



# The Planets Preservation Planning workflow and the planning tool Plato

Hannes Kulovits

Vienna University of Technology

<http://www.ifs.tuwien.ac.at/~kulovits>

- Preservation Planning
  - Evaluation of potential actions
- The Planets Preservation Planning Workflow
  - Workflow walkthrough
  - Requirements definition
  - The planning tool Plato
- Requirements definition exercise
  - Groups, scenarios, tasks
  - Schedule
- Demonstration: Plato

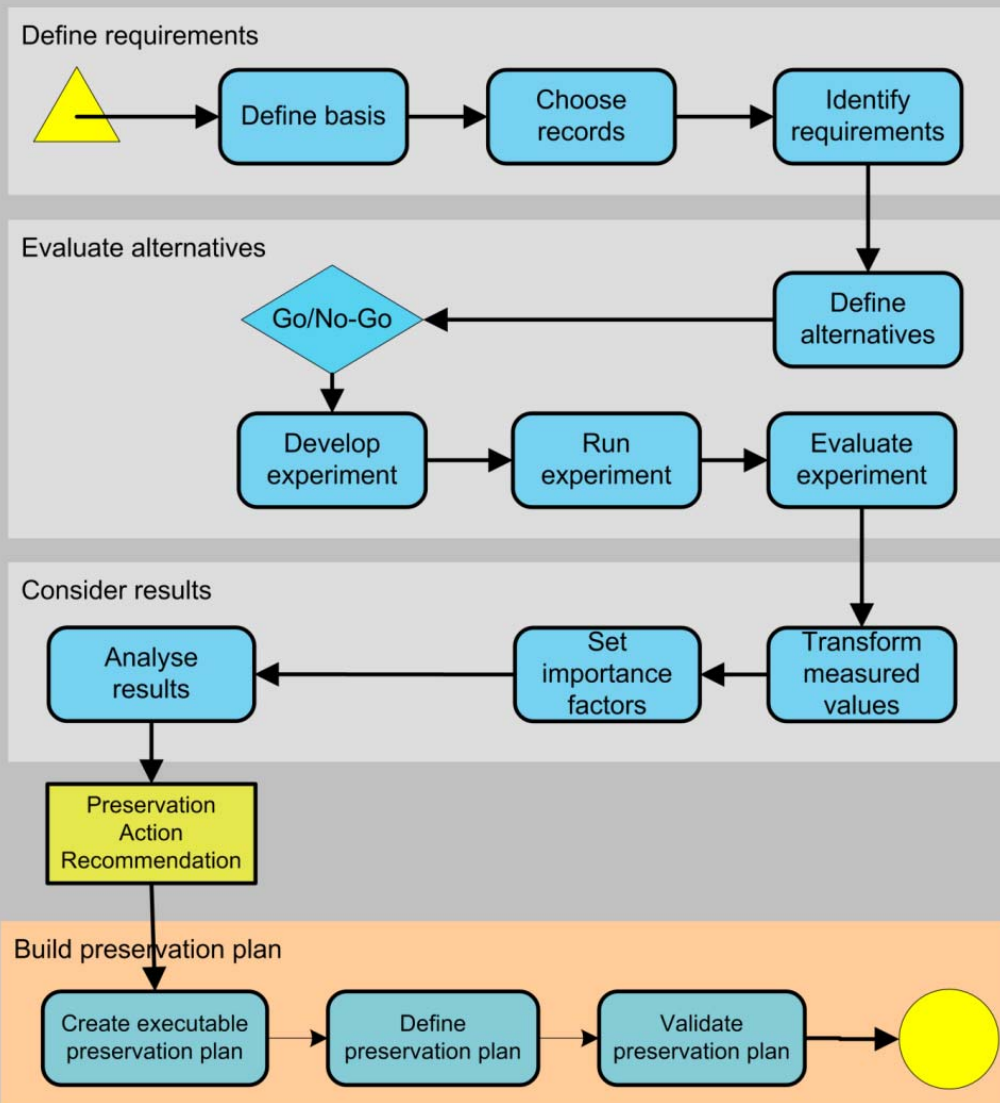
# Evaluating preservation strategies

- 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

# Planets Preservation Planning Workflow

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

## Preservation Planning workflow

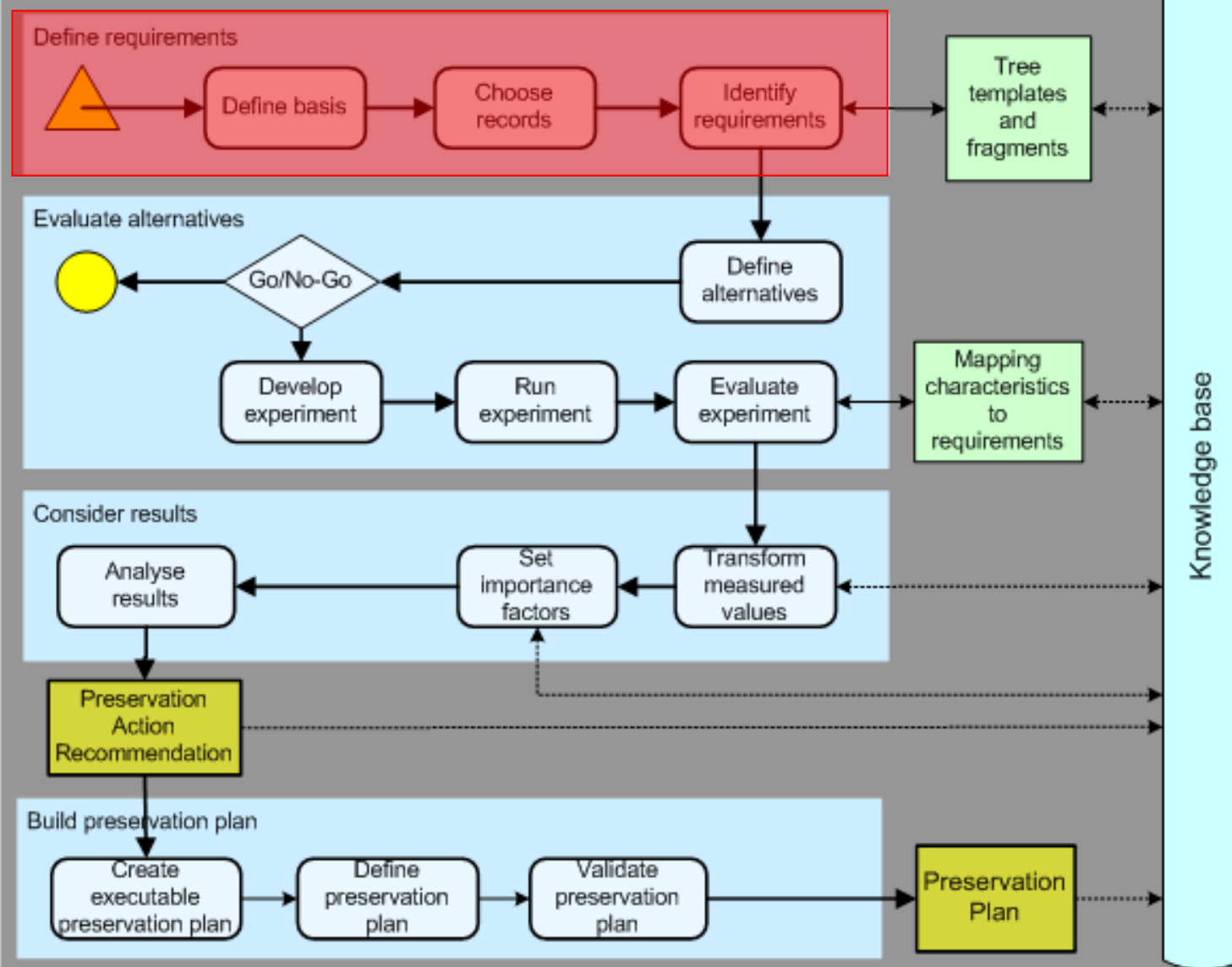


# Preservation Planning in Plato

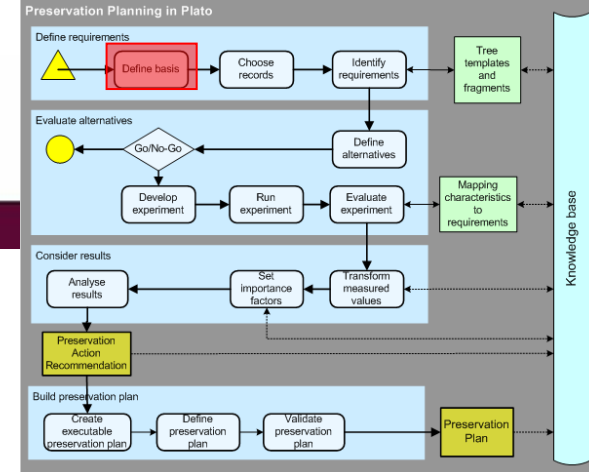
- Web based planning tool implementing the Planets preservation planning workflow
- Publicly available
- Automation of the planning process
  - Integration of registries and services for
    - File format identification
    - Preservation action (migration, emulation...)
    - Characterisation and comparison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

