wePreserve

The Planets Preservation Planning workflow and the planning tool Plato

March, 27th 2009

Hannes Kulovits

Vienna University of Technology
http://www.ifs.tuwien.ac.at/~kulovits

Information Society

digital preservation Europe

planets

CASPAR

nestor

Outline

- Preservation Planning
  - Evaluation of potential actions
- The Planets Preservation Planning Workflow
  - Workflow walkthrough
  - Requirements definition
  - The planning tool Plato

wePreserve

Evaluating preservation strategies

- Variety of solutions and tools exist
- Each strategy has unique strengths and weaknesses
- Requirements vary across settings
- Decision on which solution to adopt is complex
- Documentation and accountability is essential

- Preservation planning assists in decision making
- Evaluating preservation strategies on representative samples according to specific requirements and criteria

Information Society

digital preservation europe

planets

CASPAR

nestor

wePreserve

Scenario: Changes in user community

- Repository of electronic publications
- Policy: 90% of users can access all published reports
- Usage profile: 98% of users *can not* view `dvi` files
- Content profile: 5% of published reports in `dvi` format

- Mission: Build and execute a plan for preserving access to these documents for the designated user community

Information Society

digital preservation europe

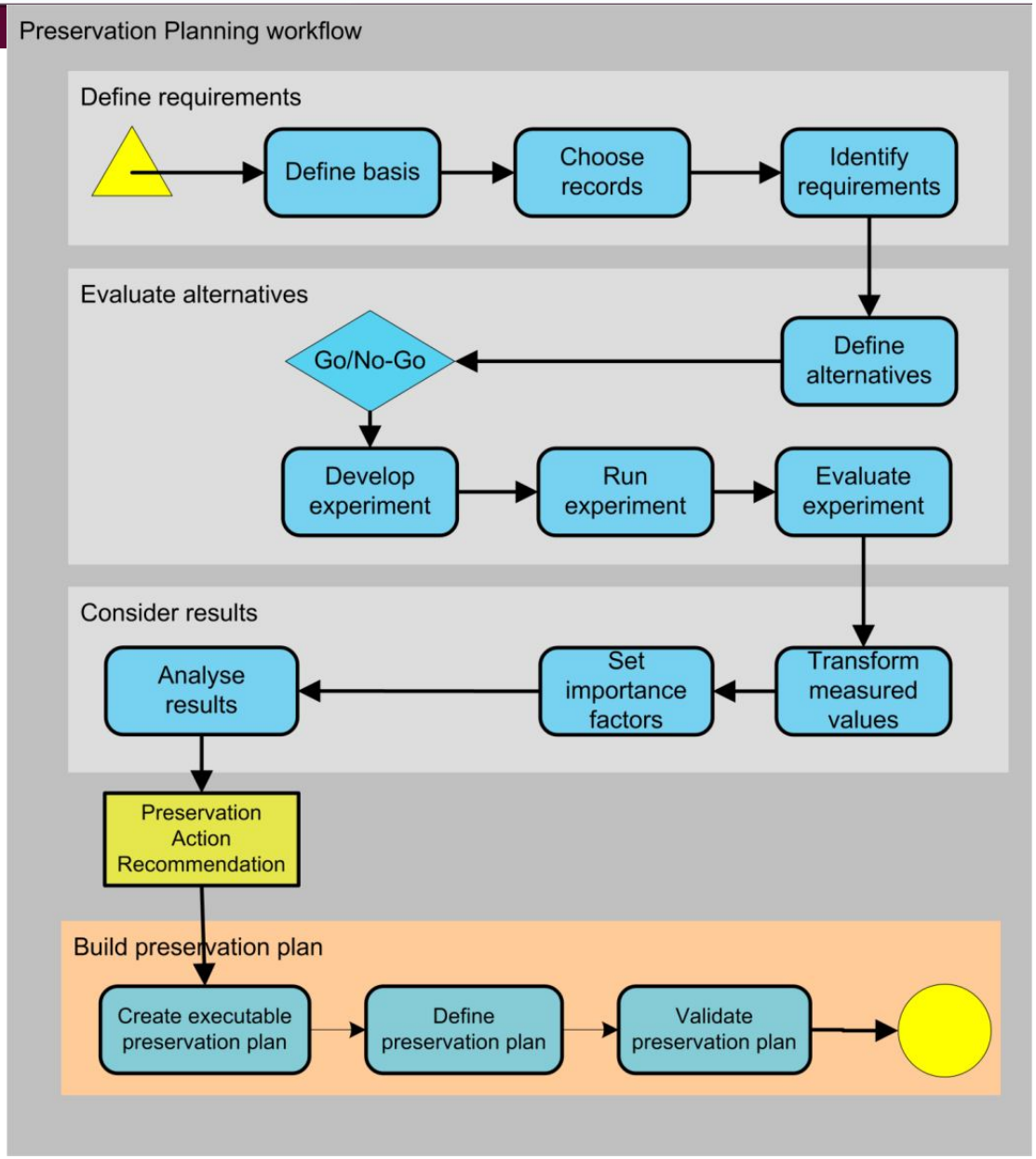
CASPAR

nestor

wePreserve

Planets Preservation Planning Workflow

- Define requirements
- Evaluate potential actions
- Analyse results
- Build a preservation plan



wePreserve

Preservation Planning in Plato

- Web based planning tool implementing the Planets preservation planning workflow
- Publicly available
- Automation of the planning process
  - Integration of registries and services for
    - File format identification
    - Preservation action (migration, emulation…)
    - Characterisation and comparison
- Knowledge base to support planning

Information Society

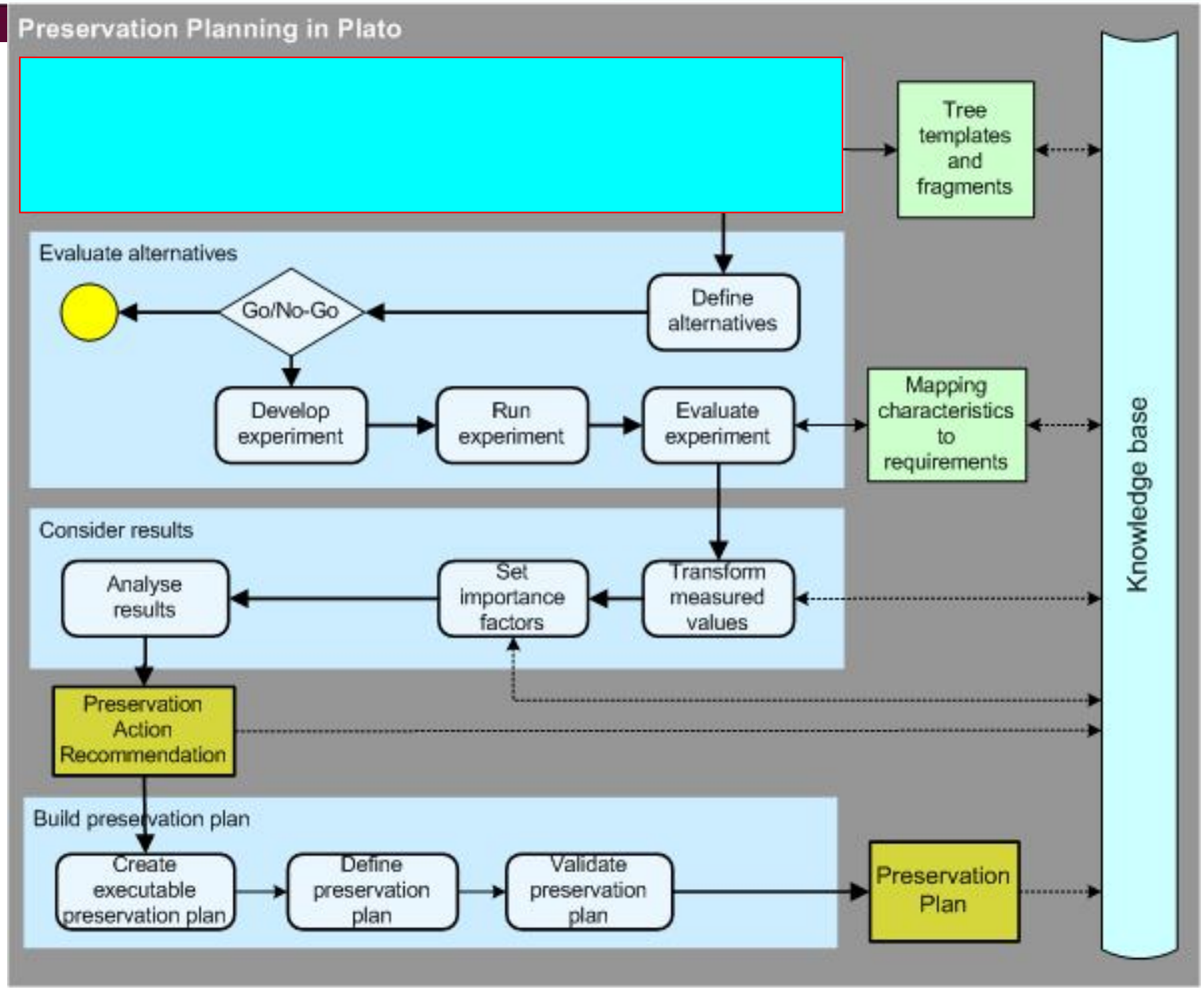
digital preservation europe

planets

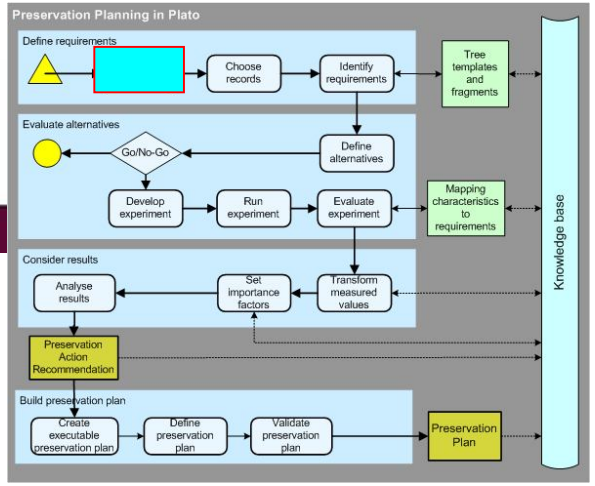
CASPAR

nestor

PP Workflow


Preservation Planning in Plato
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results
Set importance factors
Transform measured values
Preservation Action Recommendation
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan
Knowledge base

wePreserve



Define basis

- Document basic assumptions and constraints
  - Types of objects
  - Purpose of planning
  - Mandates and designated community
  - Applying policies
  - Triggers that initiated the planning process

