](#) ✓
 - Humidity [evaluate](#) ✓
 - Storage media [evaluate](#) ✓

Alternative U-Matic:

Alternative MPEG:

Alternative DPS:

Alternative Std. DV:

Alternative SVHS:

Alternative NTSC-VHS:

Alternative Hi8:

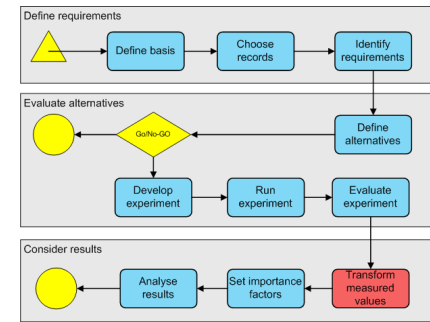
Alternative Beta Cam:

Alternative Digi-Beta:

Alternative PAL-VHS:

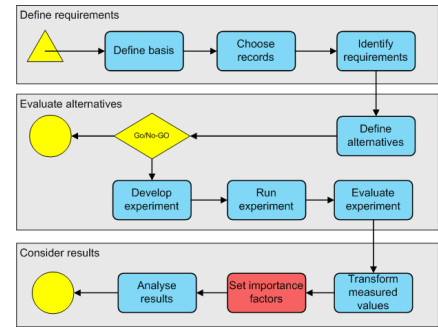
OK

Transform measured values



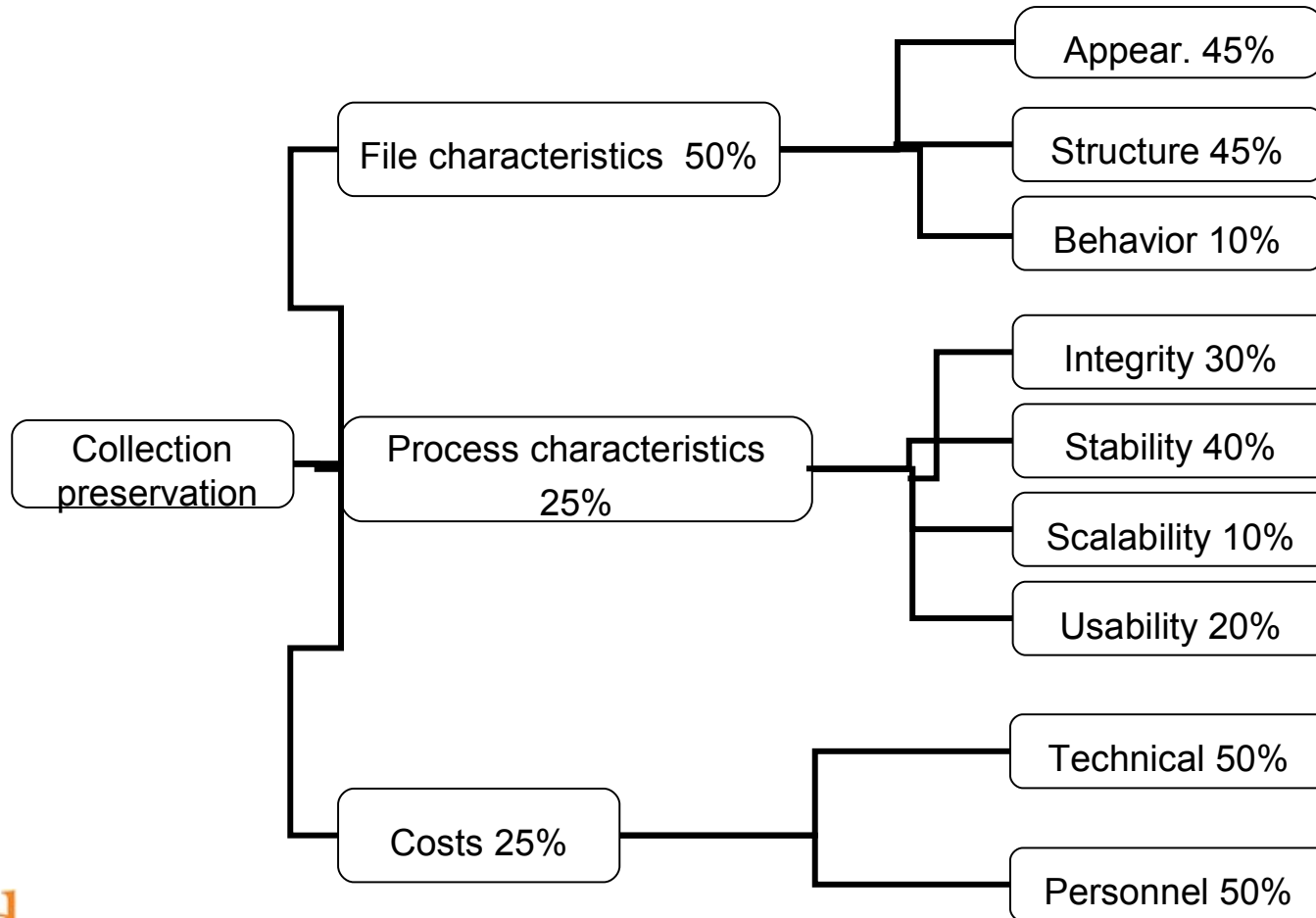
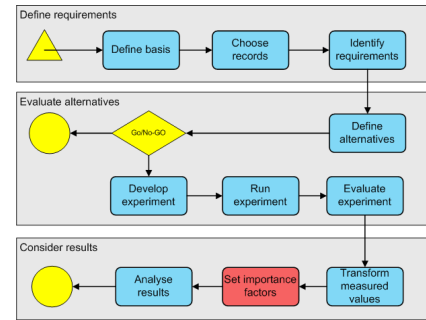
- Measures come in seconds, euro, bits, goodness values,...
- Need to make them comparable
- Transform measured values to uniform scale
- Transformation tables for each leaf criterion
- Linear transformation, logarithmic, special scale
- Scale 1-5 plus "not-acceptable"