# PP Workflow

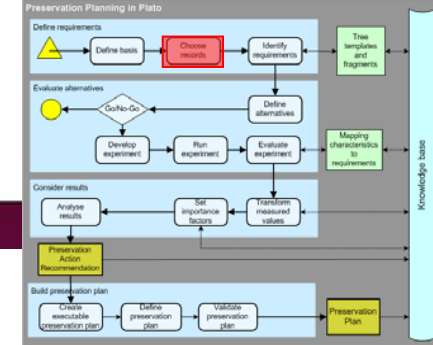
## Preservation Planning in Plato



# Define basis



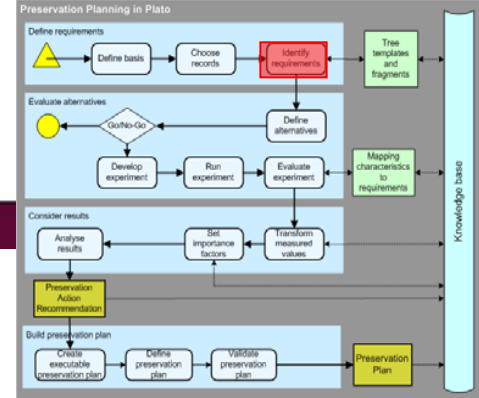
- Document basic assumptions and constraints
  - Types of objects
  - Purpose of planning
  - Mandates and designated community
  - Applying policies
  - Triggers that initiated the planning process



# 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

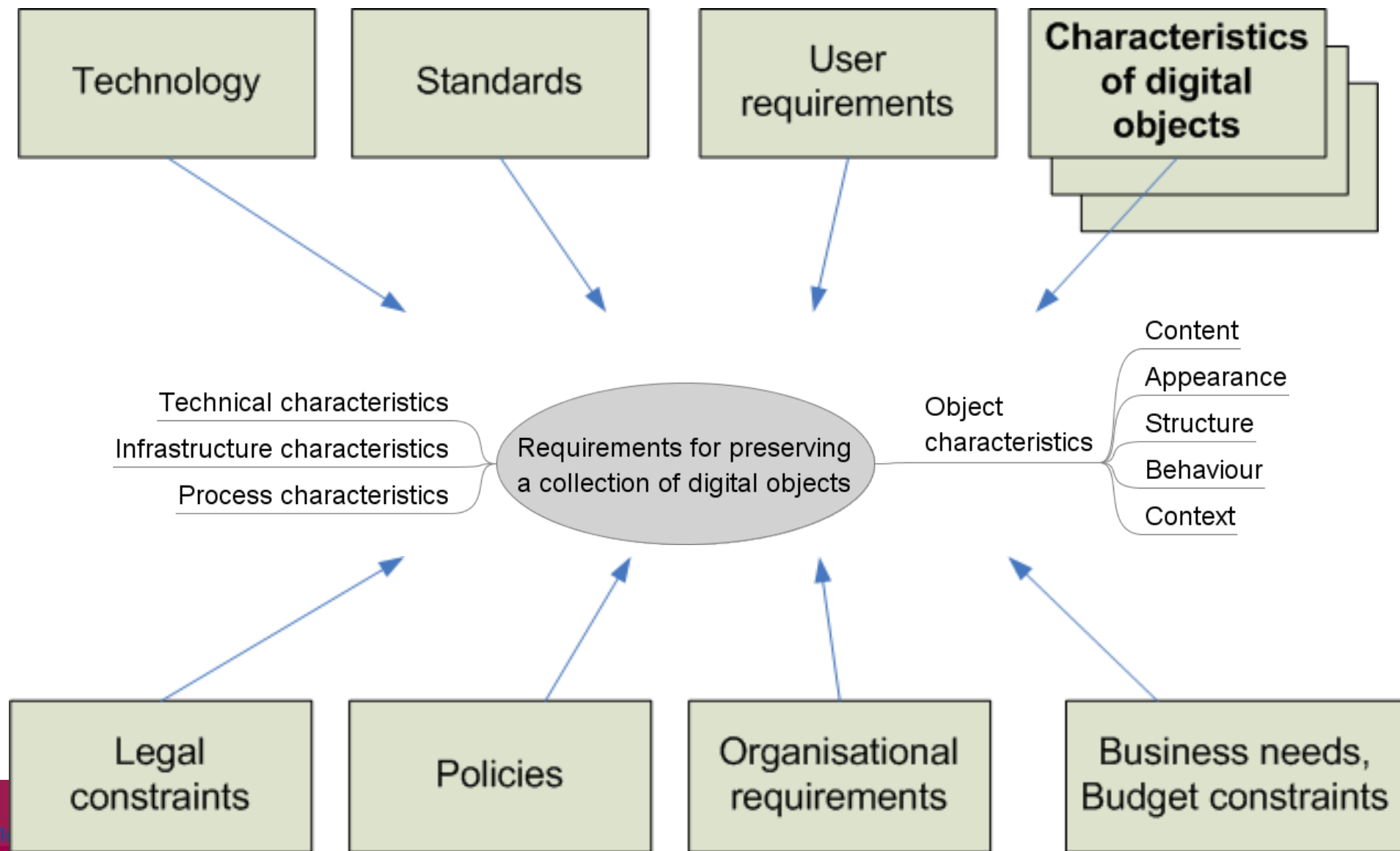




# Identify requirements

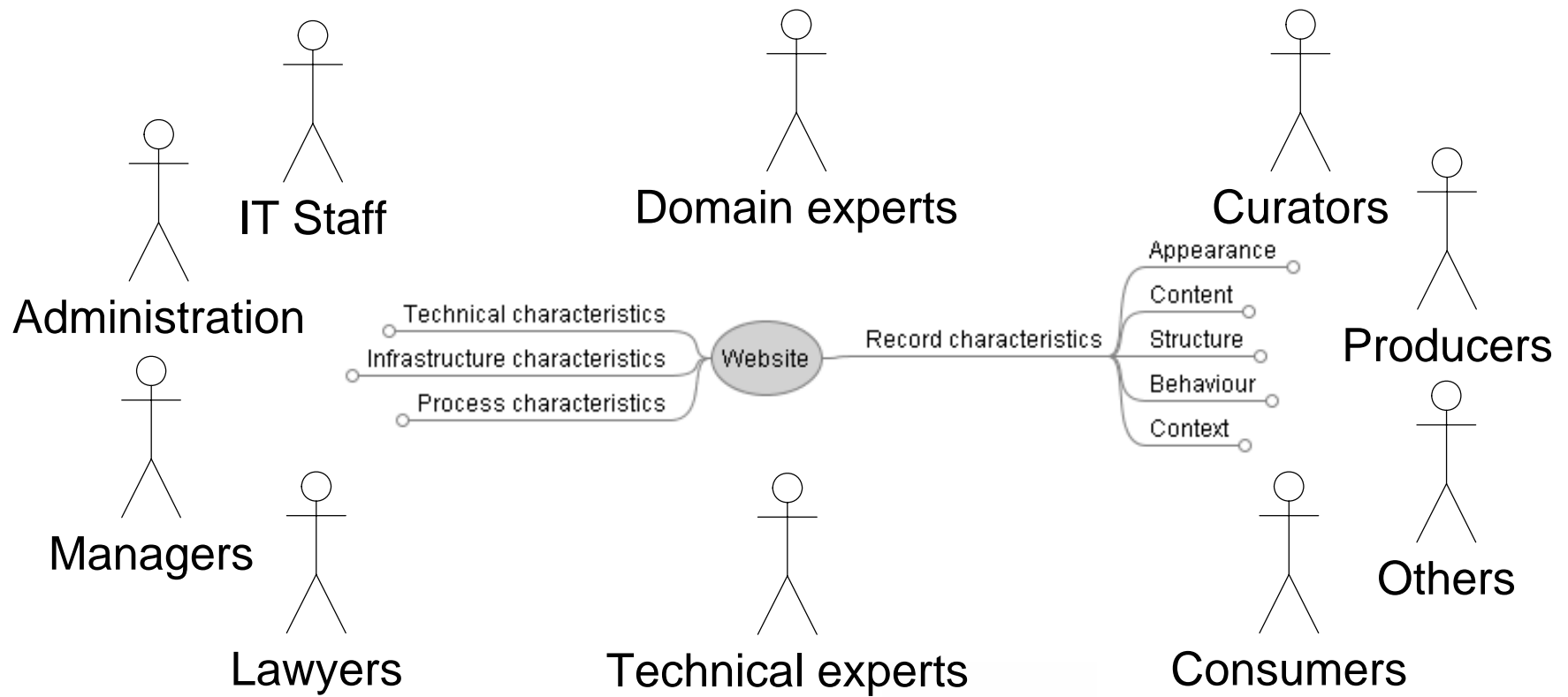
- Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain
- Usually four major groups:
  - object characteristics (content, metadata ...)
  - record characteristics (context, relations, ...)
  - process characteristics (scalability, error detection,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