Information Society

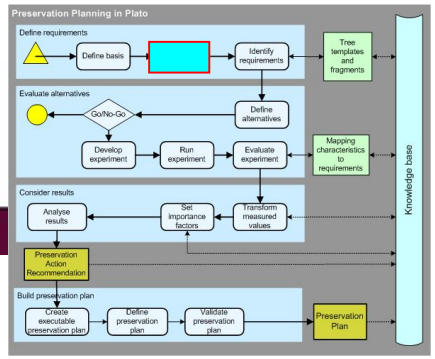
digital preservation europe

planets

CASPAR

nestor

wePreserve



Choose sample objects/records

- Define the set of objects that are the subject of preservation planning
  - Size of the collection
  - Growth rate
  - Object format
  - …
- Specify representative sample objects that cover the variety of significant properties and technical characteristics

Information Society

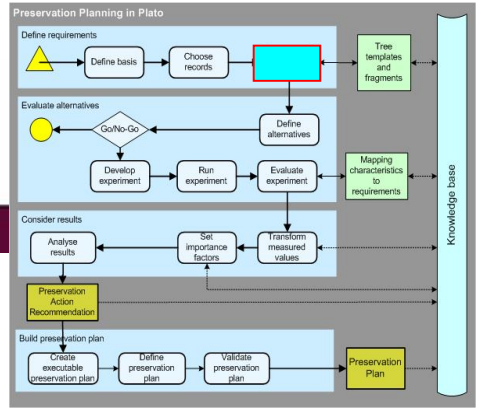
digital preservation europe

planets

CASPAR

nestor

wePreserve



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

Information Society

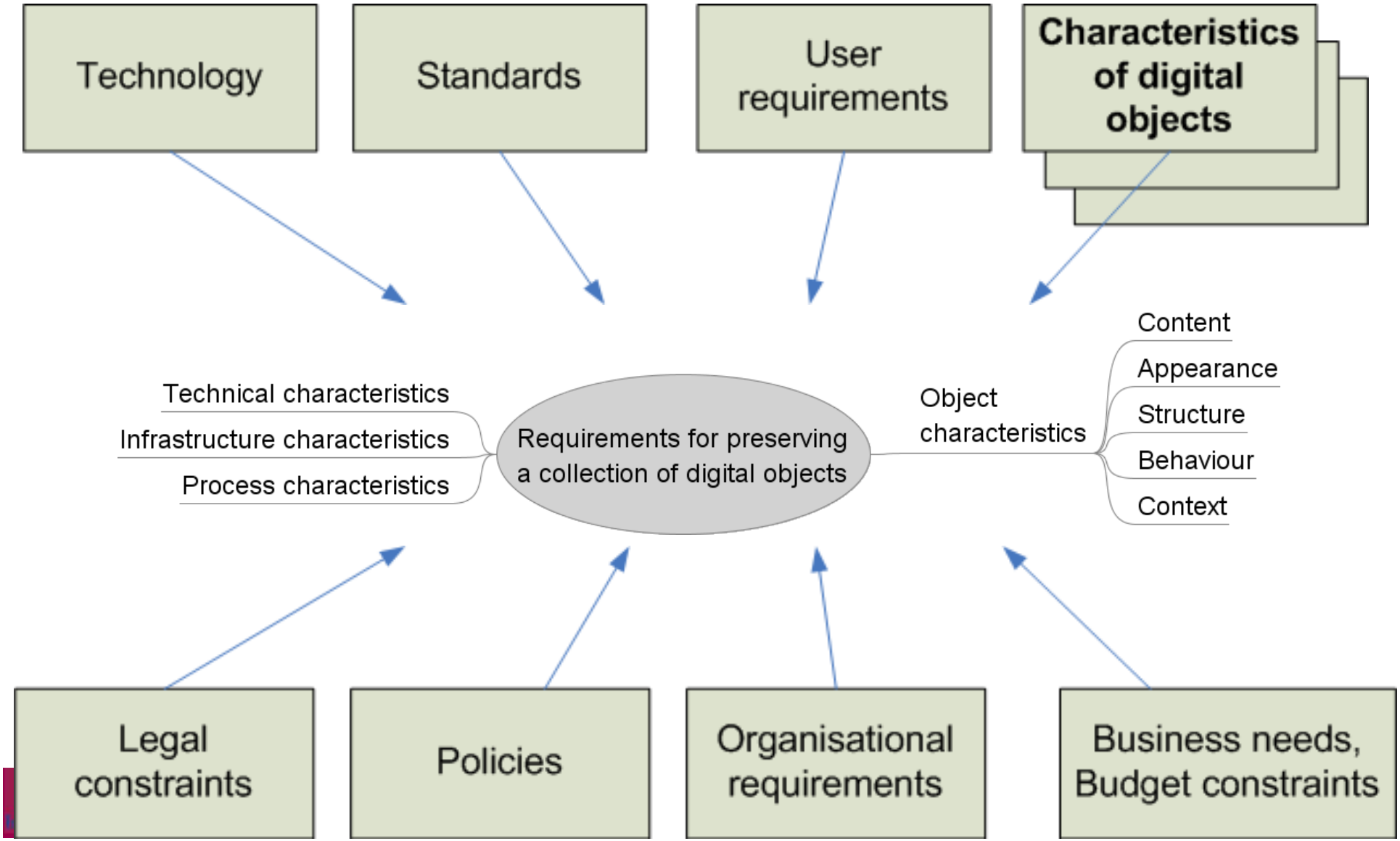
digital preservation europe

planets

CASPAR

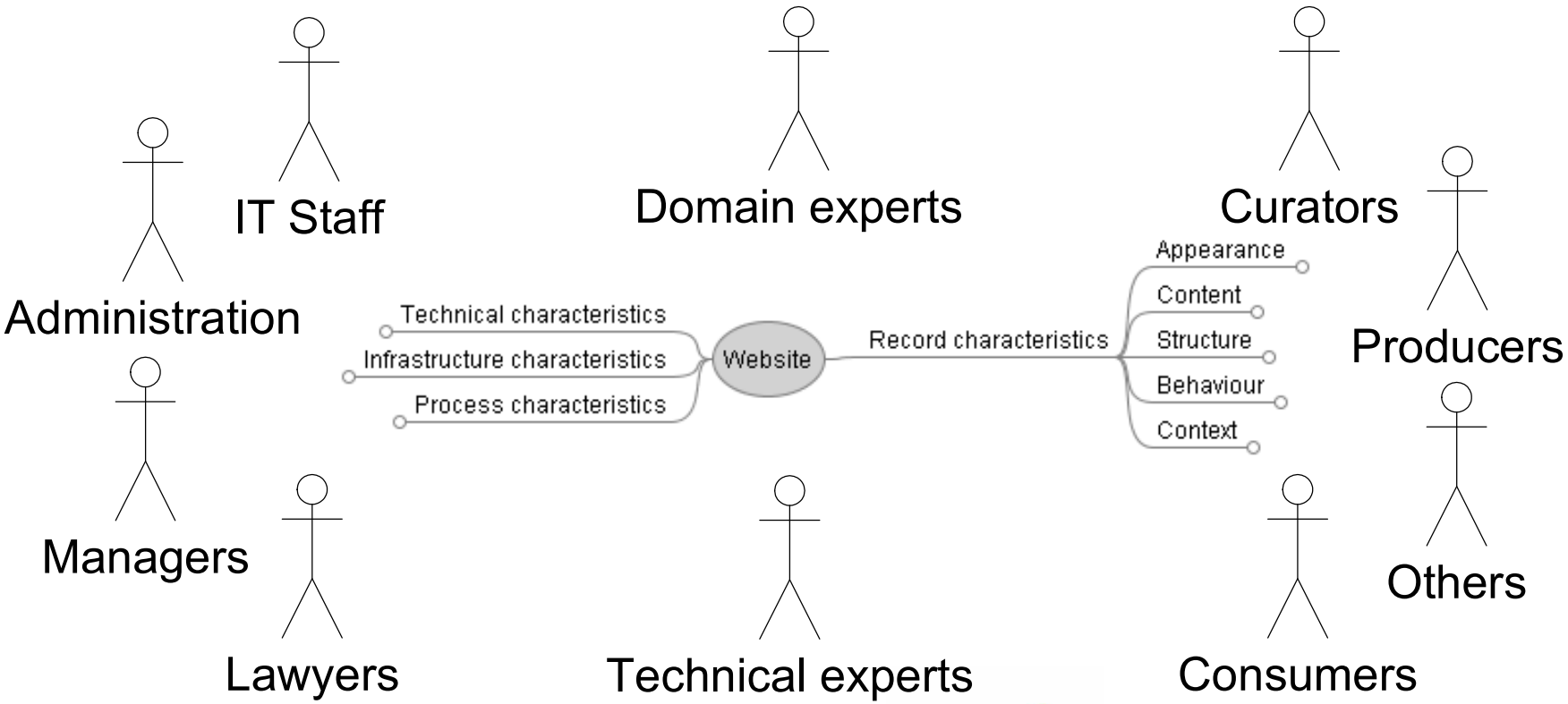
nestor

Influence Factors

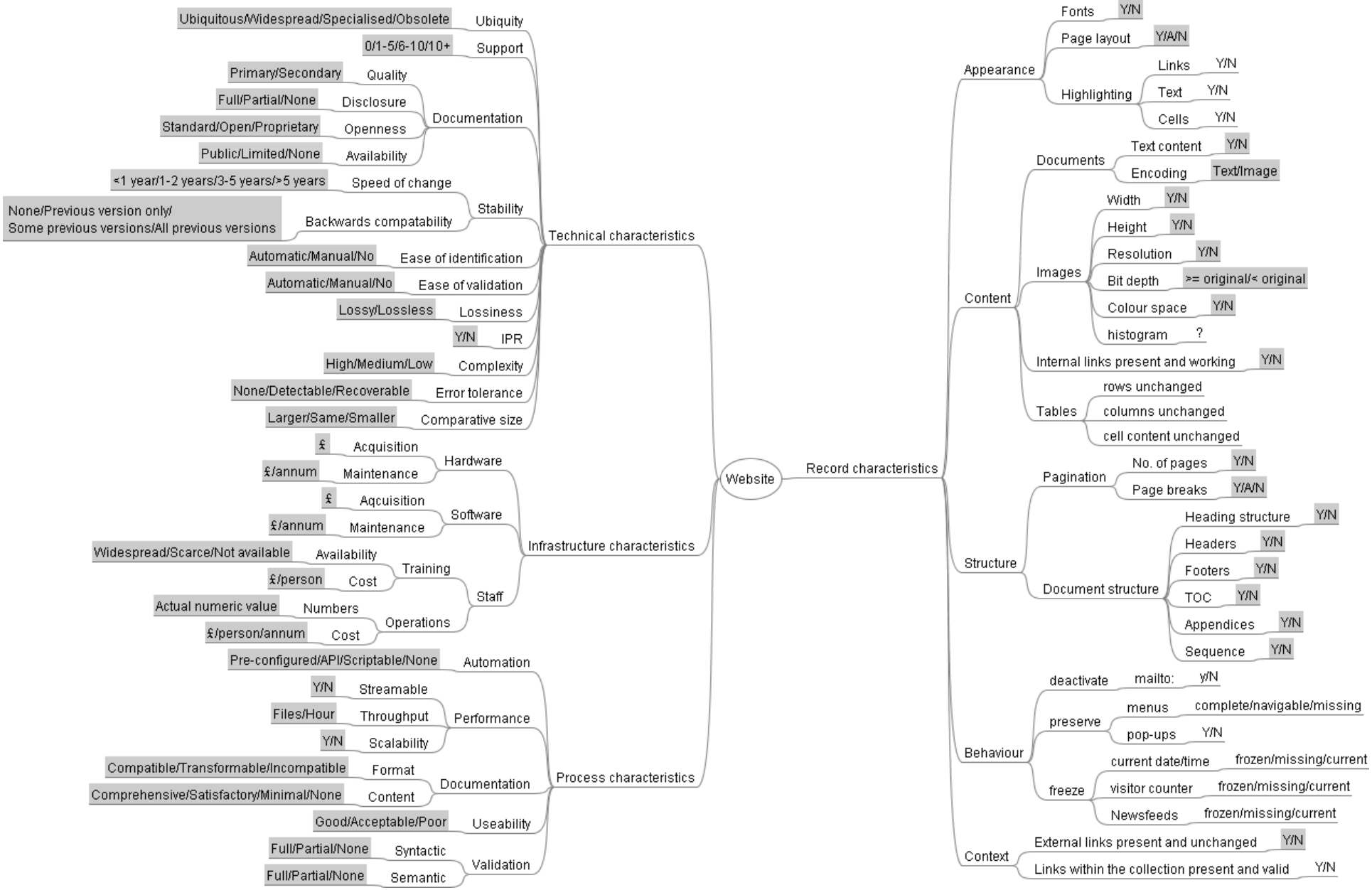


Stakeholders

- Input needed from a wide range of persons, depending on the institutional context and the collection

