

# Preservation Planning in PLANETS

London  
June 25, 2007

Christoph Becker  
Vienna University of Technology

# Agenda

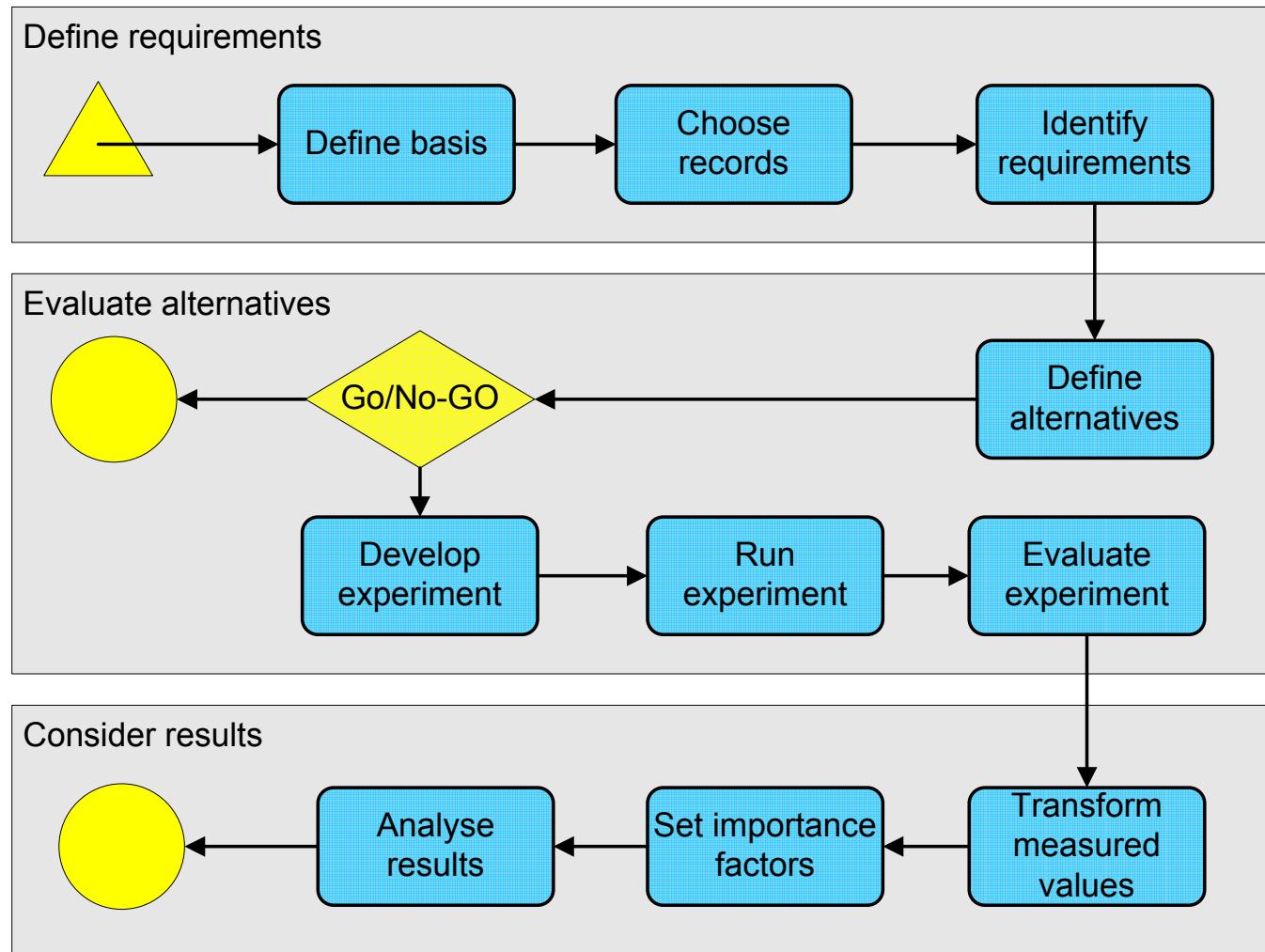
- Why Preservation Planning?
- Preservation Planning in the PLANETS project
- Decision Support for Preservation Planning
  - Workflow
  - Case Studies

- Variety of solutions and tools exist
  - Each strategy has its unique strengths and weaknesses
  - Requirements vary across settings
- 
- Decision on which solution to adopt is very complex
  - Documentation and accountability is essential
- 
- Preservation planning assists in decision making
    - Identify requirements
    - Evaluate options
    - Build strategies