# Influence Factors

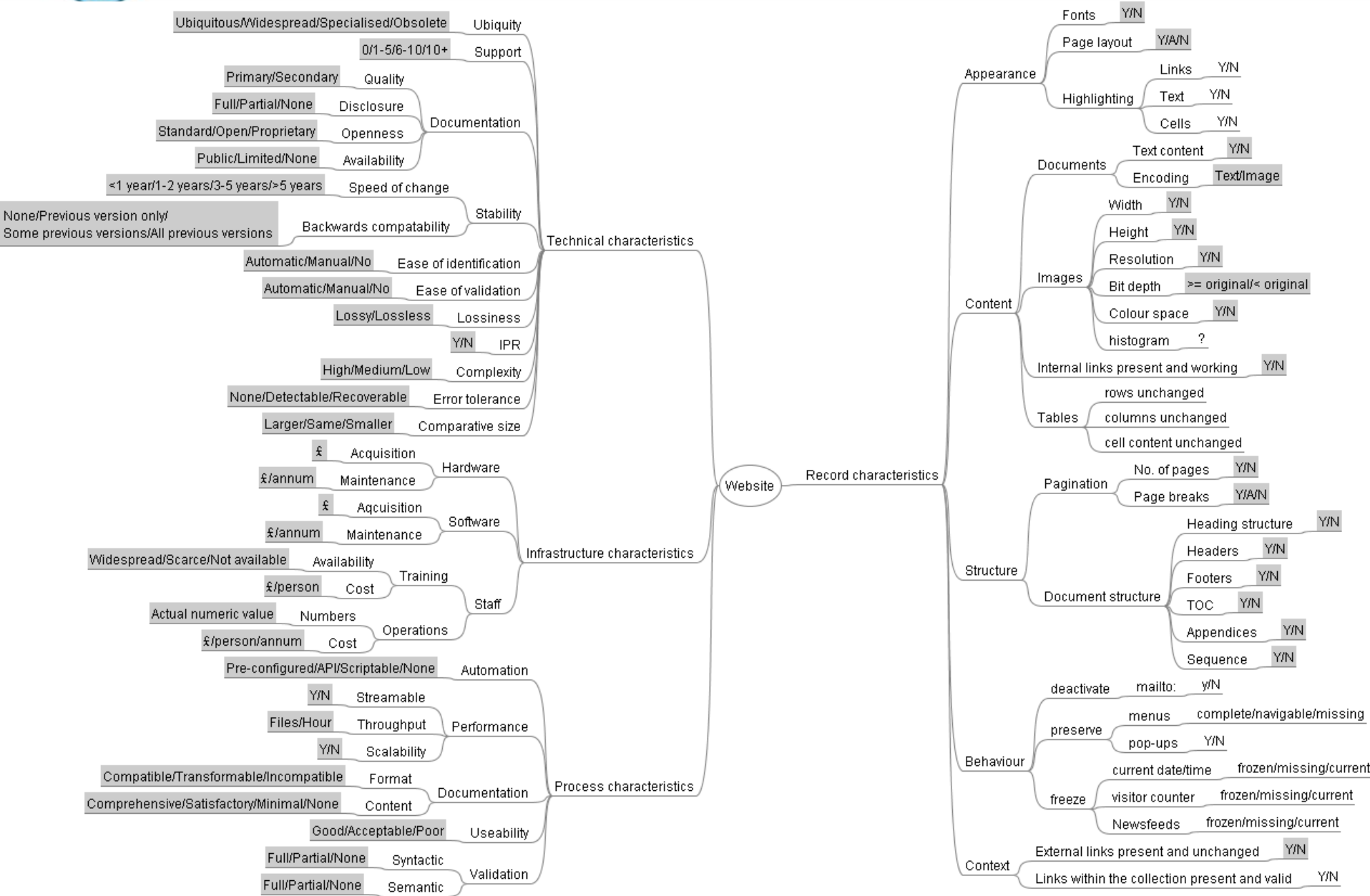


# Stakeholders

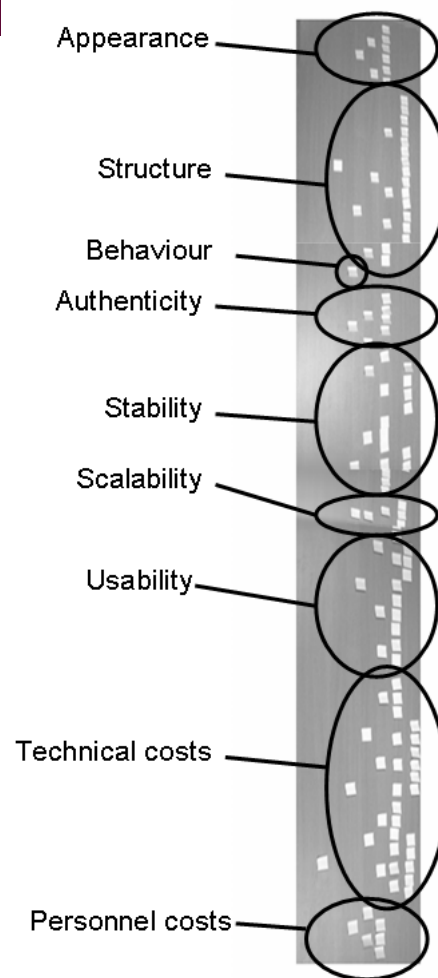
- Input needed from a wide range of persons, depending on the institutional context and the collection