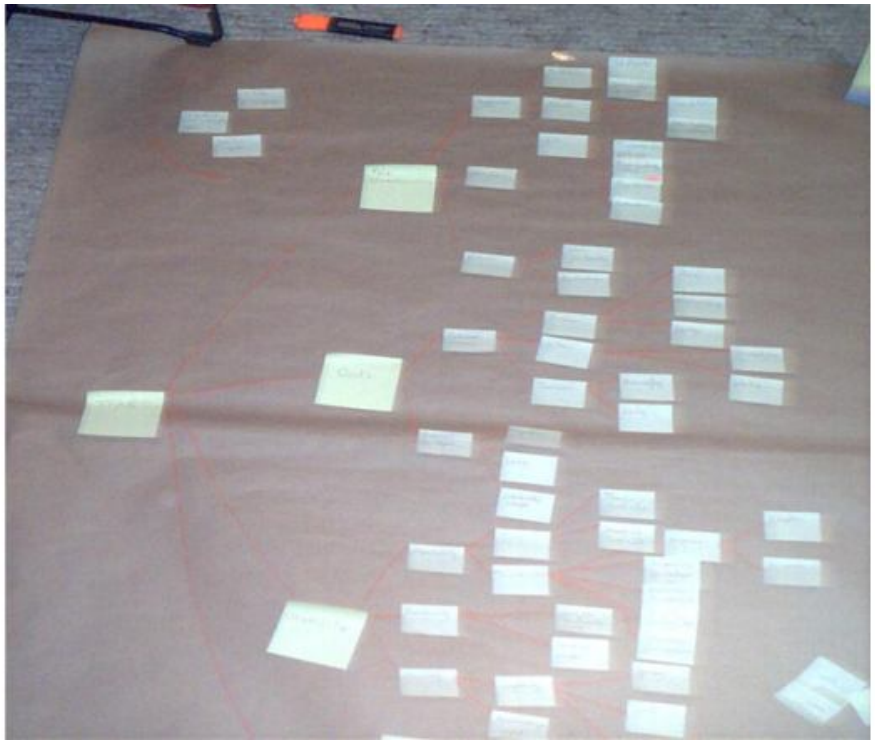


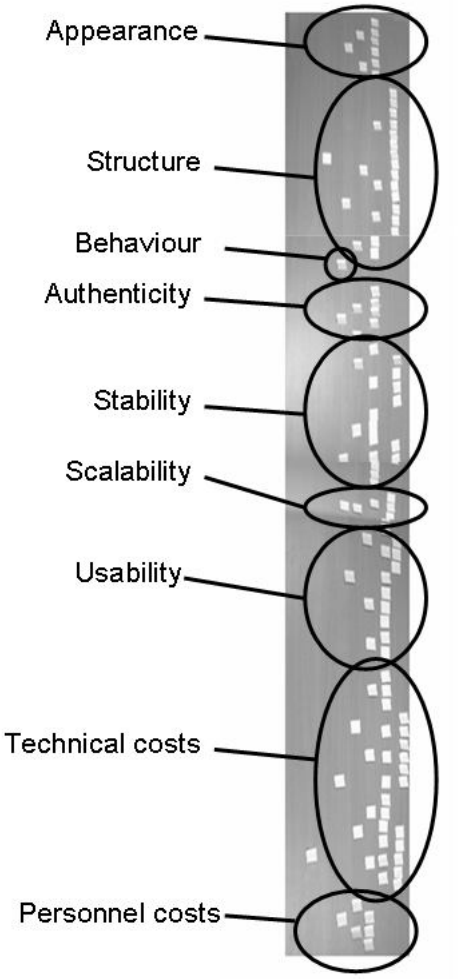
An Objective Tree



wePreserve

Analog…





Information Society

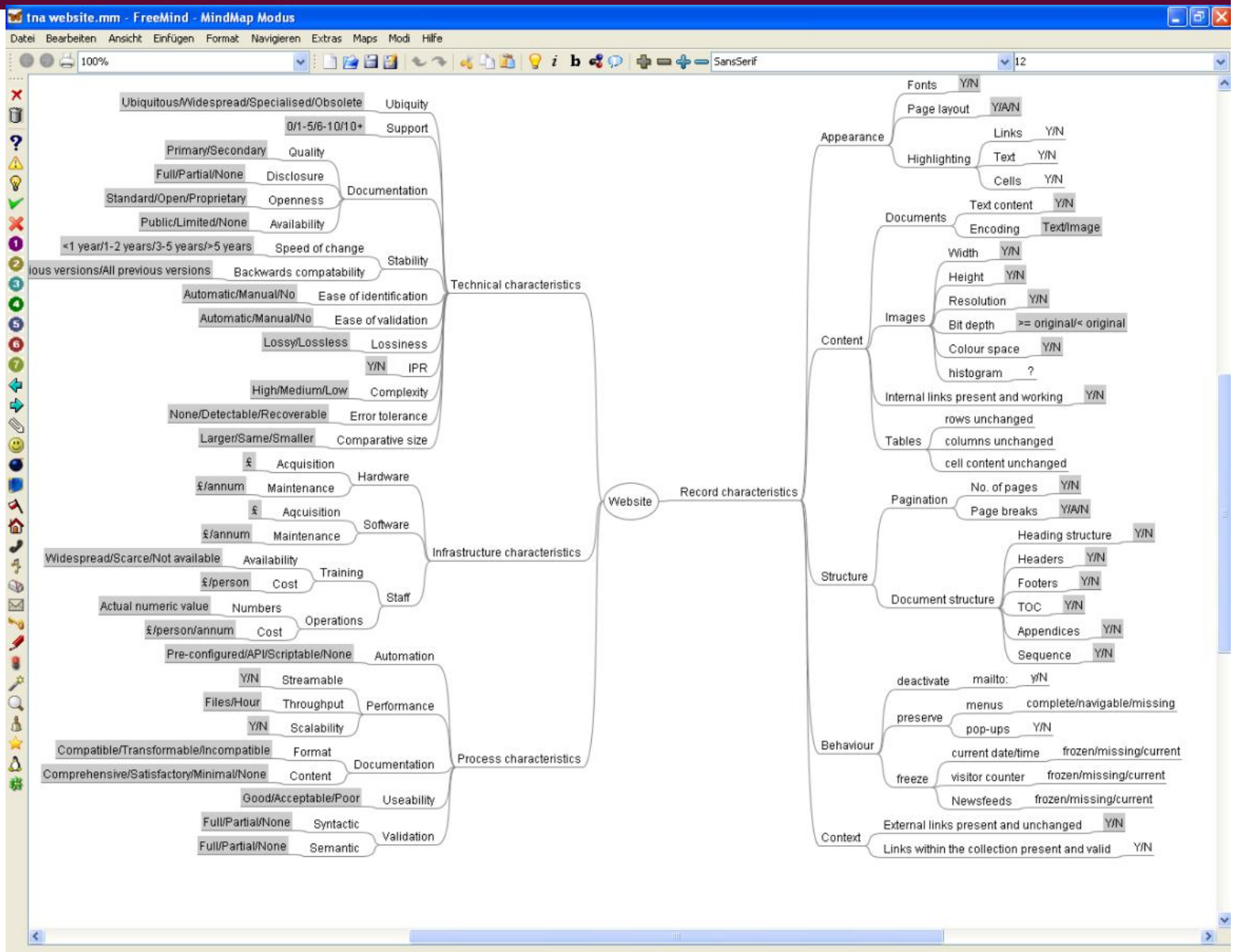
digital preservation europe

planets

CASPAR

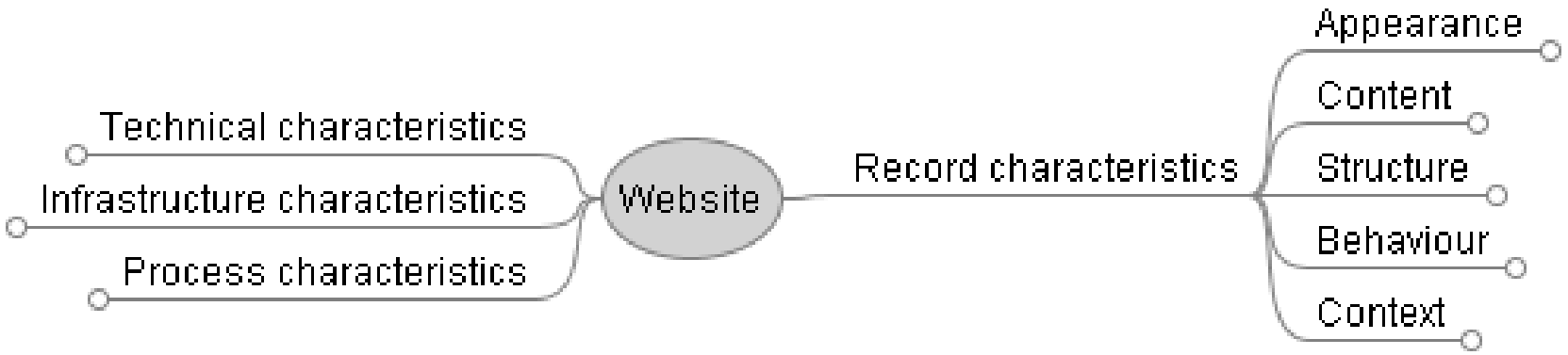
nestor

… or born-digital



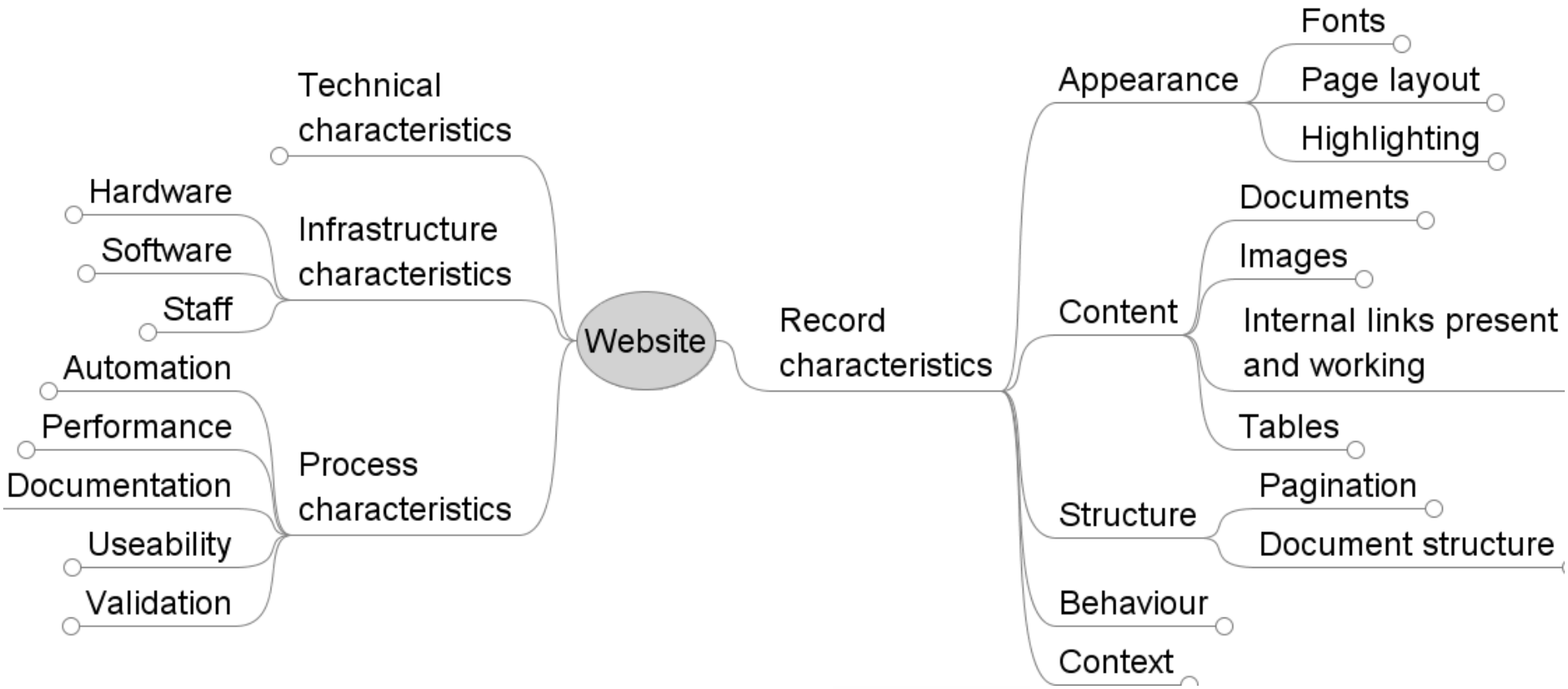
Case Study: Web archiving

- Static web pages from the public domain
- Includes documents in formats such as doc, pdf