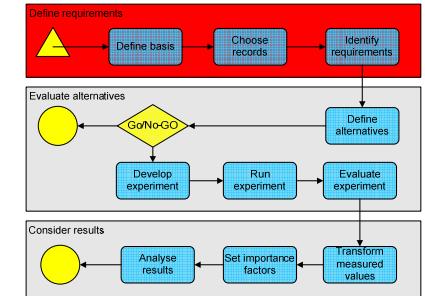
- Organisational policy and strategy models
  - Usage models
  - Collection profiling
  - Proactive planning
- 
- Decision support for preservation planning
  - Tools to support and automate preservation planning

- Support evaluation of preservation strategies
  - Identify requirements for different settings
  - Execute experiments and evaluate outcome against well-defined requirements and needs
  - Well-documented requirements and evaluation
- 
- Based on DELOS Testbed
    - Dutch Testbed
    - Utility Analysis
    - Structured model for preservation requirements

# Process Overview



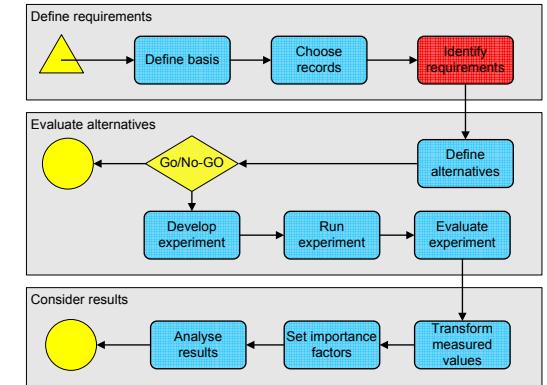
# Phase 1: Define requirements



- 1. Define basis**
  - Collection
  - Institutional settings
- 2. Choose sample objects/records**
  - Representative for the objects in the collection
- 3. Define requirements**
  - Identify requirements
  - Create objective tree
  - Assign measurable units

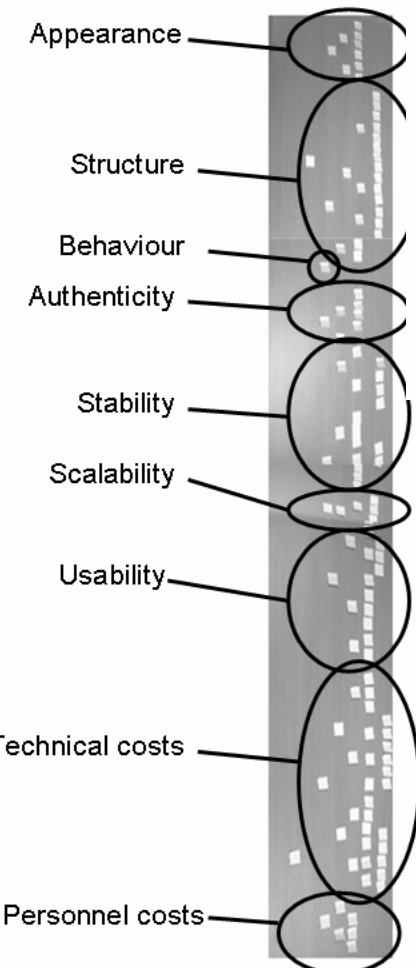
# Identify Requirements

- Identify requirements and goals
- Tree structure
- Start from high-level goals and break down to specific criteria
  
- Usually 4 top-level branches:
  - object characteristics (*content, metadata...*)
  - record characteristics (*context, relations...*)
  - process characteristics (*scalability, error detection, ...*)
  - costs (*set-up, per object, HW/SW, personnel, ...*)

