

Digital Preservation Planning
July 29 2008, London, UK



The Planets Preservation Planning workflow and the planning tool Plato

organized in cooperation with DPC

Christoph Becker

Vienna University of Technology

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



Outline

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



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

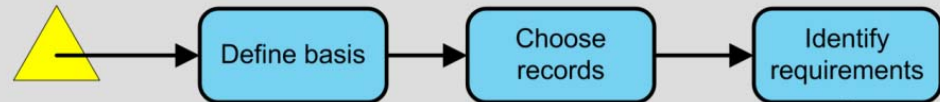


Planets Preservation Planning Workflow

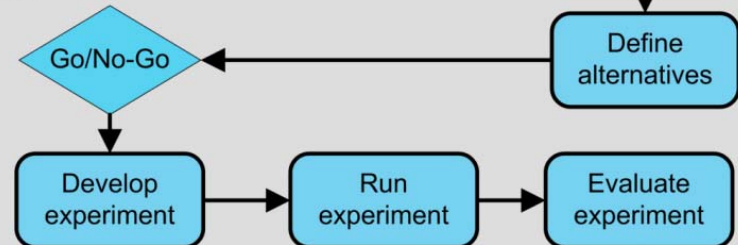
- ❑ Define requirements
- ❑ Evaluate potential actions
- ❑ Analyse results
- ❑ Build a preservation plan