- Images
- No interactive content shall be preserved



wePreserve

A bit more detail…



digital preservation europe

planets

CASPAR

nestor

Assign Measurable Units

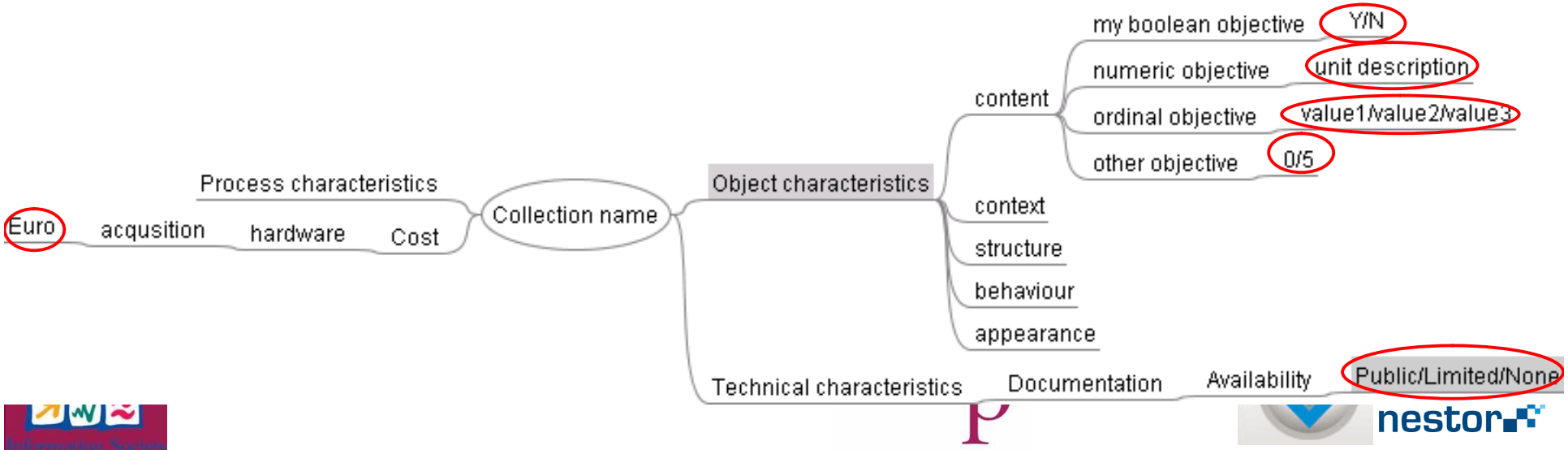
- Leaf criteria should be objectively measurable
  - Seconds per object
  - Euro per object
  - Bits of colour depth

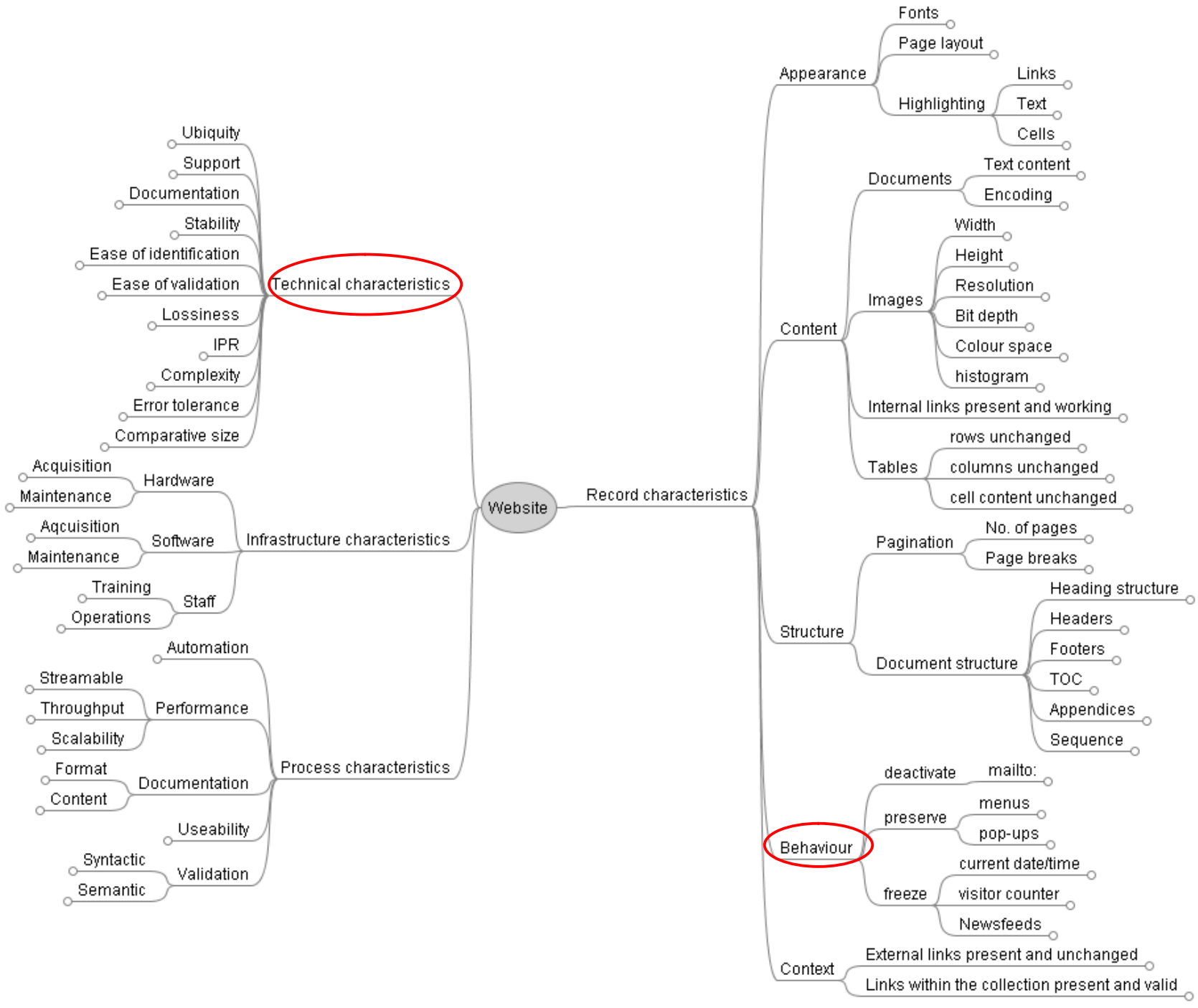
- Subjective scales where necessary
  - Adoption of file format
  - Amount of (expected) support