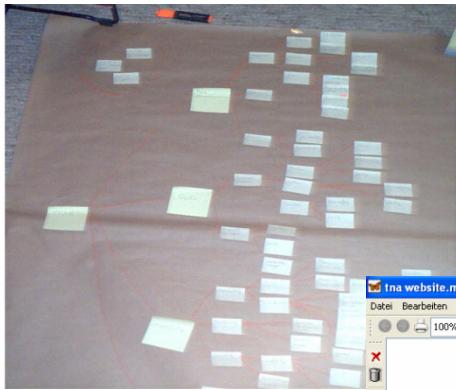




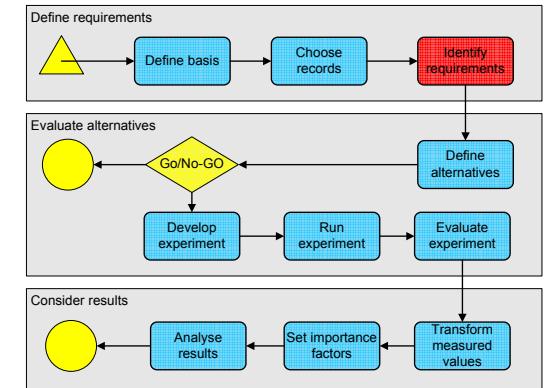
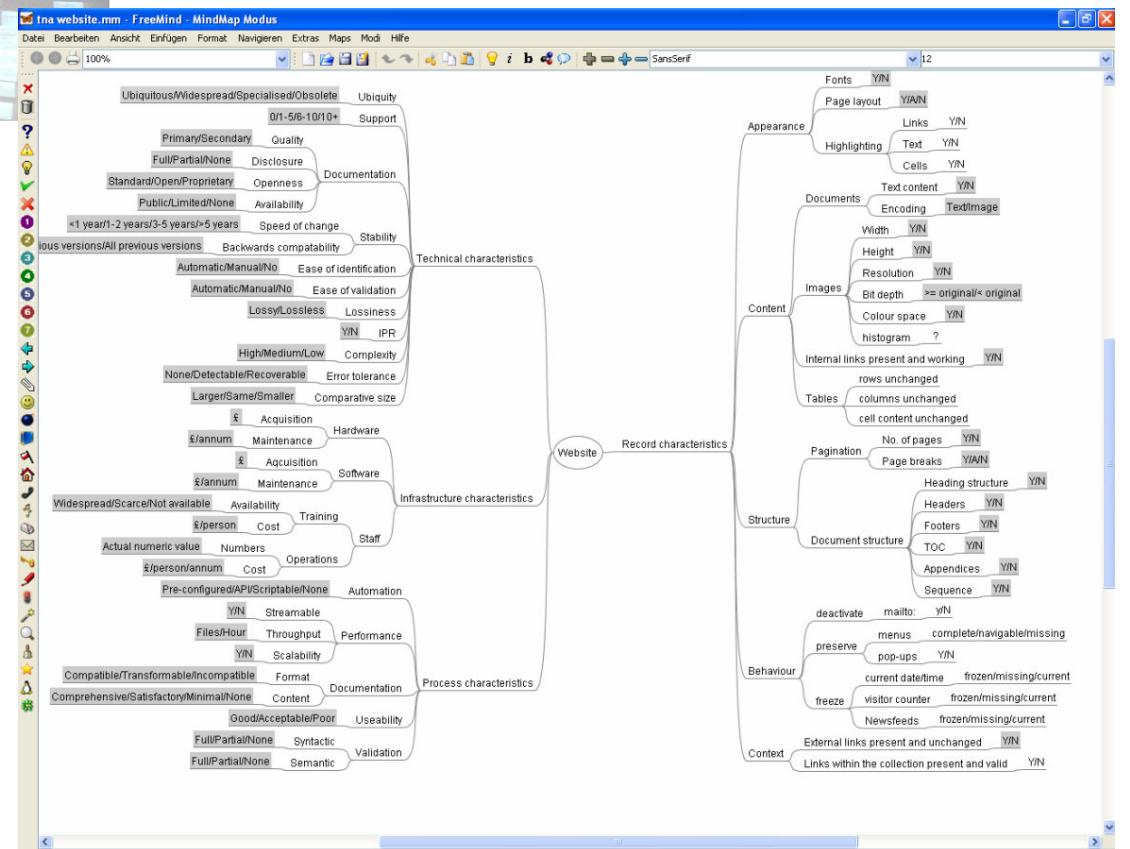
# Identify Requirements