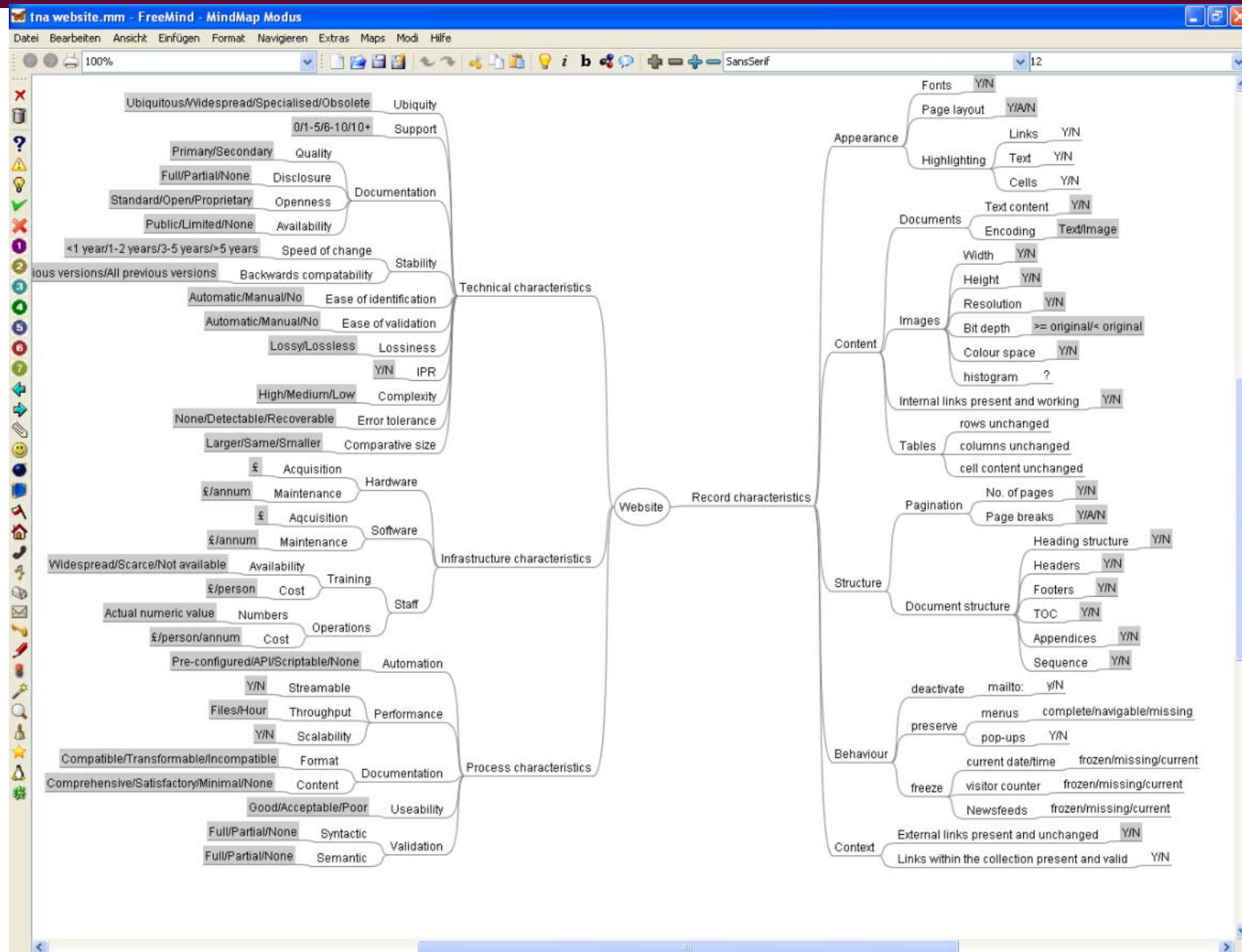
# An Objective Tree



# Analog...

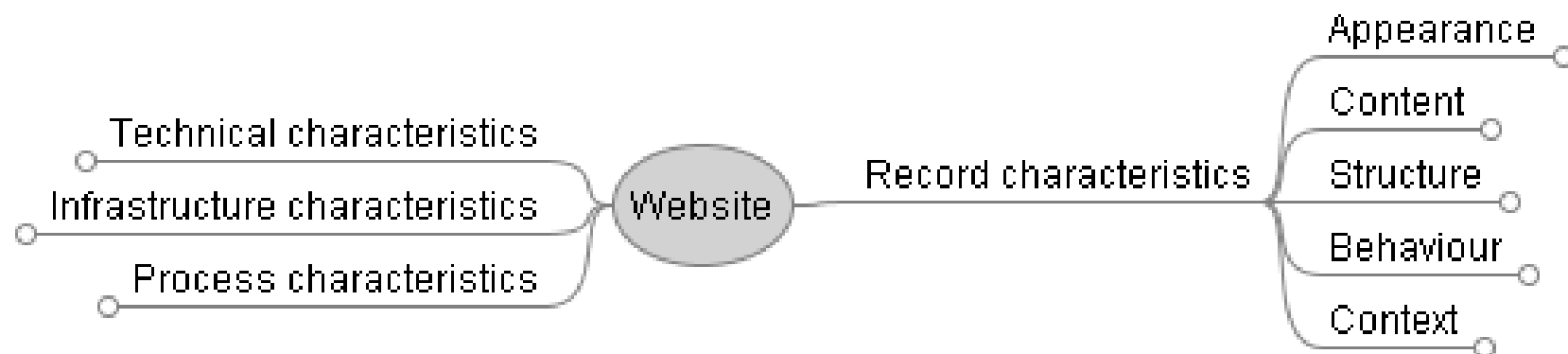


# ... or born-digital



## 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



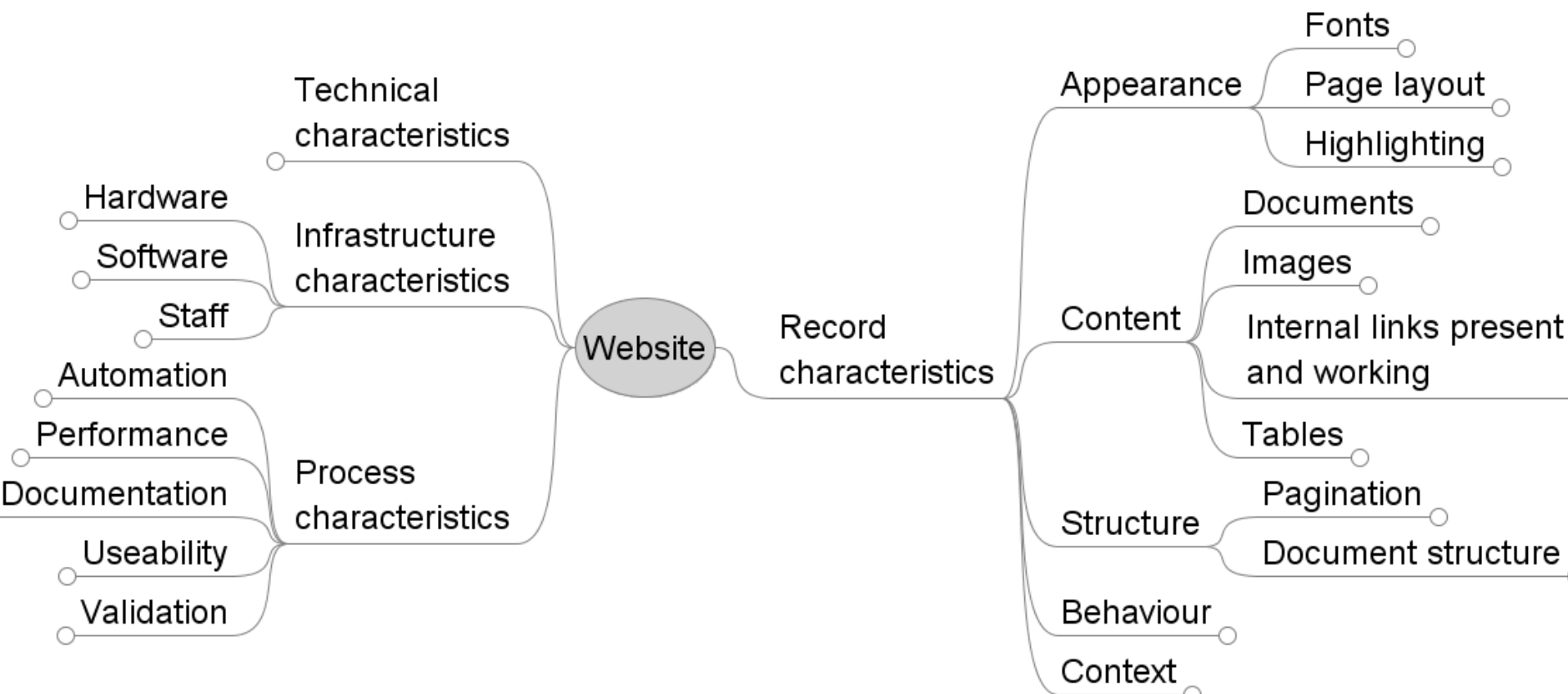


# Object characteristics

- Content
- Structure
- Appearance
- Behaviour
- Context



# A bit more detail...



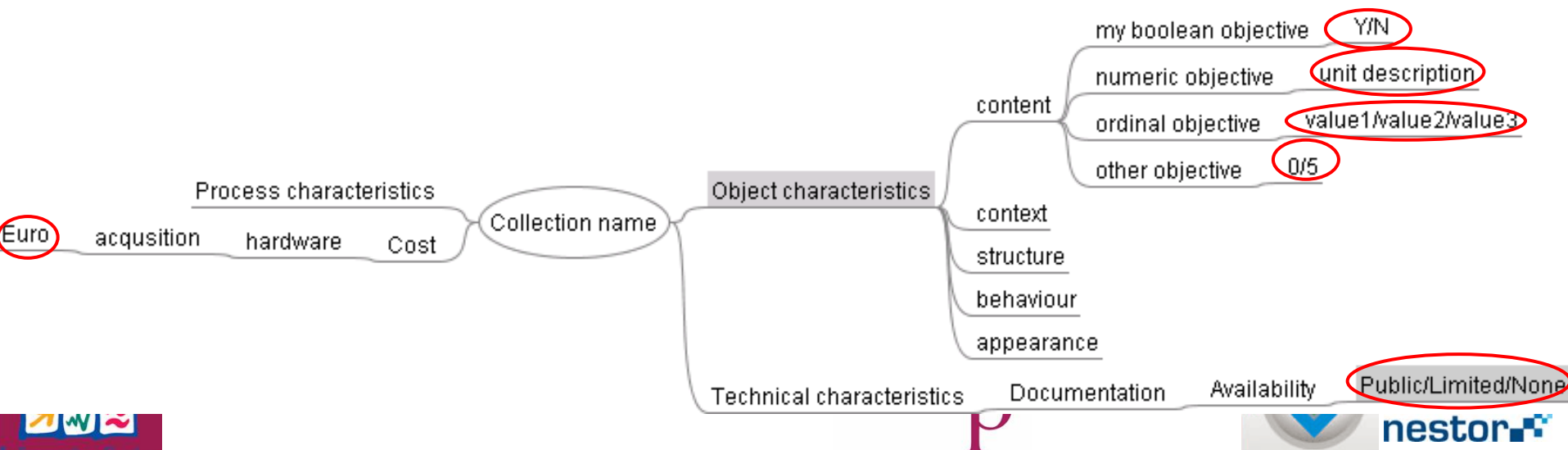
# Assign Measurable Units

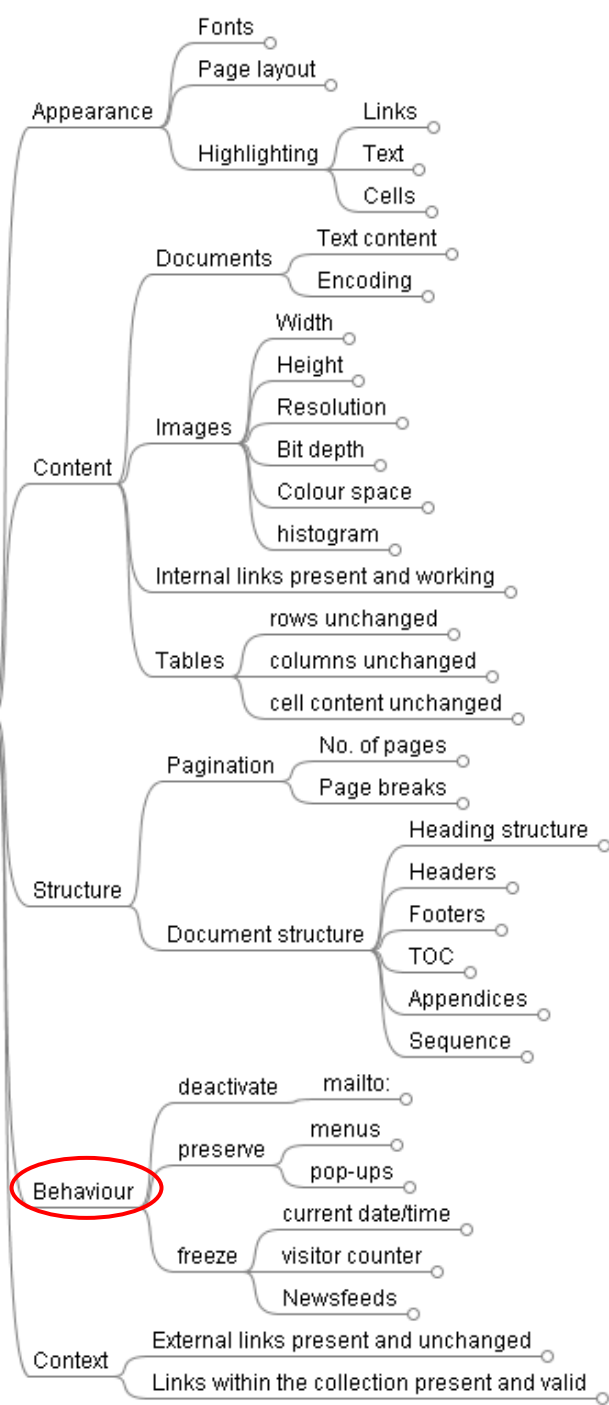
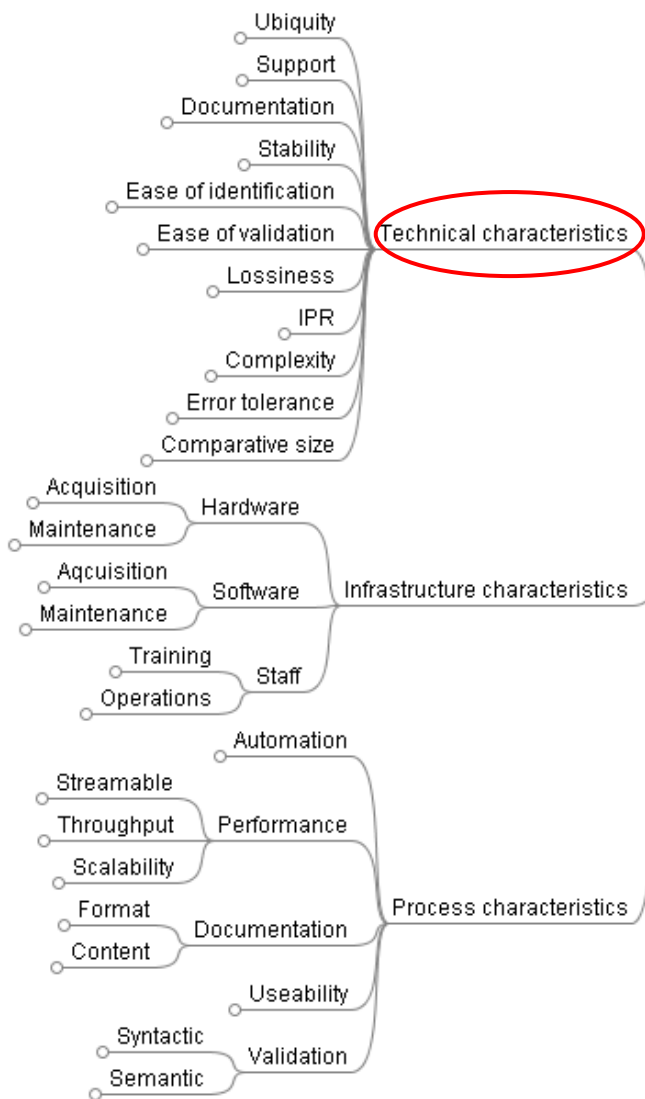
- Leaf criteria should be objectively measurable
  - Seconds per object
  - Euro per object
  - Bits of colour depth
- Subjective scales where necessary
  - Adoption of file format
  - Amount of (expected) support