Preservation Planning workflow

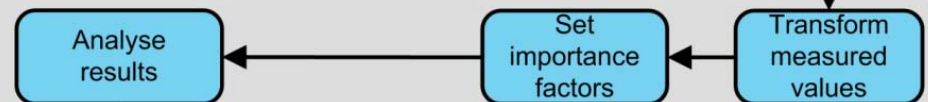
Define requirements



Evaluate alternatives

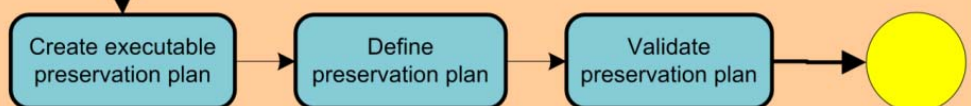


Consider results



Preservation
Action
Recommendation

Build preservation plan

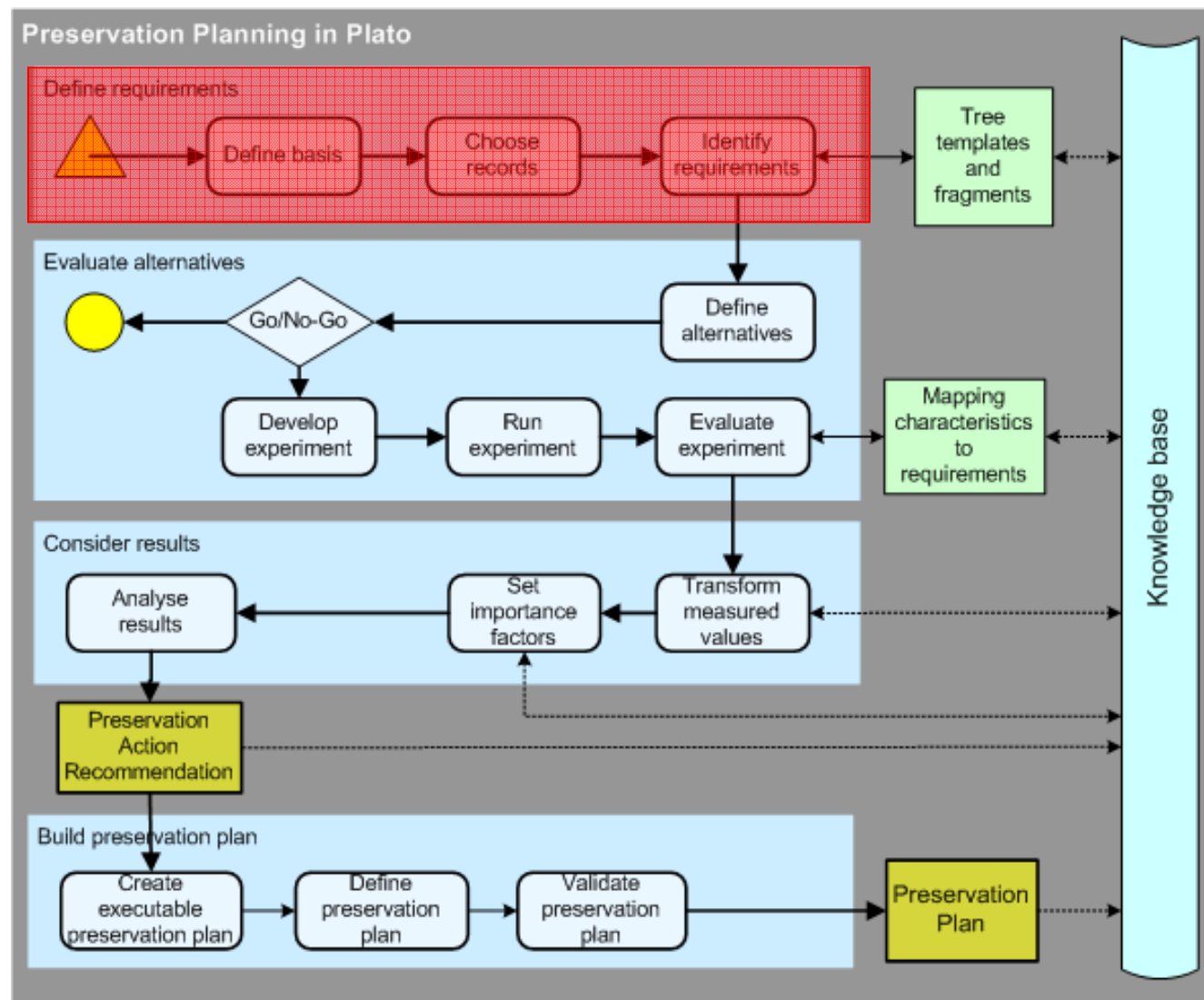


Preservation Planning in Plato

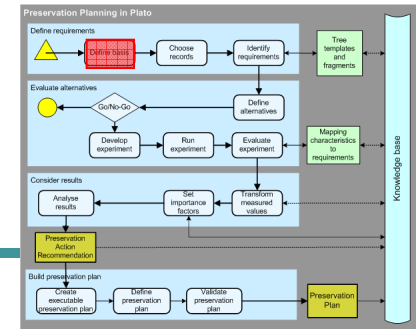
- ❑ Web based planning tool implementing the Planets preservation planning workflow
- ❑ Publicly available
- ❑ Automation of the planning process
 - Integration of registries and services for
 - File format identification
 - Preservation action (migration, emulation...)
 - Characterisation and comparison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



PP Workflow

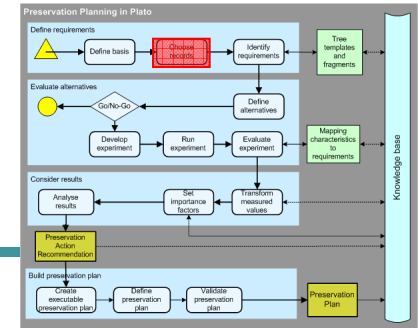


Define basis



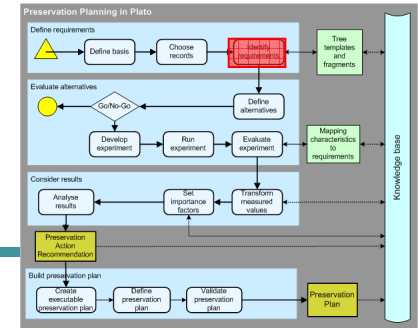
- Document basic assumptions and constraints
 - Types of objects
 - Purpose of planning
 - Mandates and designated community
 - Applying policies
 - Triggers that initiated the planning process