Set importance factors



- Definition which criteria are more important
- Depends on individual preferences and requirements
- Adaptation for each implementation
- High influence on the final ranking
- Aggregation of weights

Set importance factors



DELOS DP TESTBED

Navigation

- ◆ [Create new Project](#)
- ◆ [1.\) Define Basis](#)
- ◆ [2.\) Choose Record](#)
- ◆ [3.\) Define Objective Tree](#)
- ◆ [Assign measurable units](#)
- ◆ [4.\) Define Alternatives](#)
- ◆ [5.\) Specify Ressources](#)
- ◆ [6.\) Go/No-Go](#)
- ◆ [7.\) Develop Experiment](#)
- ◆ [8.\) Run Experiment](#)
- ◆ [9.\) Evaluate Alternatives](#)
- ◆ [10.\) Transformation Table](#)
- ◆ [11.\) Weight Criteria](#)
- ◆ [12.\) Aggregate Alternatives](#)
- ◆ 13.) Final Ranking/Sensitivity Analysis

Output

- ◆ [XML Output](#)

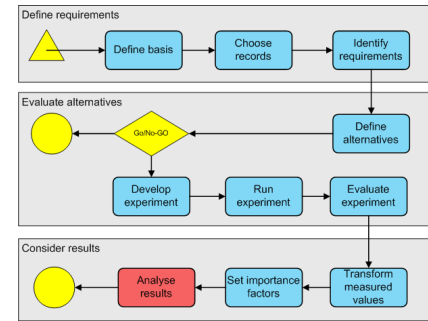
Contact

- ◆ BugReport [Stephan Strodl](#)
- ◆ Website [DELOS DP Testbed](#)
- ◆ Last Update: 20.1.2006

XI. Weight Criteria

[-] Root	1.0		
[-] File Char.	0.5	/0.0	weight ✓
[-] Appearance	0.45	/0.0	weight ✓
[-] Quality	0.7	/0.0	weight ✓
[-] Functionality	0.1	/0.0	weight ✓
Picture-audio synchronisation	0.200	/0.200	weight ✓
[-] Structure	0.45	/0.0	weight ✓
[-] Metadata	0.4	/0.0	weight ✓
[-] Structure Storage	0.35	/0.0	weight ✓
[-] Location	0.05	/0.0	weight ✓
[-] Formating	0.2	/0.0	weight ✓
[-] Behaviour	0.1	/0.0	weight ✓
[-] Process Char.	0.25	/0.0	weight ✓
[-] Cost	0.25	/0.0	weight ✓