➤ Quantitative results

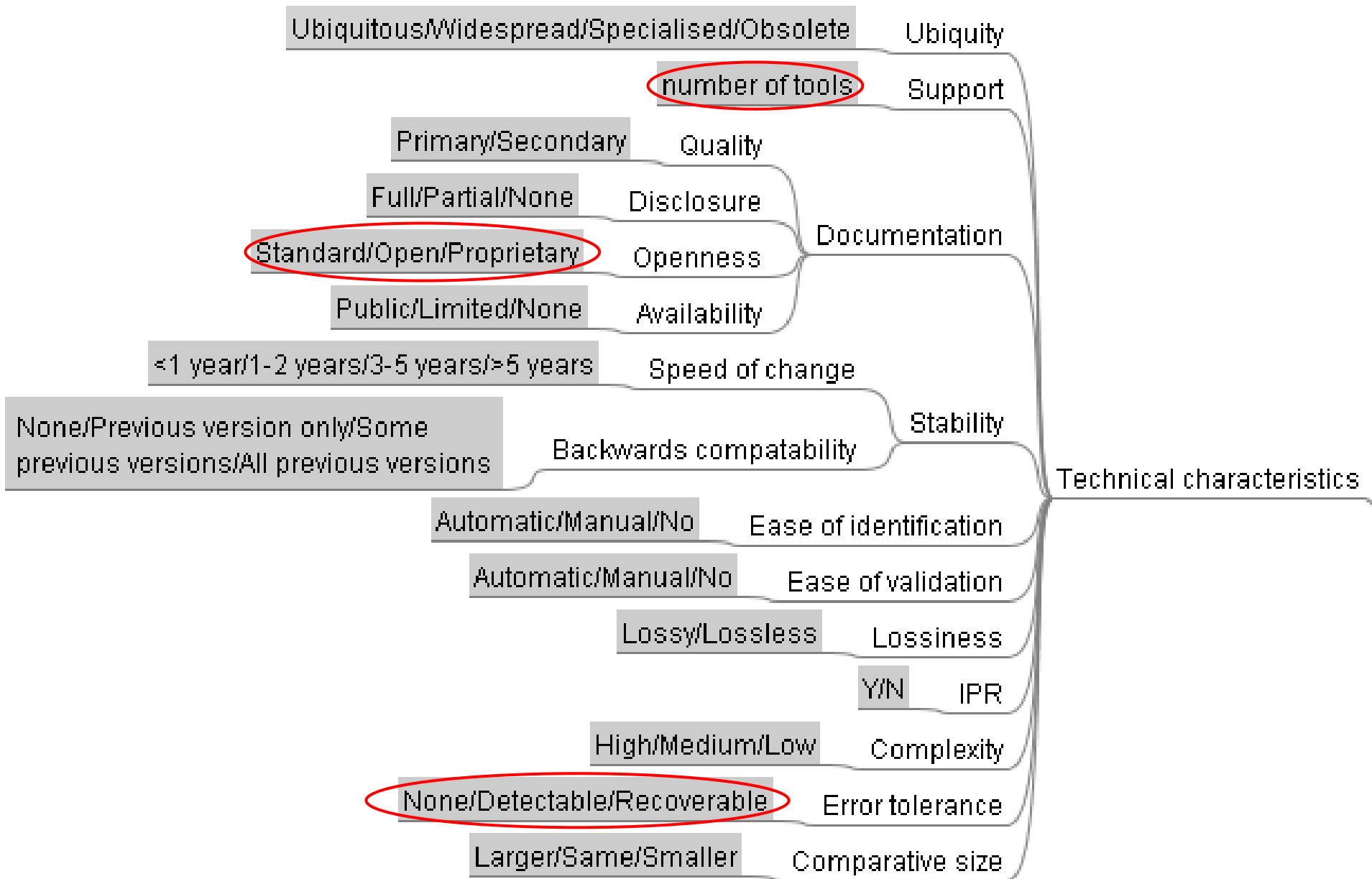
# Types of scales

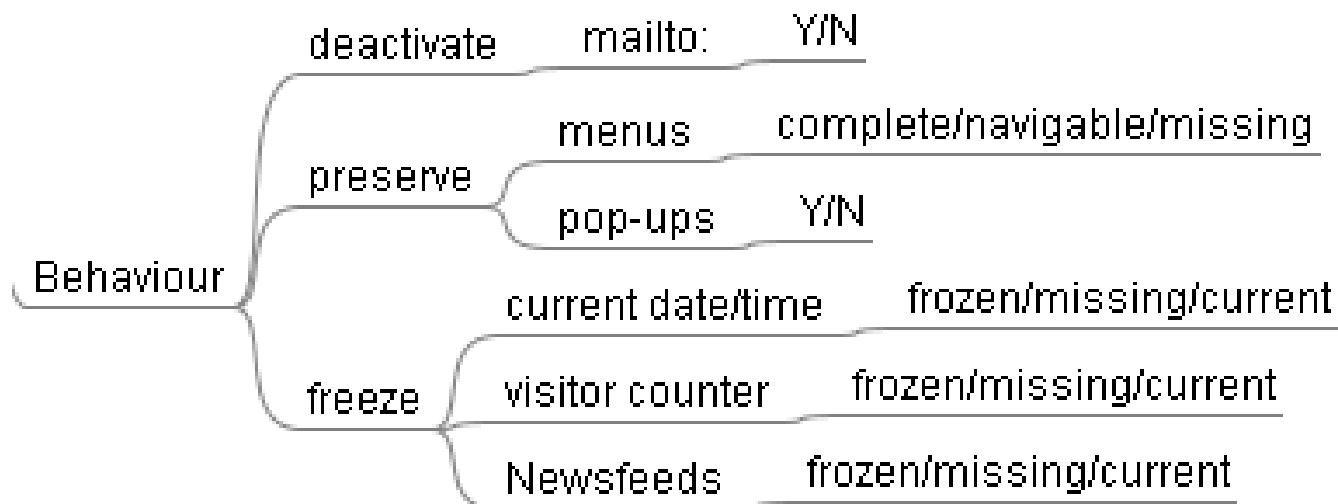
- Numeric (unit)
- Yes/No (Y/N)
- Yes/Acceptable/No (Y/A/N)
- Ordinal: define the possible values (good/bad/ugly)
- Subjective 0-to-5 (0/5)