Choose sample objects/records



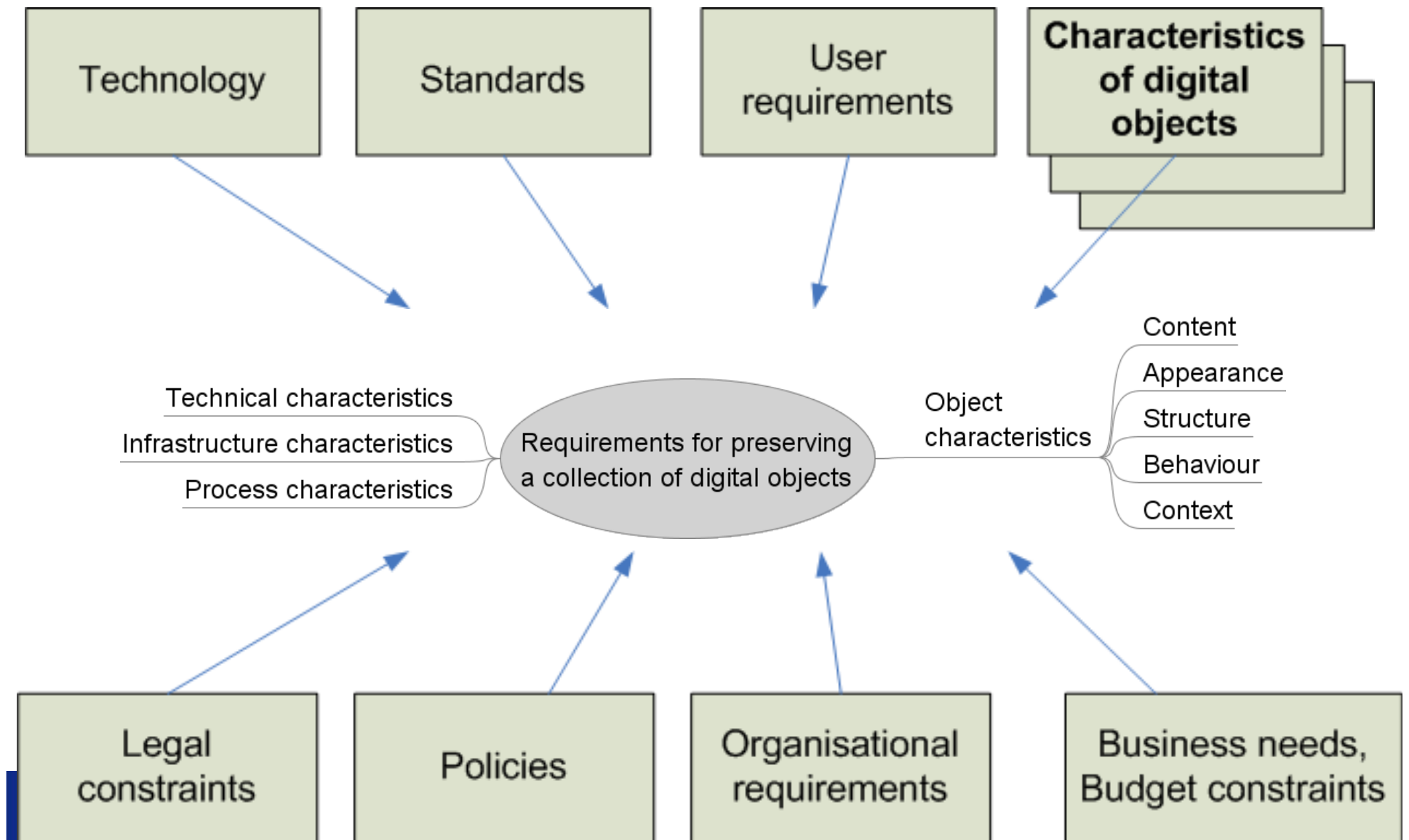
- Define the set of objects that are the subject of preservation planning
 - Size of the collection
 - Growth rate
 - Object format
 - ...
- Specify representative sample objects that cover the variety of significant properties and technical characteristics