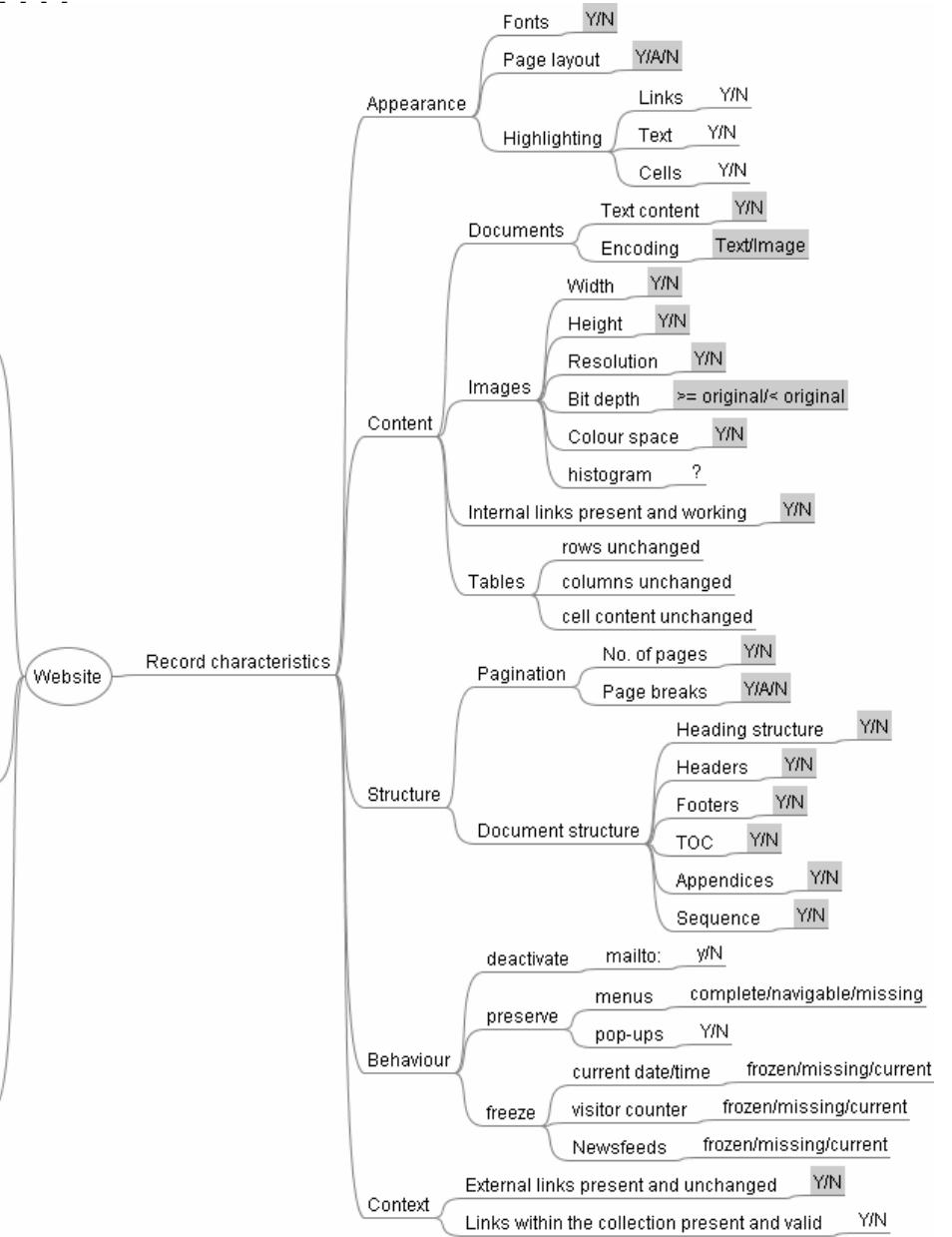
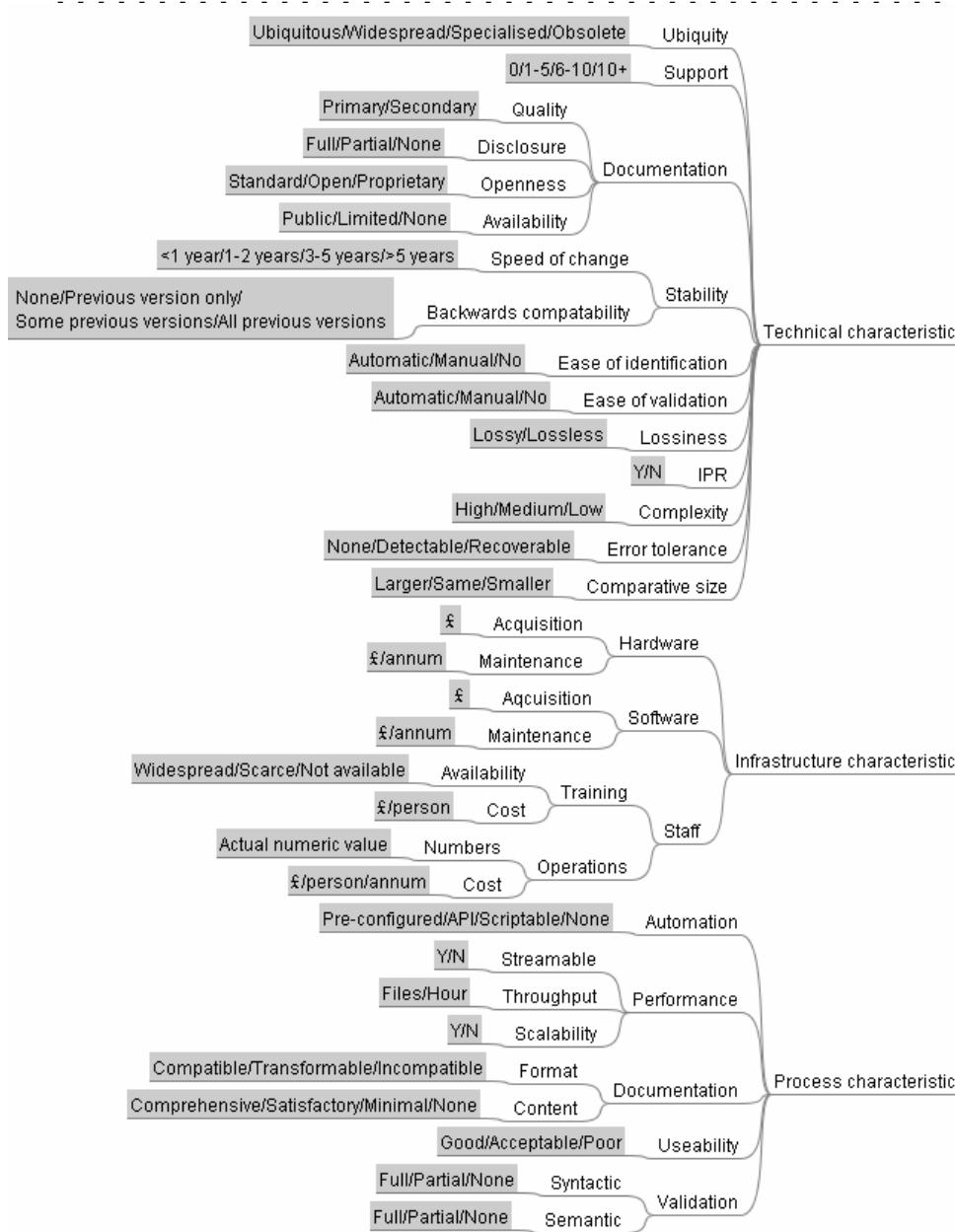
... or  
born-  
digital