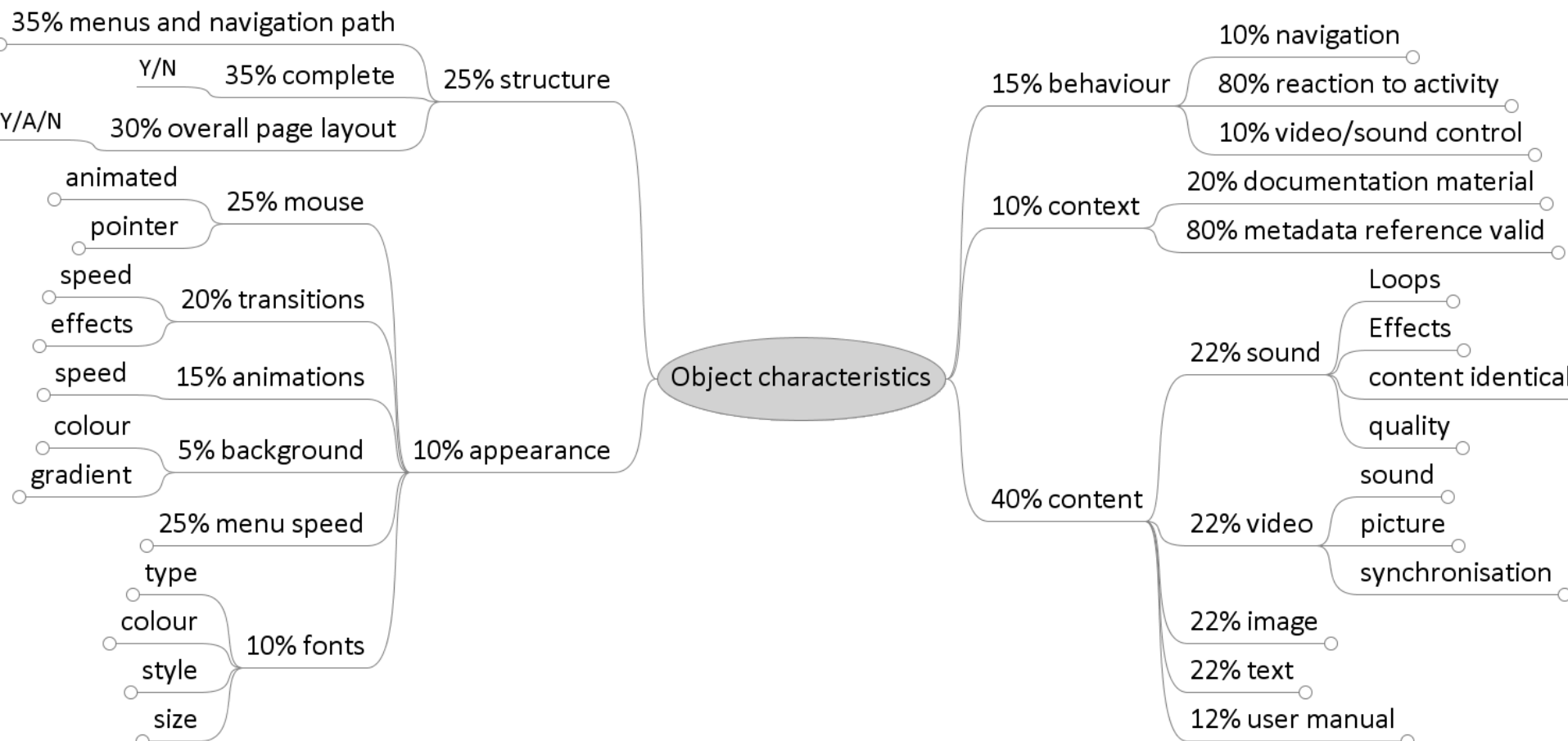
# File format characteristics





- 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



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

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



# Objective Tree



## PLANETS Preservation Planning Tool (Plato)

Institute of Software Technology and Interactive Systems

Project | Define Requirements | Evaluate Requirements | Consider Results |

Loaded project: PP4 workshop - The National Archive

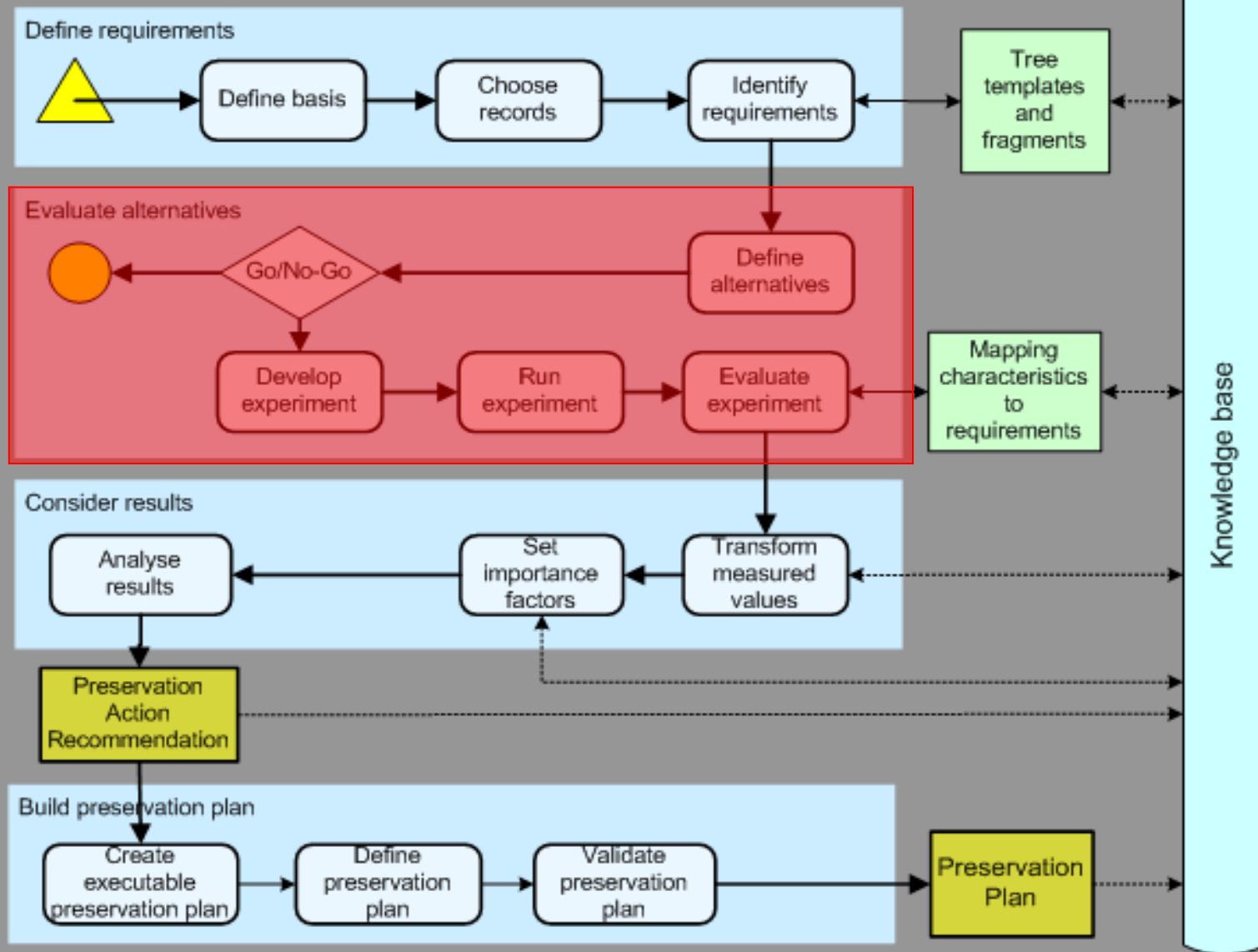
### Identify Requirements

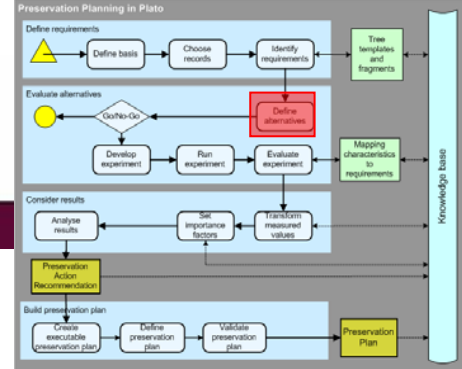
[Expand All](#) | [Collapse All](#)  
[Website](#)

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

# PP Workflow

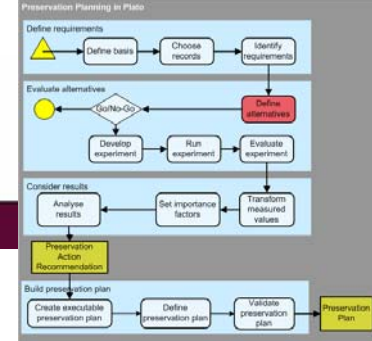
## Preservation Planning in Plato





## 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



# 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



Show Preservation Services

Planets Service Registry



Show Preservation Services

CRiB Service Registry



Show Preservation Services

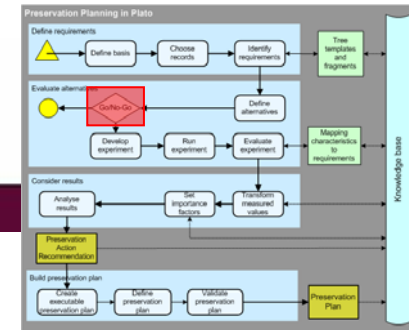
	Preservation Action	Target Format	Info
<input type="checkbox"/>	JPG > BMP	Windows Bitmap, version 3.0	JPG>BMP
<input checked="" type="checkbox"/>	JPG > TIF	Tagged Image File Format, version 3	JPG>BMP>TIF
<input type="checkbox"/>	JPG > TIF #2	Tagged Image File Format, version 3	JPG>TIF
<input checked="" type="checkbox"/>	JPG > TIF_2	Tagged Image File Format, version 3	JPG>TIF_2
<input type="checkbox"/>	JPG > PNG	Portable Network Graphics, version 1.0	JPG>PNG
<input type="checkbox"/>	JPG > JP2	JPEG 2000	JPG>JP2

Create alternatives for selected services



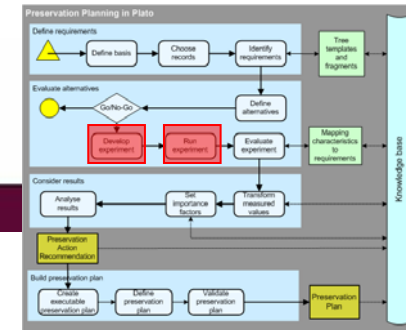
# Specify resources

- 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 readressed or not?