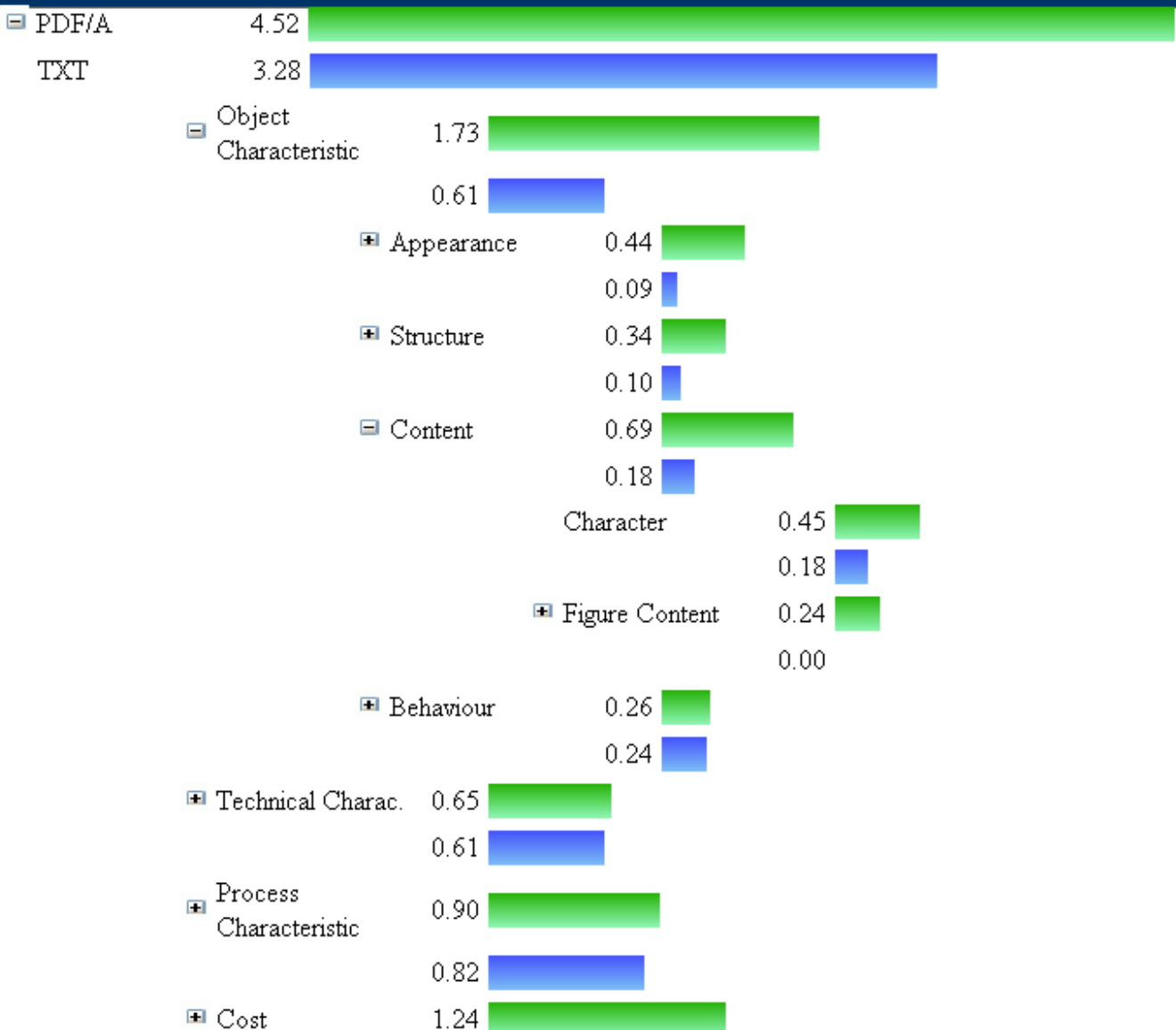
Analyse results



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

XII. Aggregate Alternatives



alternative

- PDF/A
- RTF ConvertDoc
- TXT

aggregation

sum

depth *

- horizontal
- vertical
- join

show

Step11

Step13

DELOS DP TESTBED

VIII. Sensitivity Analysis

Rang	Alternative	Gesamtnutzwert
Position 1:	PDF/A	4.52
Position 2:	TIFF	4.29
Position 3:	EPS	4.24
Position 4:	JPEG2000	4.17
Position 5:	RTF_Adobe	3.42
Position 6:	RTF ConvertDoc	3.38
Position 7:	TXT	3.28