## Analog...

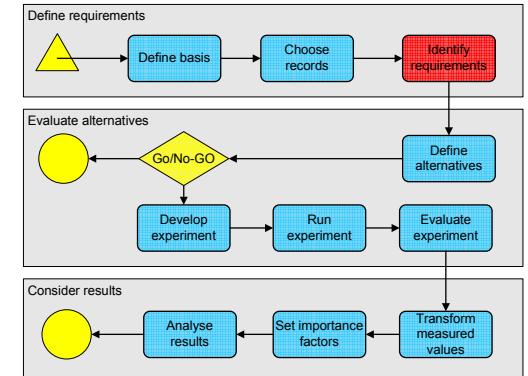


# Public Website (TNA)

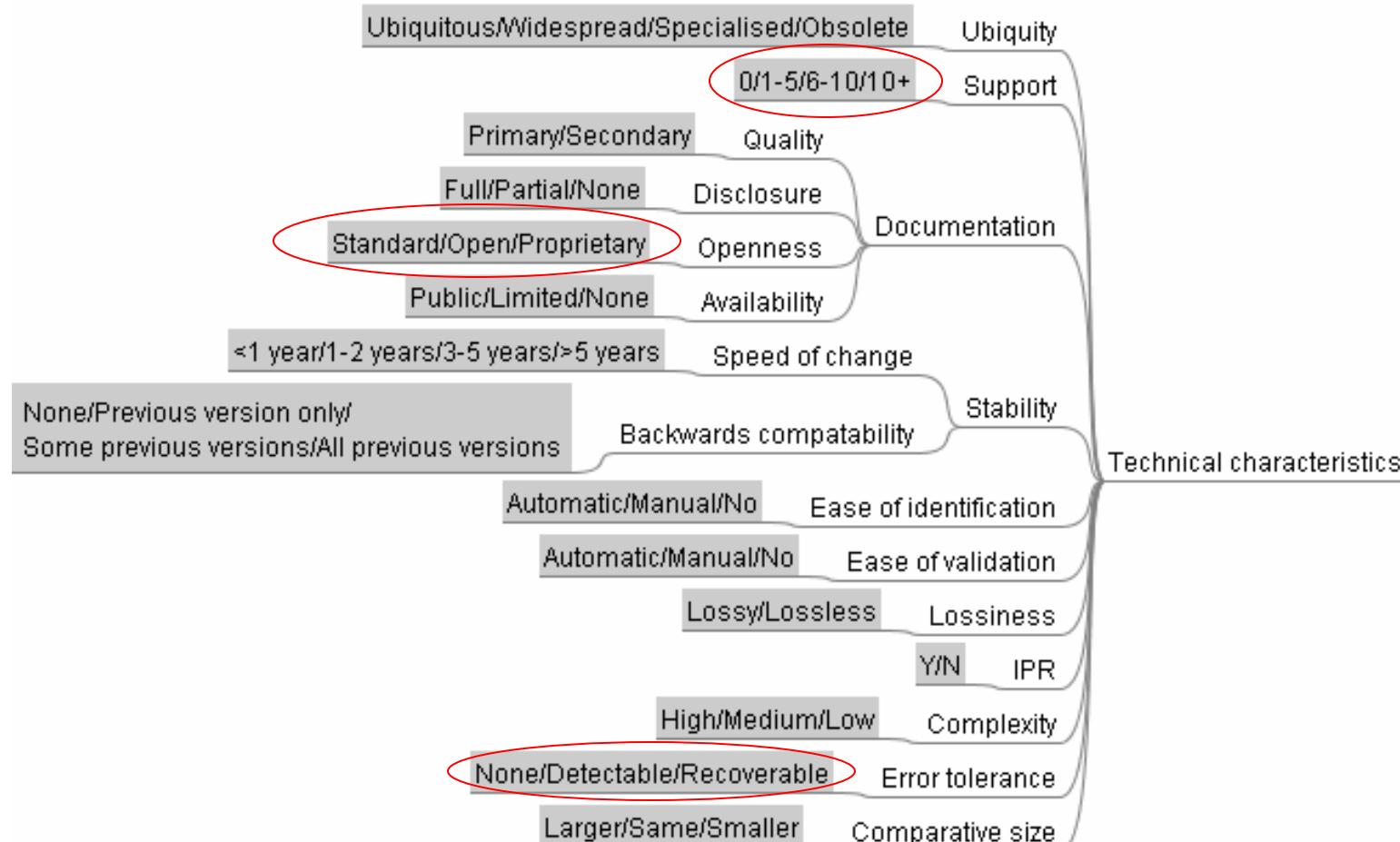


# Assign Measurable Units

- Leaf criteria should be objectively measurable
    - Seconds per object
    - Euro per object
    - Bits of colour depth
  - Subjective scales where necessary
    - Adoption of file format
    - Amount of (expected) support
- Quantitative results



# File Format Characteristics



# Tool Support

PLANETS Preservation Planning Tool (*PLATO*)

PlanningTool > Home

Logout [planets]

Expand All | Collapse All  
X Website

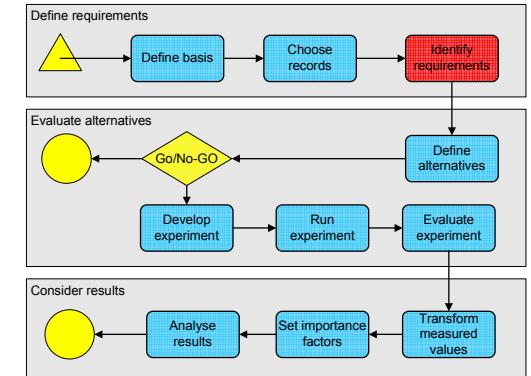
Focus	Node	Scale	Restriction	Unit
	Website			
X	Record characteristics			
X	▶ Appearance			
X	▶ Content			
X	▶ Structure			
X	▶ Behaviour			
X	▶ Context			
	Technical characteristics			
X	Previous			
X	▶ Ubiquity	Ordinal	Ubiquitous/Widespread/Specialised/Obsolete	
X	▶ Support	Ordinal	0/1-5/6-10/10+	number of tools
X	▶ Documentation			
X	▶ Stability			