➢ Quantitative results

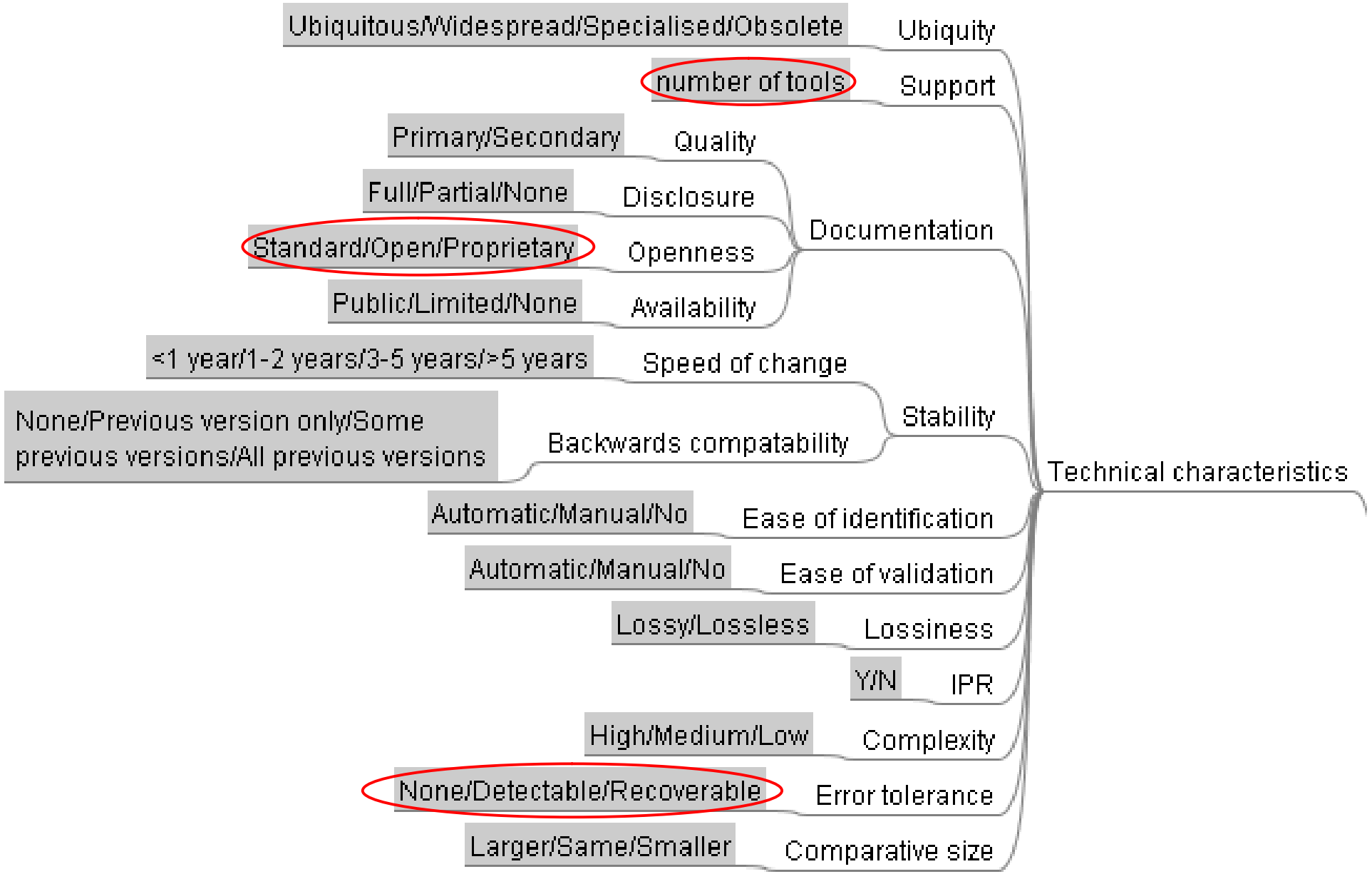
Types of scales

- Numeric (unit)
- Yes/No (Y/N)
- Yes/Acceptable/No (Y/A/N)
- Ordinal: define the possible values (good/bad/ugly)
- Subjective 0-to-5 (0/5)




Website
Technical characteristics
Ubiquity
Support
Documentation
Stability
Ease of identification
Ease of validation
Lossiness
IPR
Complexity
Error tolerance
Comparative size
Infrastructure characteristics
Hardware
Acquisition
Maintenance
Software
Aqcuisition
Maintenance
Staff
Training
Operations
Process characteristics
Automation
Performance
Streamable
Throughput
Scalability
Documentation
Format
Content
Useability
Validation
Syntactic
Semantic
Record characteristics
Appearance
Fonts
Page layout
Highlighting
Links
Text
Cells
Content
Documents
Text content
Encoding
Images
Width
Height
Resolution
Bit depth
Colour space
histogram
Internal links present and working
Tables
rows unchanged
columns unchanged
cell content unchanged
Structure
Pagination
No. of pages
Page breaks
Document structure
Heading structure
Headers
Footers
TOC
Appendices
Sequence
Behaviour
deactivate
mailto:
preserve
menus
pop-ups
freeze
current date/time
visitor counter
Newsfeeds
Context
External links present and unchanged
Links within the collection present and valid

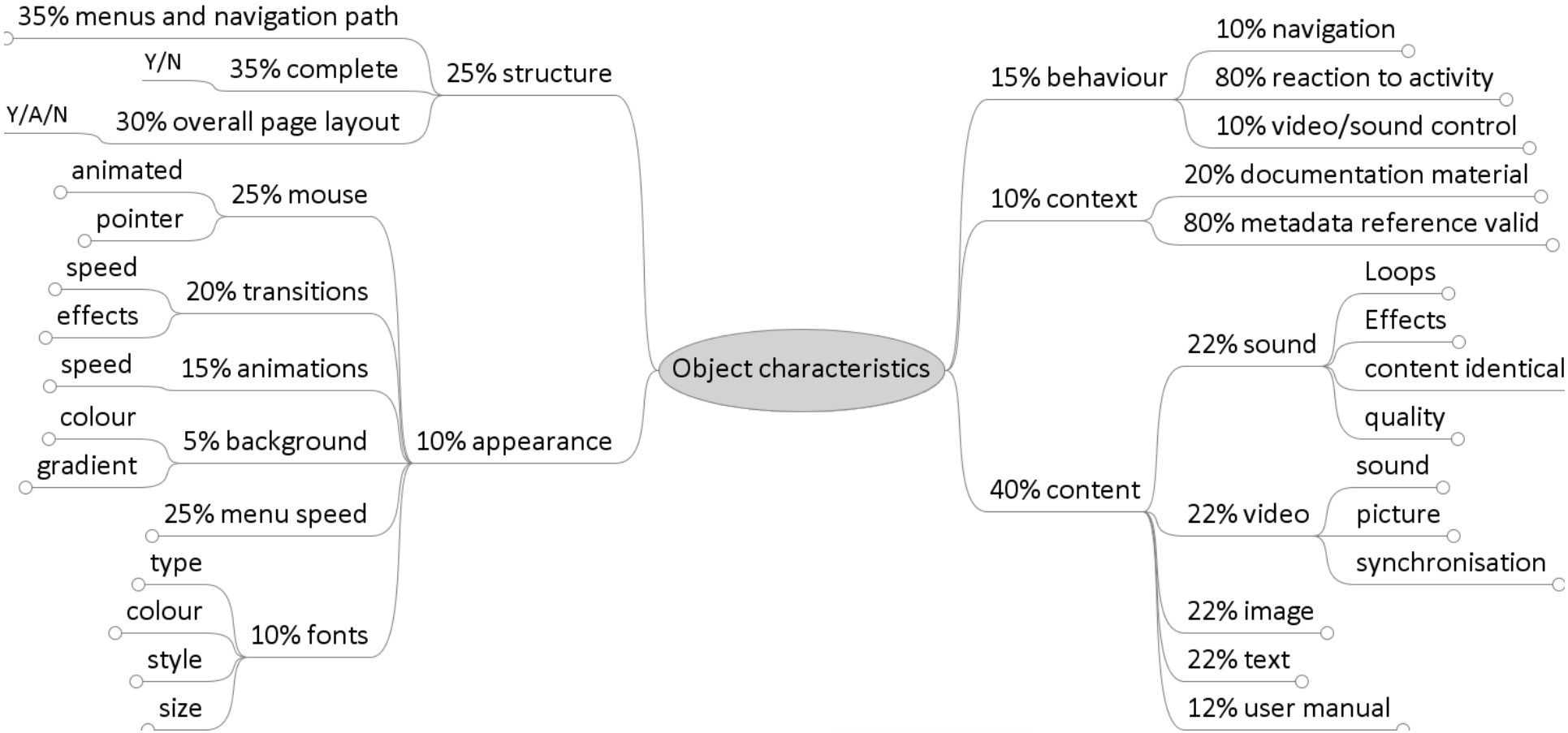
File format characteristics



Behaviour

- **Visitor counter** and similar things can be
  - Frozen at the point of harvesting
  - Left out
  - Still counting while being accessed in the archive (Is this desirable?)