Variance

Aggregation

Method

Navigation

- ◆ Create new Project
- ◆ 1.) Define Basis
- ◆ 2.) Choose Record
- ◆ 3.) Define Objective Tree
[Assign measurable units](#)
- ◆ 4.) Define Alternatives
- ◆ 5.) Specify Ressources
- ◆ 6.) Go/No-Go
- ◆ 7.) Develop Experiment
- ◆ 8.) Run Experiment
- ◆ 9.) Evaluate Alternatives
- ◆ 10.) Transformation Table
- ◆ 11.) Weight Criteria
- ◆ 12.) Aggregate Alternatives
- ◆ 13.) Final Ranking/Sensitivity
[Analysis](#)

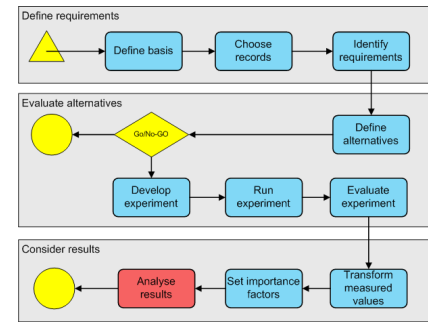
Output

- ◆ [XML Output](#)

Contact

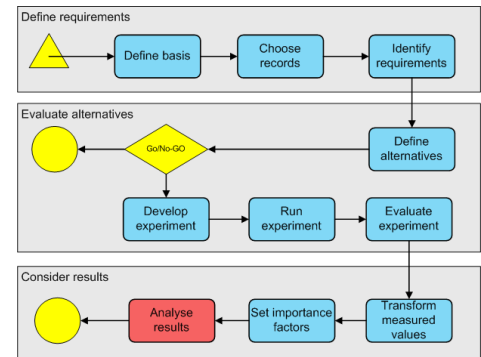
- ◆ BugReport
[Stephan Strodl](#)
- ◆ Website
[DELOS DP Testbed](#)
- ◆ Last Update: 20.1.2006

Analyse results



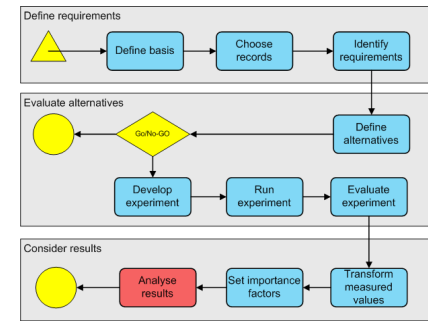
- Single performance value for each alternative to rank the alternatives
- Single performance values for each alternative for each sub-set of criteria to identify the best combination of alternatives
- Sensitivity Analysis: Analysis of the influence of small changes in the weight on the final value
- Basis for making Informed, well-documented, repeatable, accountable decisions

Analyse results



- 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

Consider results



- The review of the results may help to refine
 - The evaluation process/procedure
 - The preservation planning environment itself
 - The evaluation metrics
 - Understanding of the essential characteristics of the objects,
 - and identify further evaluations, experiments
- The review should take into account all previous work done in the preservation planning environment
- The review should look at both the technical and intellectual aspects of digital objects

- PLANETS is working on tool support
 - Integration of distributed preservation services
 - Preservation action services operate on objects (migration, emulation,...)
 - Preservation characterisation services describe objects
 - Web service infrastructure
 - Improved automation and usability
 - Integration of emerging technologies

- Templates/fragment libraries for objective trees

Conclusions

- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation for informed and accountable decisions
- Set of templates to assist institutions
- Generic workflow that can easily be integrated in different institutional settings

<http://www.ifs.tuwien.ac.at/dp>