X	▶ Ease of identification			
X	▶ Ease of validation			
X	▶ Lossiness			
X	▶ IPR	Boolean		
	Next 9 - 11 of 11			
X	▶ Infrastructure characteristics			
X	▶ Process characteristics			
X	▶ Automation	Ordinal	Automatic/Manual/No	
X	▶ Performance			
X	▶ Streamable	Boolean		
X	▶ Throughput	Float		files per hour
X	▶ Scalability	Boolean		
	▶ Documentation			
X	▶ Format	Ordinal	Compatible/Transformable/Incompatible	
X	▶ Content	Ordinal	Comprehensive/Satisfactory/Minimal/None	
X	▶ Usability	Ordinal	Good/Acceptable/Poor	
	▶ Validation			
X	▶ Syntactic	Ordinal	Full/Partial/None	
X	▶ Semantic	Ordinal	Full/Partial/None	

© 2007 Institute of Software Technology and Interactive Systems: «office bears»

Quick Access: [Help]

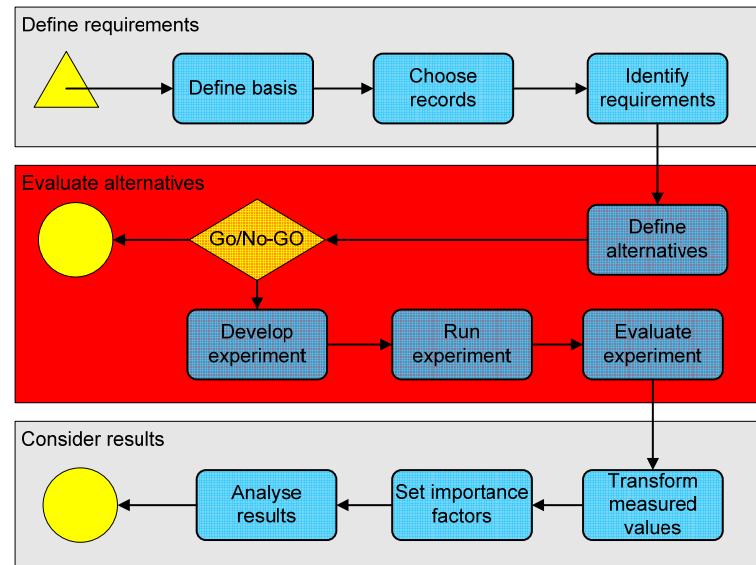
# Summary Phase 1

- **Defined basis**
  - Collection
  - Institutional context
- **Sample records chosen**
- **Defined requirements**
  - Objective tree
  - Measurable criteria
  - Basis for evaluation