Interactive multimedia

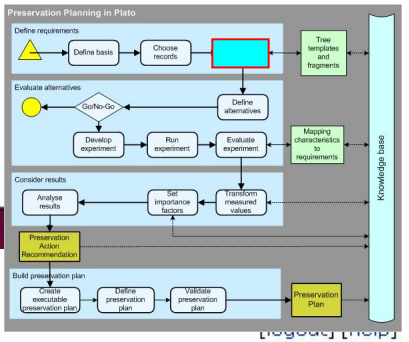


Behaviour

- Interactive presentations exhibit two facets
  - Graph-like navigation structure
  - Navigation along the paths

Node	Scale	Restriction
Object characteristics		
behaviour		
navigation	Ordinal	interactive and integrated/navigatable/none
reaction to activity		
mouse		
position	Boolean	
clicks	Boolean	
keyboard	Boolean	
video/sound control		
structure		
menus and navigation path	Ordinal	complete and free/partial (linear)/none
complete	Boolean	
overall page layout	Ordinal	Y/A/N

Objective Tree



PLANETS Preservation Planning Tool *(Plato)*

Institute of Software Technology and Interactive Systems

[logout] [help]

Project | Define Requirements :: Evaluate Requirements :: Consider Results | Loaded project: PP4 workshop - The National Archive

Identify Requirements

Expand All | Collapse All

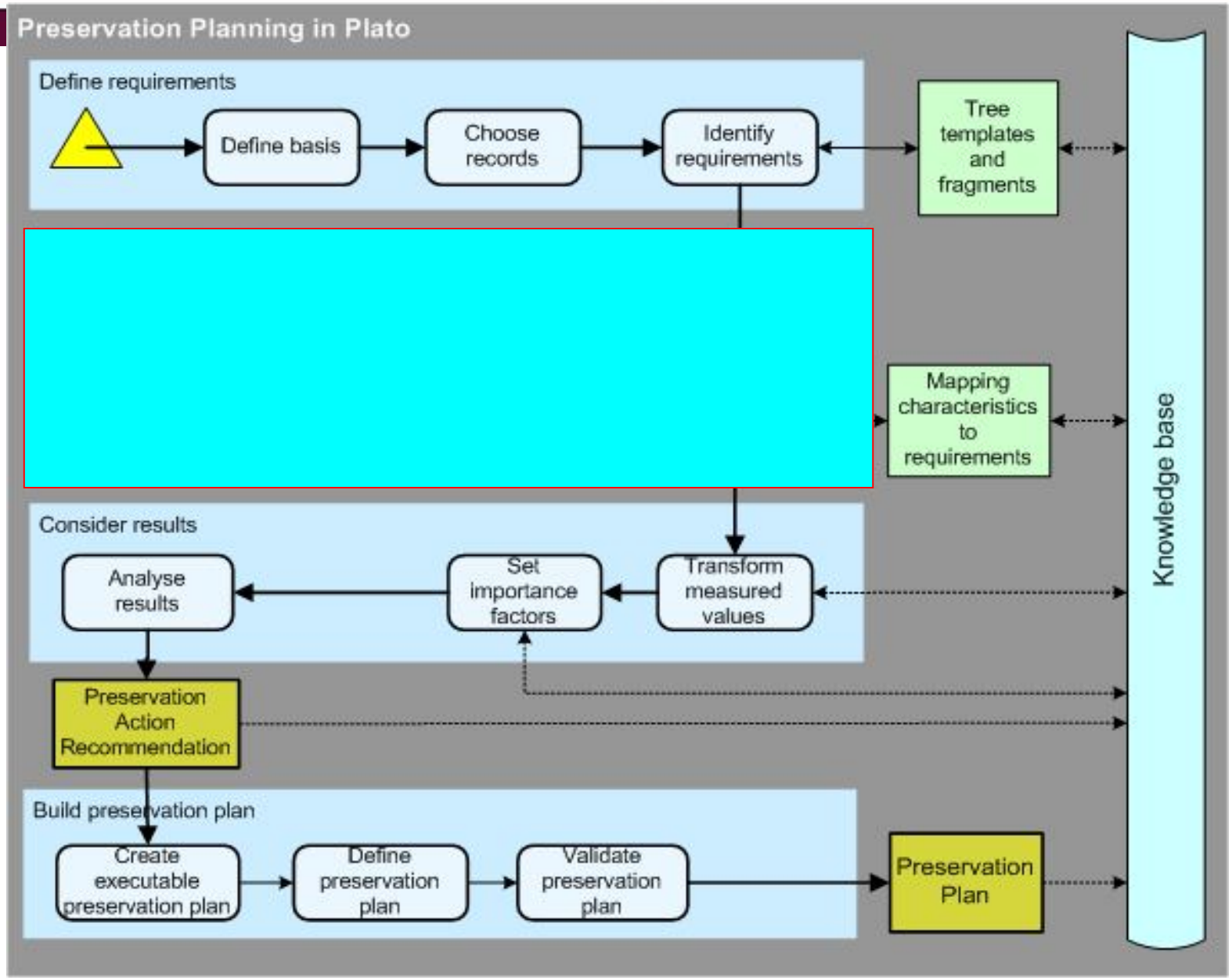
Website

Focus	Node	Single	Scale	Restriction	Unit
	▼ Website				
X	▼ Record characteristics				
X	▶ Appearance				
X	▶ Content				
X	▶ Structure				
X	▼ Behaviour				
X	▼ deactivate				
X	▶ mailto:	☐	Boolean	Yes/No	
X	▼ preserve				
X	▶ menus	☐	Ordinal	complete/navigable/missing	
X	▶ pop-ups	☐	Boolean	Yes/No	
X	▼ freeze				
X	▶ current date/time	☐	Ordinal	frozen/missing/current	
X	▶ visitor counter	☐	Ordinal	frozen/missing/current	
X	▶ Newsfeeds	☐	Ordinal	frozen/missing/current	
X	▶ Context				
X	▼ Technical characteristics				
X	▶ Ubiquity	☐	Ordinal	Ubiquitous/Widespread/Specialised/Obs	
X	▶ Tool Support	☐	Positive Number		Number of tools
X	▶ Documentation				
X	▶ Stability				
X	▶ Ease of identification	☐	Ordinal	Automatic/Manual/No	
X	▶ Ease of validation	☐	Ordinal	Automatic/Manual/No	
			Ordinal	Lossy/Lossless	

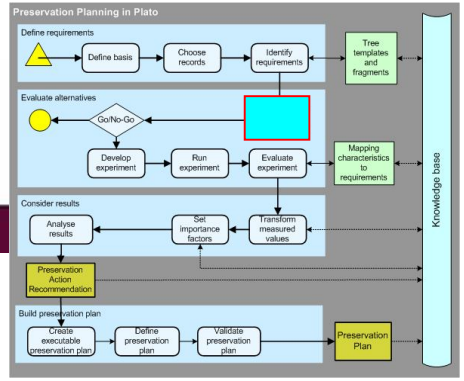
© 2007 Institute of Software Technology and Interactive Systems: «office bears»

Quick Access:

PP Workflow


Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Mapping characteristics to requirements
Consider results
Analyse results
Set importance factors
Transform measured values
Preservation Action Recommendation
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan
Knowledge base

wePreserve



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

Information Society

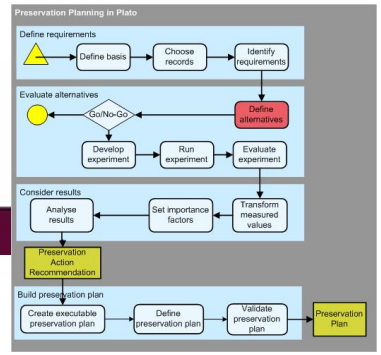
digital preservation europe

planets

CASPAR

nestor

wePreserve



Discovering possible actions

Create alternatives from applicable services

Sample record #1 has format JPEG File Interchange Format, 1.01.
You can look up services that are able to handle this object type in the following registries:

Planets Preservation Action Tool registry

planets

Show Preservation Services

Planets Service Registry

planets

Show Preservation Services

CRiB Service Registry

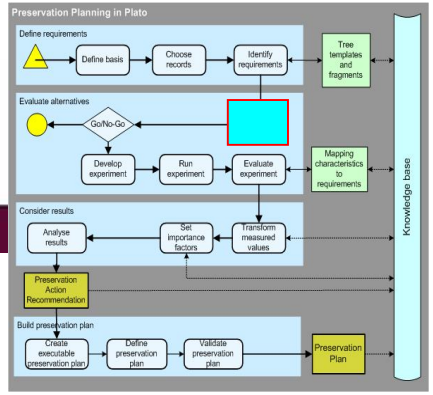
Show Preservation Services

	Preservation Action	Target Format	Info
☐	JPG > BMP	Windows Bitmap, version 3.0	JPG>BMP
☑	JPG > TIF	Tagged Image File Format, version 3	JPG>BMP>TIF
☐	JPG > TIF #2	Tagged Image File Format, version 3	JPG>TIF
☑	JPG > TIF_2	Tagged Image File Format, version 3	JPG>TIF_2
☐	JPG > PNG	Portable Network Graphics, version 1.0	JPG>PNG
☐	JPG > JP2	JPEG 2000	JPG>JP2

Create alternatives for selected services

stor

wePreserve



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,...)

Information Society

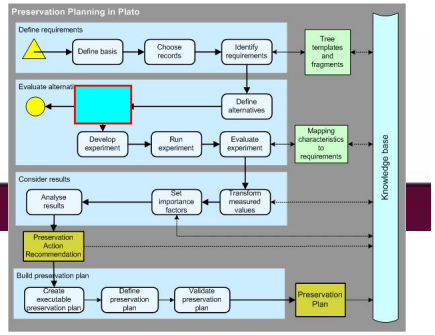
digital preservation europe

planets

CASPAR

nestor

wePreserve



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 readdressed or not?