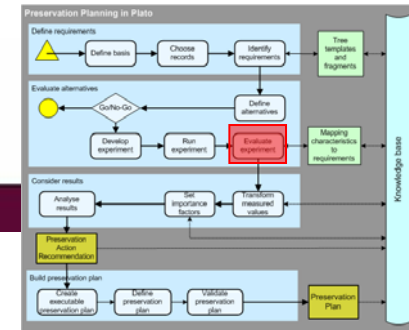


# Develop and run experiment

## ➤ 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



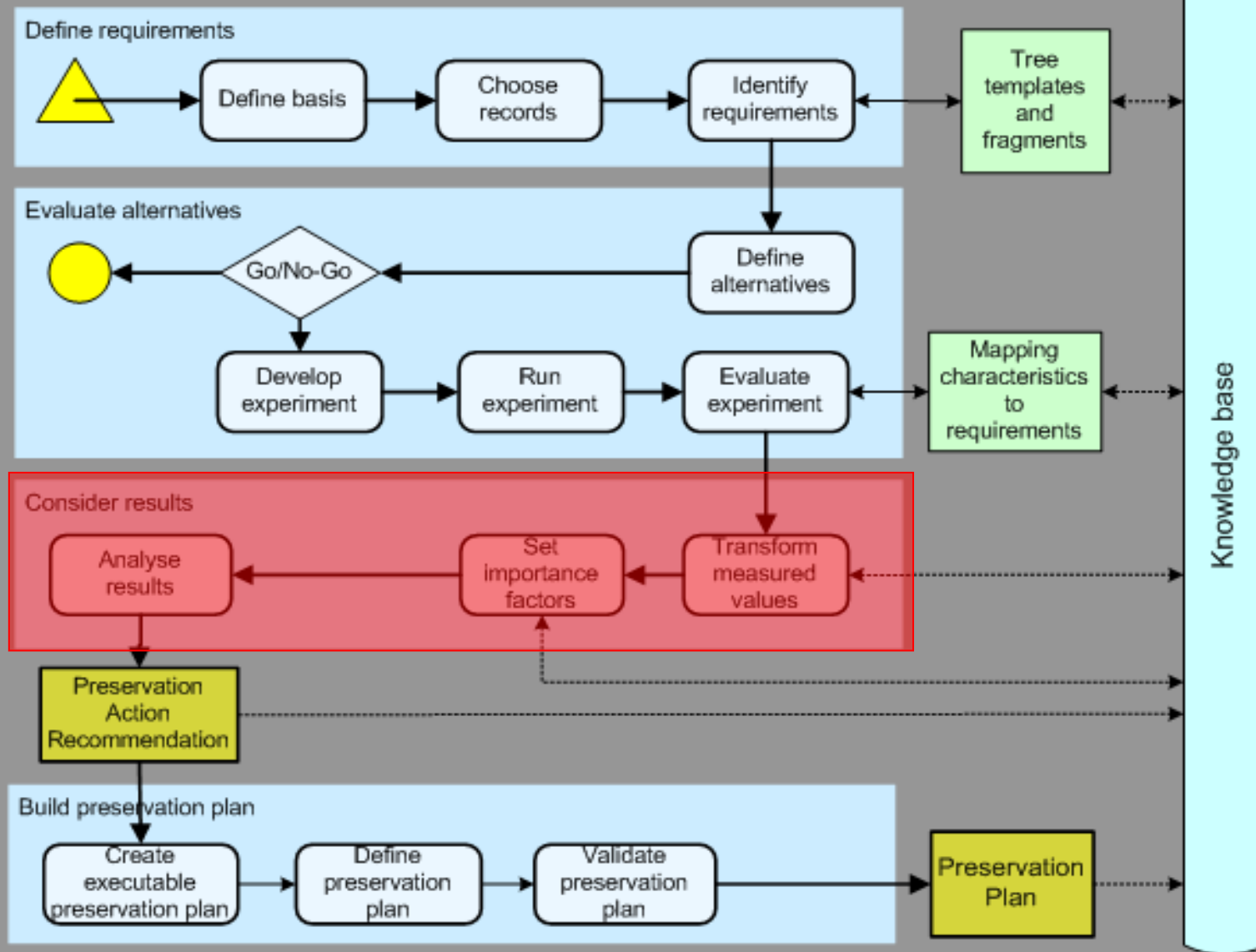
# Evaluate experiment

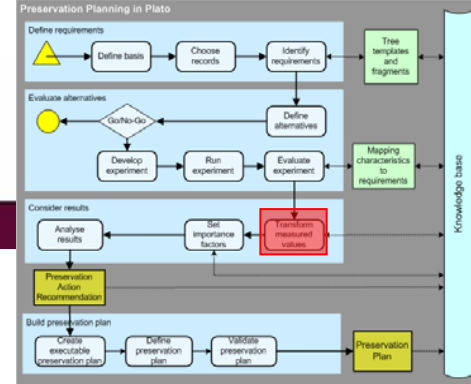
- Evaluate the outcome of each alternative for each leaf of the objective tree
- The evaluation will identify
  - Need for repeating the process
  - Unexpected (or undesired) results
- Includes both technical and intellectual aspects
- Evaluation may include comparing the results of more than one experiment/evaluation.



# PP Workflow

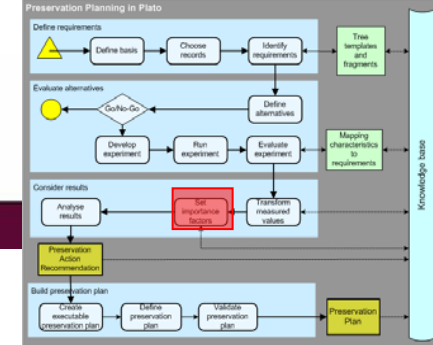
## Preservation Planning in Plato





# 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



# 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

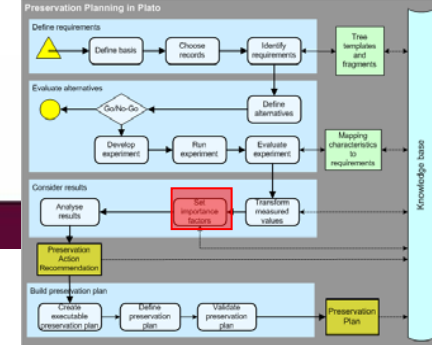
Balance weights automatically ☒

[Expand All](#) | [Collapse All](#)

**Object characteristics**

Focus	Name	Weight	Lock	Total weight
	▼ Object characteristics	0		1
X	▶ behaviour	0	<input checked="" type="checkbox"/>	0.15
X	▶ structure	0	<input checked="" type="checkbox"/>	0.25
X	▶ context	0	<input type="checkbox"/>	0.1
X	▶ appearance	0	<input type="checkbox"/>	0.1
X	▶ content	0	<input checked="" type="checkbox"/>	0.4

Save Proceed



# 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

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

Analyse results

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

Conclusion

Recommendation

Recommendation:

Reasoning:

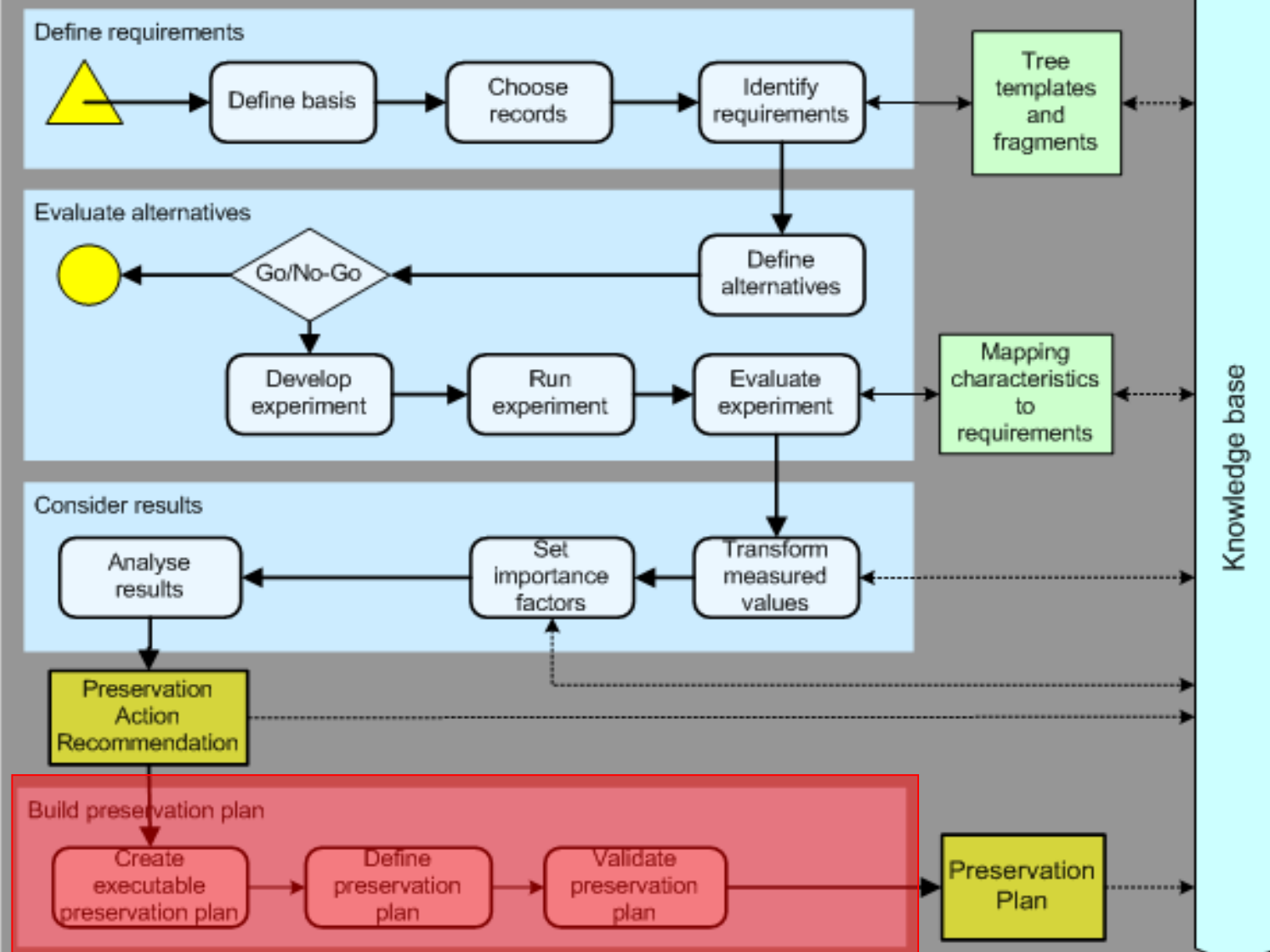


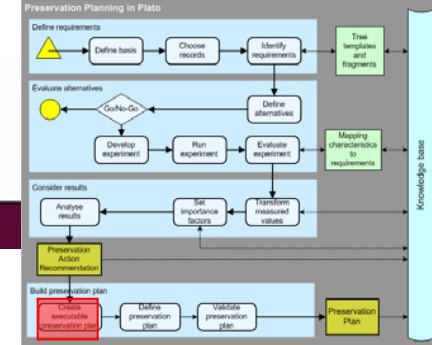
Effects of applying this strategy:



# PP Workflow

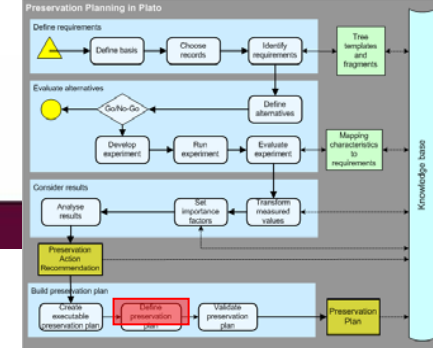
## Preservation Planning in Plato





# 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...



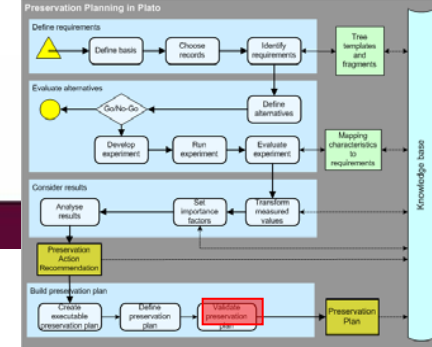
# Define preservation plan

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

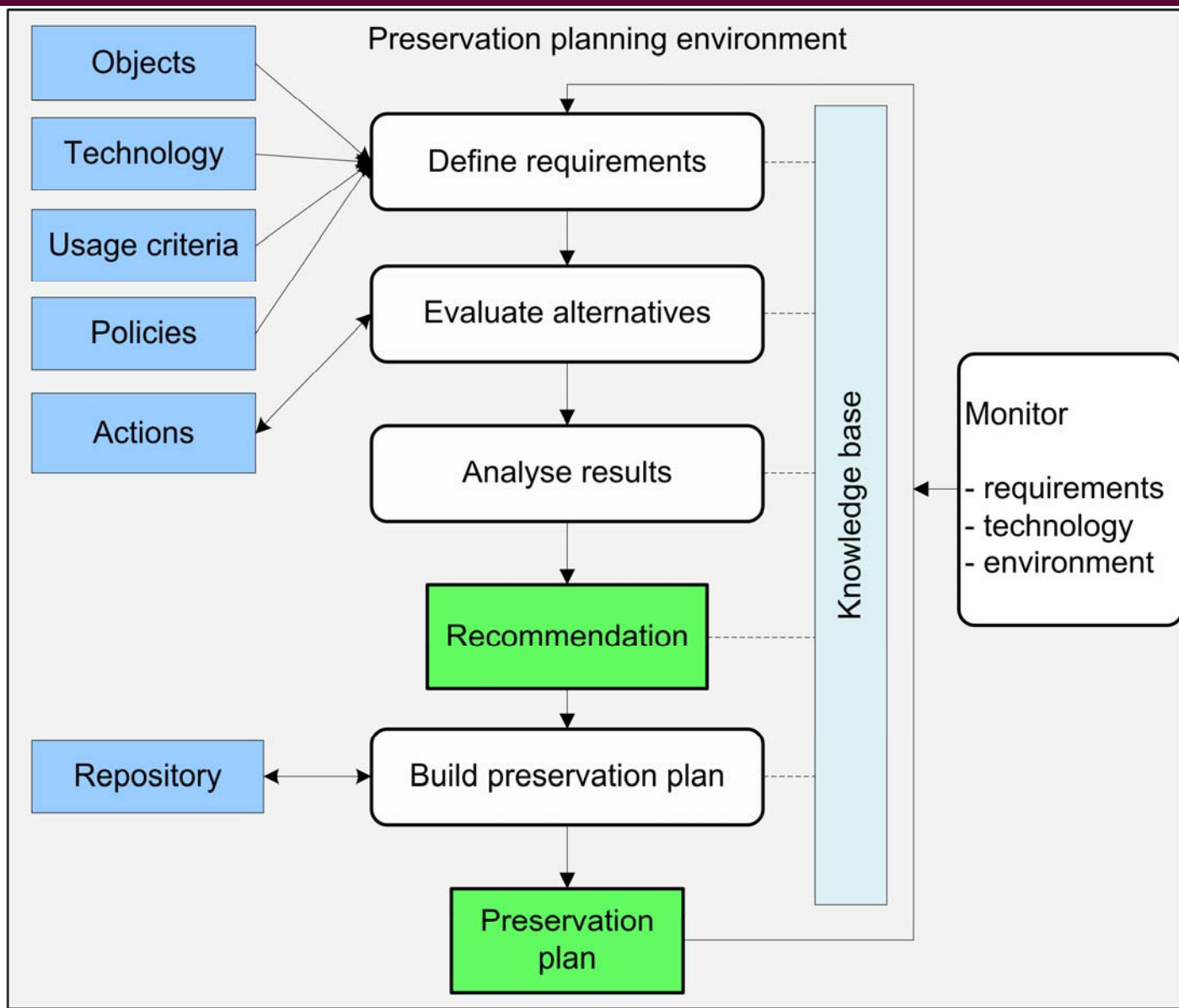


# Validate preservation plan

- Validate all elements of the plan
- Check for completeness
- Formally approve the plan and put it into action
- Continuous review and monitoring is necessary!



# Summary



# Questions?

[kulovits@ifs.tuwien.ac.at](mailto:kulovits@ifs.tuwien.ac.at)

[www.ifs.tuwien.ac.at/dp/plato](http://www.ifs.tuwien.ac.at/dp/plato)

[www.planets-project.eu](http://www.planets-project.eu)

- ☐ Preservation Planning
  - Evaluation of potential actions
- ☐ The Planets Preservation Planning Workflow
  - Underlying methodology
  - Workflow walkthrough
  - The planning tool Plato
- ☐ Break-out session: Requirements definition
  - Groups
  - Scenarios
  - Schedule
- Demonstration

# Practice time!

- Part 1: Think...
  - ...about
    - Your collection, your objects
    - The designated community, organisation...
    - Requirements
  - Document that shortly to have a common basis
  - Create a draft objective tree
- Part 2: Draw...
  - Refine the tree structure and complete it
  - Think about the significant properties of the objects in the specific scenario
  - Assign measurable units
  - Set high-level importance factors

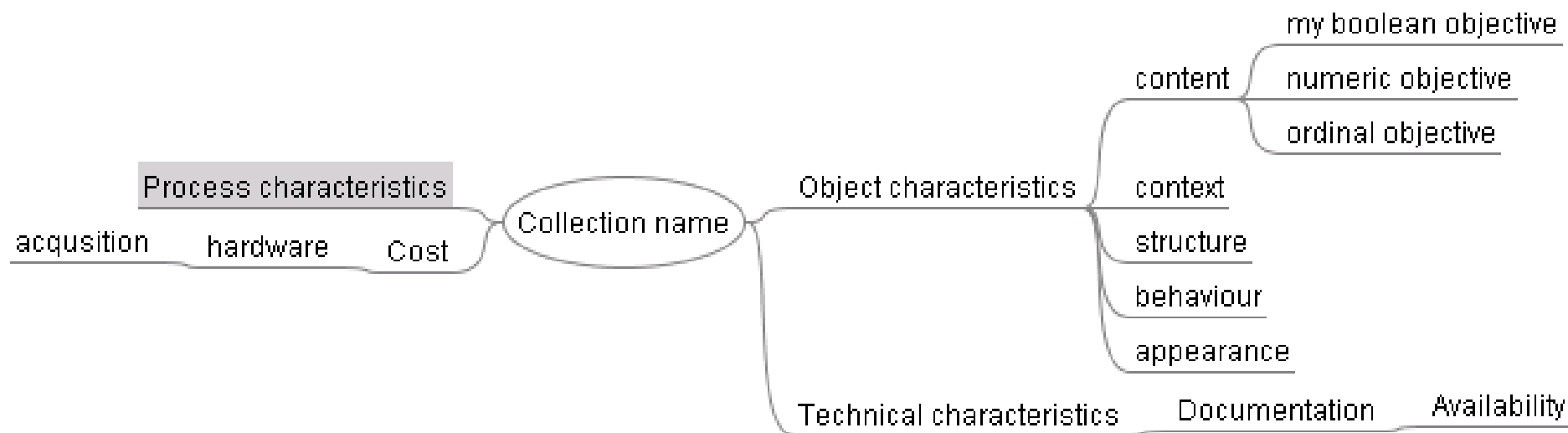
- Form the same groups as in the previous preservation planning exercise
- Use the results from the previous exercise as a starting point

## How to construct the tree

- With the open-source mind-mapping tool Freemind
  - Java required
  - Freemind is installed in 20 seconds
- With post-it notes
  - Please recreate the tree in FreeMind at the end (for the discussion session)

# Tree template

- This is *one* way to start
- Add (and remove) criteria as you like
- Adapt hierarchy as you deem appropriate





# Questions?

[kulovits@ifs.tuwien.ac.at](mailto:kulovits@ifs.tuwien.ac.at)

[www.ifs.tuwien.ac.at/dp/plato](http://www.ifs.tuwien.ac.at/dp/plato)

[www.planets-project.eu](http://www.planets-project.eu)

# Practice time!

- Part 1: Think...
  - ...about
    - Your collection, your objects
    - The designated community, organisation...
    - Requirements
  - Document that shortly to have a common basis
  - Create a draft objective tree
- Part 2: Draw...
  - Refine the tree structure and complete it
  - Think about the significant properties of the objects in the specific scenario
  - Assign measurable units
  - Set high-level importance factors

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.