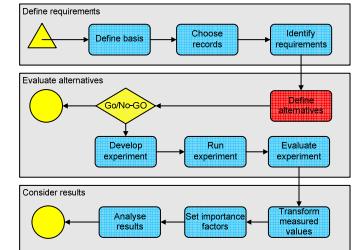


## Phase 2: Evaluate Alternatives

5. Define Alternatives
6. Go/No-Go decision
7. Develop experiment
8. Run experiment
9. Evaluate experiment

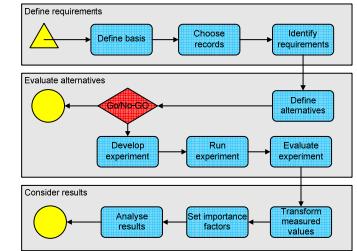


# Define Alternatives



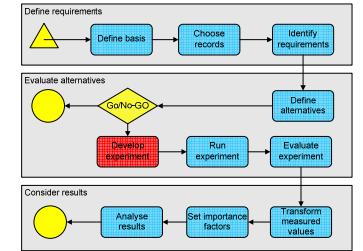
- Given the type of object and requirements, what strategies would be best suitable/are possible?
  - Migration
  - Emulation
  - Other?
- For each alternative precise definition of
  - Which tool (OS, version)
  - Which functions of the tool
  - Which parameters
  - Resources that are needed (human, technical, time and cost)

# Go/No-Go



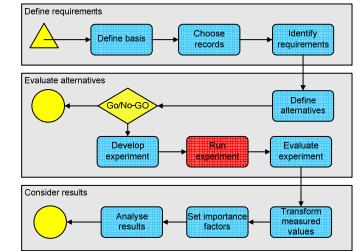
- Deliberate step for deciding whether it will be useful and cost-effective to continue, given
  - the resources to be spent (people, money)
  - the expected results
- Review of the experiment and evaluation process
  - Given the objectives, is the experiment
    - Complete?
    - Correct?

# Develop Experiment



- Define for each alternative
  - software tools and environment
  - hardware configuration
  - invocation parameters
- PLANETS will provide service discovery through registries

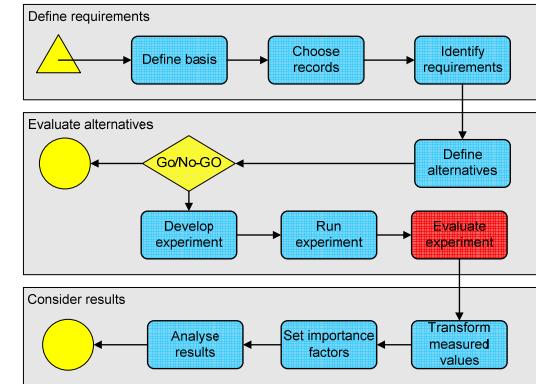
# Run Experiment



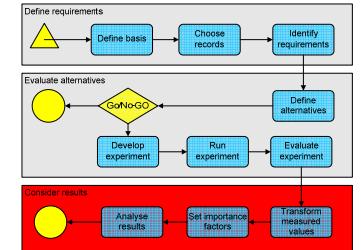
- Run experiment with the previously defined sample records
  - Manual tool execution
  - Automated service calls
- The results will be evaluated in the next stage

# Evaluate Experiment

- Evaluate how successfully the requirements are met
- Measure performance with respect to leaf criteria in the objective tree
- Document the results



# Phase 3: Consider Results

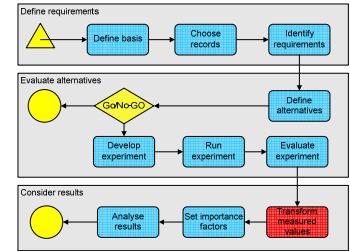


9. Transform measured values

10. Set importance factors

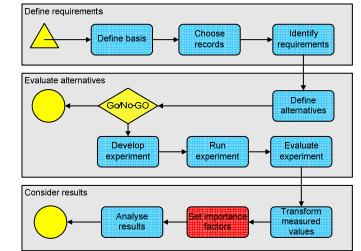
11. Analyse results

# Transform Measured Values



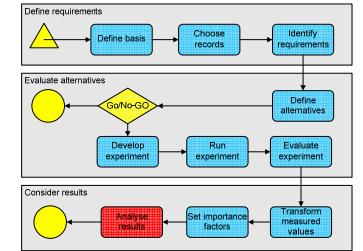
- Measures come in seconds, euro, bits, goodness values,...
- Need to make them comparable
- Transform measured values to uniform scale
- Transformation tables for each leaf criterion
- Scale 1-5 plus "not-acceptable"