Information Society

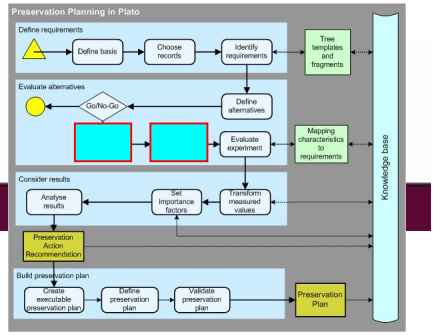
digital preservation europe

planets

CASPAR

nestor

wePreserve



Develop and run experiment

- Formulate for each experiment detailed
  - Development plan
    - steps to build and test software components
    - procedures and preparation
    - parameter settings for integrating preservation services
  - Evaluation/experiment plan (workflow/sequence of activities)
- Apply the selected potential preservation actions on the sample objects

Information Society

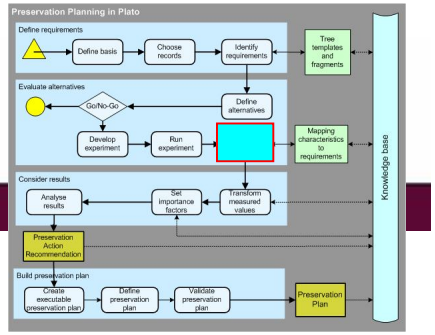
digital preservation europe

planets

CASPAR

nestor

wePreserve



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.

Information Society

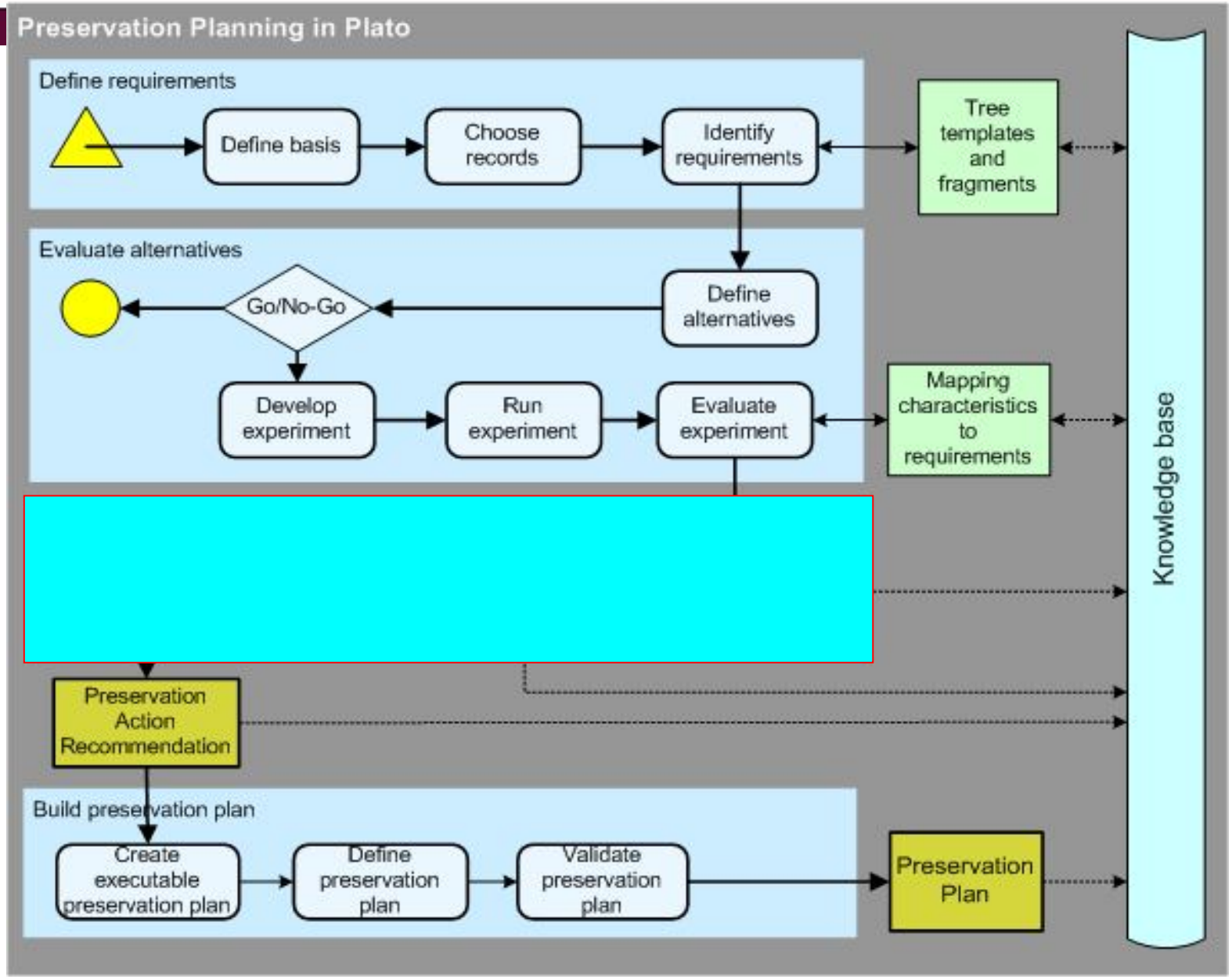
digital preservation europe

planets

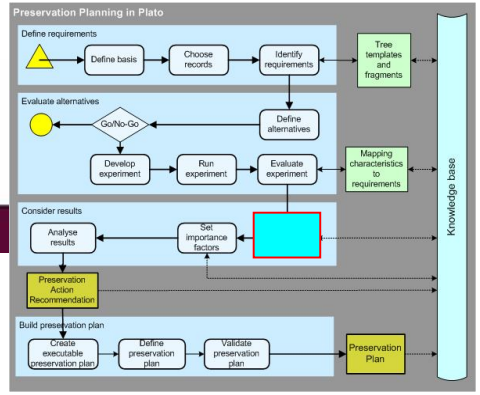
CASPAR

nestor

PP Workflow


Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Knowledge base
Preservation Action Recommendation
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan

wePreserve



Transform measured values

- Measures come in seconds, euro, bits, goodness values, …
- Need to make them comparable
- Transform measured values to uniform scale
- Target scale 0-5

Information Society

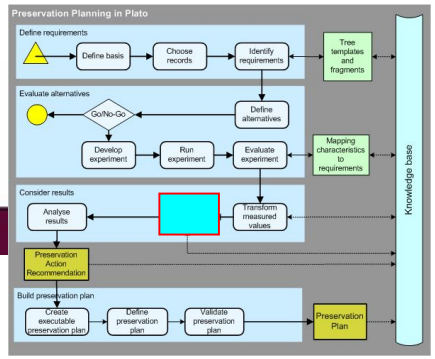
digital preservation europe

planets

CASPAR

nestor

wePreserve



Set importance factors

PLANETS Preservation Planning Tool - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://localhost:8080/plato/workflow/importancefactors.seam

PLANETS Preservation Planning Tool *(Plato)*

Project | Define Requirements :: Evaluate Requirements :: Consider Results | Project 'Minimalist

Set Importance Factors

Balance weights automatically ☑

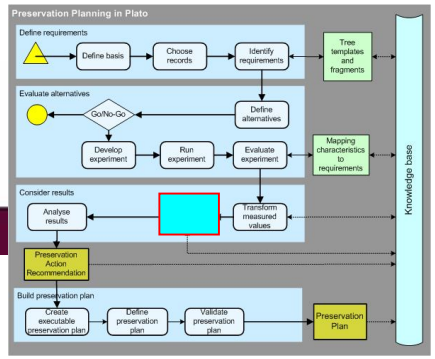
Expand All | Collapse All

Object characteristics

Focus	Name	Weight	Lock	Total weight
	▼ Object characteristics	0 — 1 — 1	☐	1
X	▶ behaviour	0 — 1 — 0.15	☑	0.15
X	▶ structure	0 — 1 — 0.25	☑	0.25
X	▶ context	0 — 1 — 0.1	☐	0.1
X	▶ appearance	0 — 1 — 0.1	☐	0.1
X	▶ content	0 — 1 — 0.4	☑	0.4

Save Proceed

wePreserve



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
- Rank alternatives according to overall performance value at root
- Performance of each alternative
  - overall
  - for each sub-criterion (branch)
- Comparison of different alternatives

Information Society

digital preservation europe

planets

CASPAR

nestor

Analyse results

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
Expand All | Collapse All
Polar bear image preservation

Focus	Name	Result
	⊟ Polar bear image preservation	TIFF (tool A): 4,78 TIFF (tool B): 4,28 PNG (tool D): 3,97
X	⊟ Process	TIFF (tool A): 1,65 TIFF (tool B): 1,16 PNG (tool D): 0,74
	Complexity	TIFF (tool A): 2,50 TIFF (tool B): 2,50 PNG (tool D): 1,25
	Cost	TIFF (tool A): 2,50 TIFF (tool B): 1,00 PNG (tool D): 1,00
X	⊟ Image properties	TIFF (tool A): 1,70 TIFF (tool B): 1,70 PNG (tool D): 1,70
	Bits of colour depth	TIFF (tool A): 5,00 TIFF (tool B): 5,00 PNG (tool D): 5,00
X	⊟ Technical characteristics	TIFF (tool A): 1,43 TIFF (tool B): 1,43 PNG (tool D): 1,53
	Official standard	TIFF (tool A): 3,50 TIFF (tool B): 3,50 PNG (tool D): 3,50
	Filesize (in Relation to Original)	TIFF (tool A): 0,83 TIFF (tool B): 0,83 PNG (tool D): 1,12