# Set Importance Factors



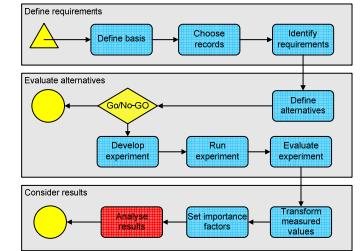
- Not all leaf criteria are equally important
- Set relative importance of all siblings in a branch
- Weights are propagated down the tree to the leaves

# 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
  - Performance values for each alternative

# Analyse Results



- Rank alternatives according to overall performance value at root
- Performance of each alternative
  - overall
  - for each sub-criterion (branch)
- Comparison of different alternatives
- Optional refinement of
  - Objective tree
  - Transformation rules
  - Importance factors
  - Alternatives

# 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](#)

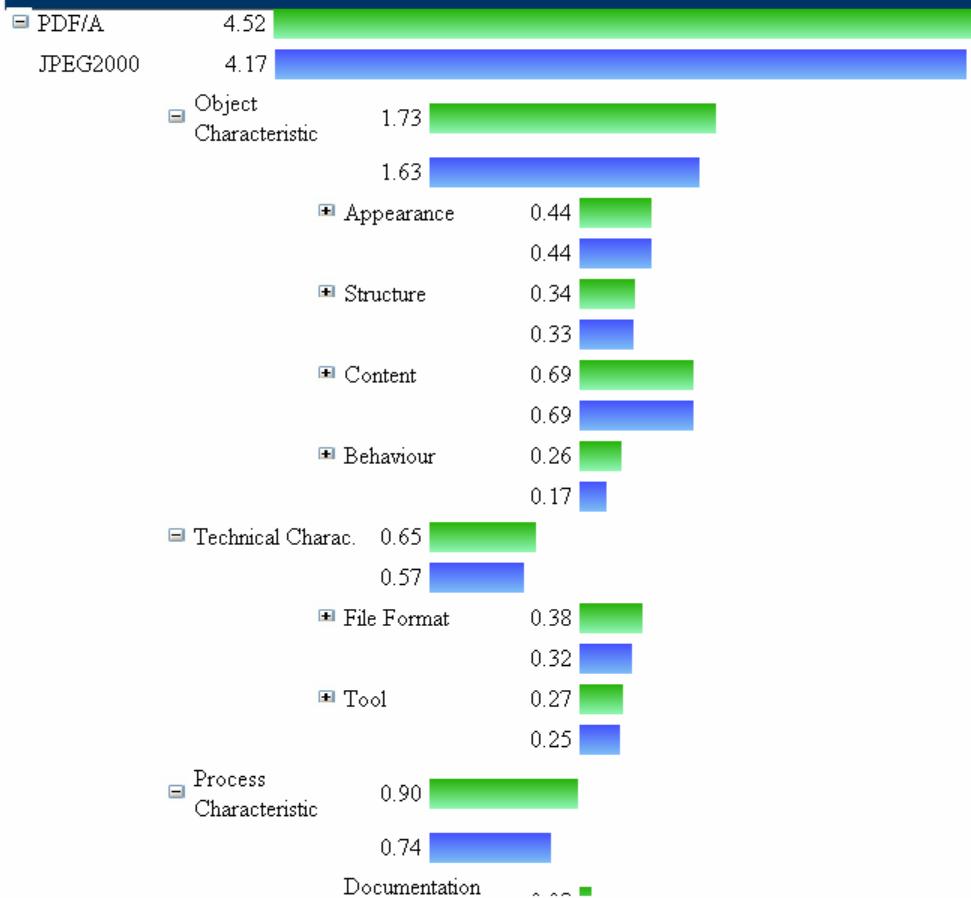
**Output**

- [XML Output](#)

**Contact**

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

## XII. Aggregate Alternatives



**alternative**

PDF/A  
RTF ConvertDoc  
TXT

**aggregation**

sum

**depth** 2

horizontal

vertical

join

show

Step11 Step13

# Summary

- Preservation planning
- Evaluate alternative preservation strategies
- Clearly defined and measurable criteria
- Well-defined experiments
- Documented and accountable decisions

# Case Study: Image Archive

---

- Austrian National Library
- 1 million images
  - 200.000 graphics
  - 800.000 photographs mostly taken between WW1 and WW2
- 200.000 digitised images at the moment
  - Different scanning techniques
- Images on a NAS
- Metadata in database
  
- Storage space
- Online access copy
- Print quality

# Next Steps

- Tool support
  - Integrate services
    - Migration
    - Emulation
    - Object characterisation
- Automate evaluation
- Continue case studies
  - Image archive (Austrian National Library)
- Template trees and fragments