Conclusion

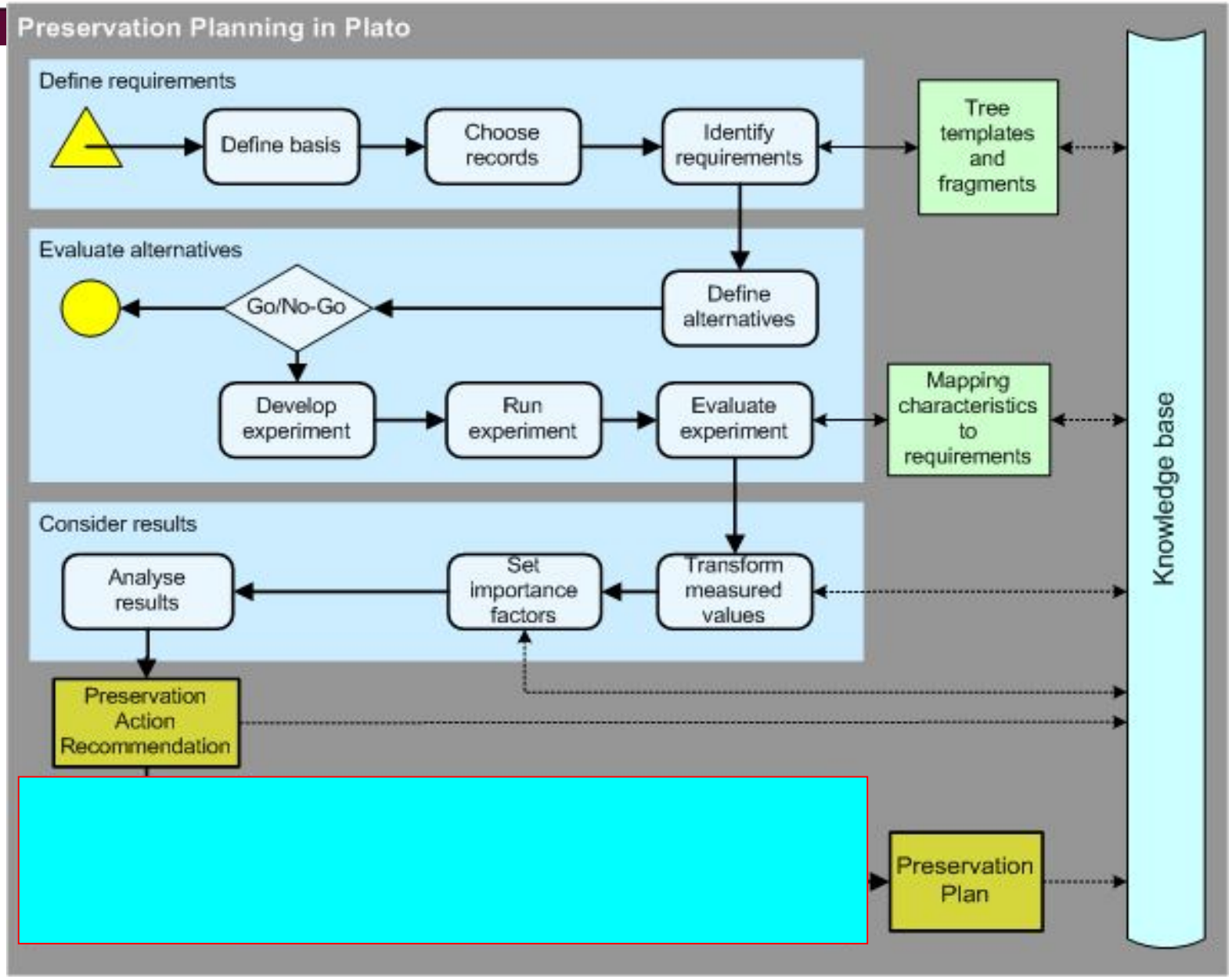
Recommendation

Recommendation:

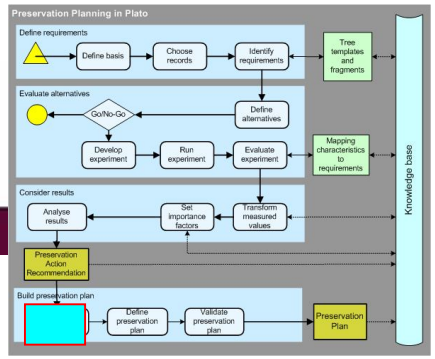
Reasoning:

Effects of applying this strategy:

PP Workflow


Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results
Set importance factors
Transform measured values
Preservation Action Recommendation
Preservation Plan
Knowledge base

wePreserve



Create executable plan

- Preservation Action Plan
- When?
  - Conditions and triggers for execution
  - Hardware and software requirements…
- What?
  - Single tool, composite workflow of services….
  - Validation and QA
  - Other actions needed, such as reporting…

Information Society

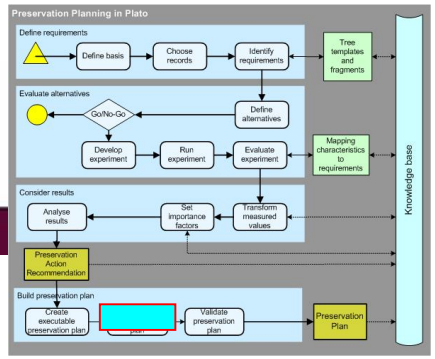
digital preservation europe

planets

CASPAR

nestor

wePreserve



Define preservation plan

- Executable action plan is not enough
  - Rules for monitoring
  - Evidence of decisions
  - Estimates of costs
  - Roles and responsibilities
  - …

Information Society

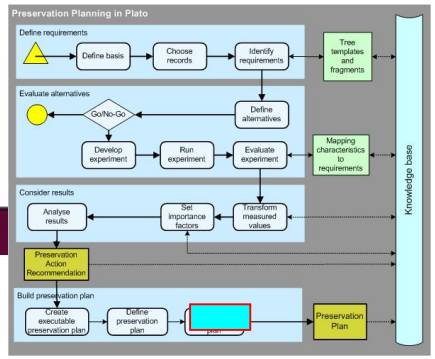
digital preservation europe

planets

CASPAR

nestor

wePreserve



Validate preservation plan

- Validate all elements of the plan
- Check for completeness
- Formally approve the plan and put it into action

- Continuous review and monitoring is necessary!

Information Society

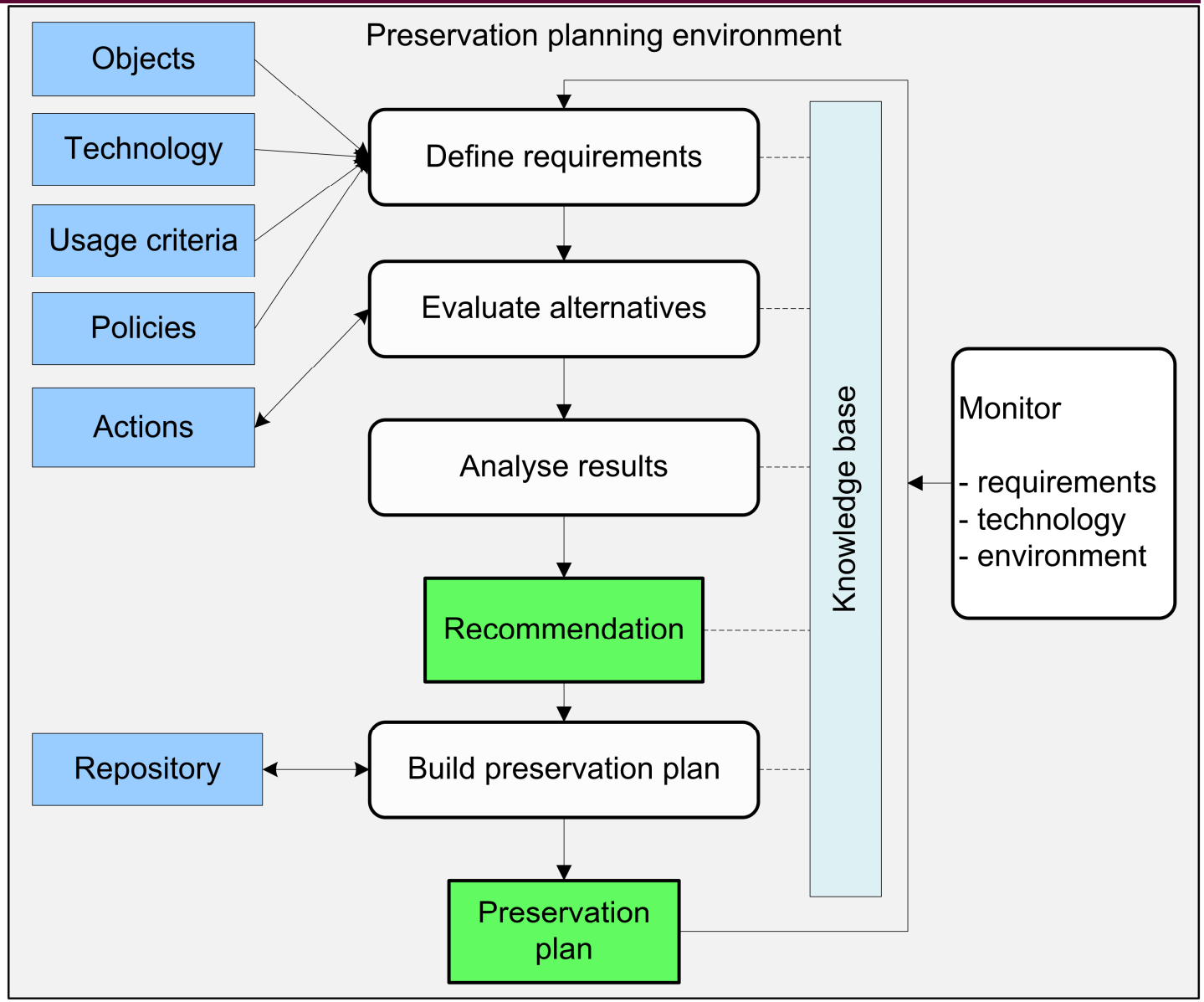
digital preservation europe

planets

CASPAR

nestor

Summary


Preservation planning environment
Objects
Technology
Usage criteria
Policies
Actions
Define requirements
Evaluate alternatives
Analyse results
Recommendation
Build preservation plan
Preservation plan
Repository
Knowledge base
Monitor
- requirements
- technology
- environment

Summary

Plato is openly available at:

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

wePreserve

Thank you

kulovits@ifs.tuwien.ac.at

www.ifs.tuwien.ac.at/dp/plato
www.planets-project.eu

digital preservation europe

planets

CASPAR

nestor

Information Society

This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License.
To view a copy of this license,
visit http://creativecommons.org/licenses/by-nc-nd/3.0/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.