Identify requirements



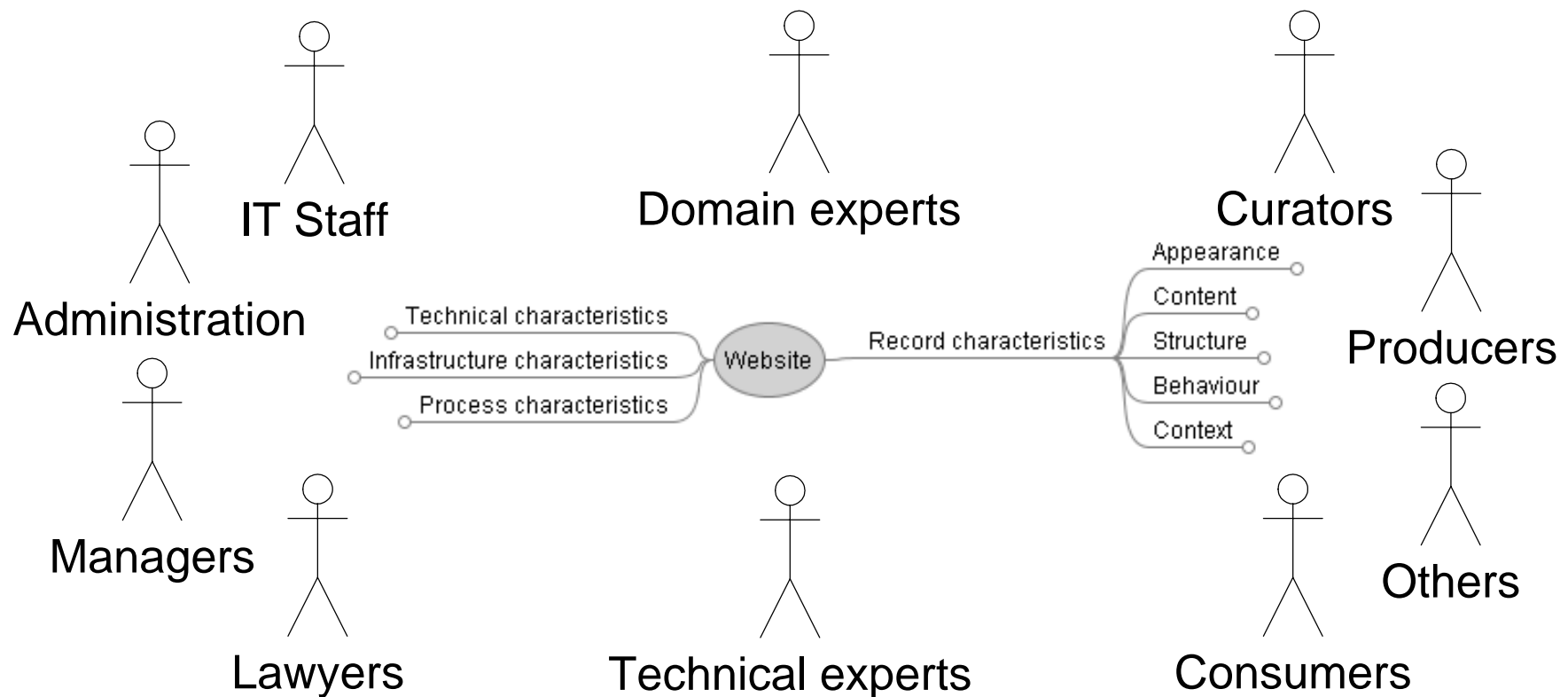
- Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain
- Usually four major groups:
 - object characteristics (content, metadata ...)
 - record characteristics (context, relations, ...)
 - process characteristics (scalability, error detection,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

Requirements and Influence Factors

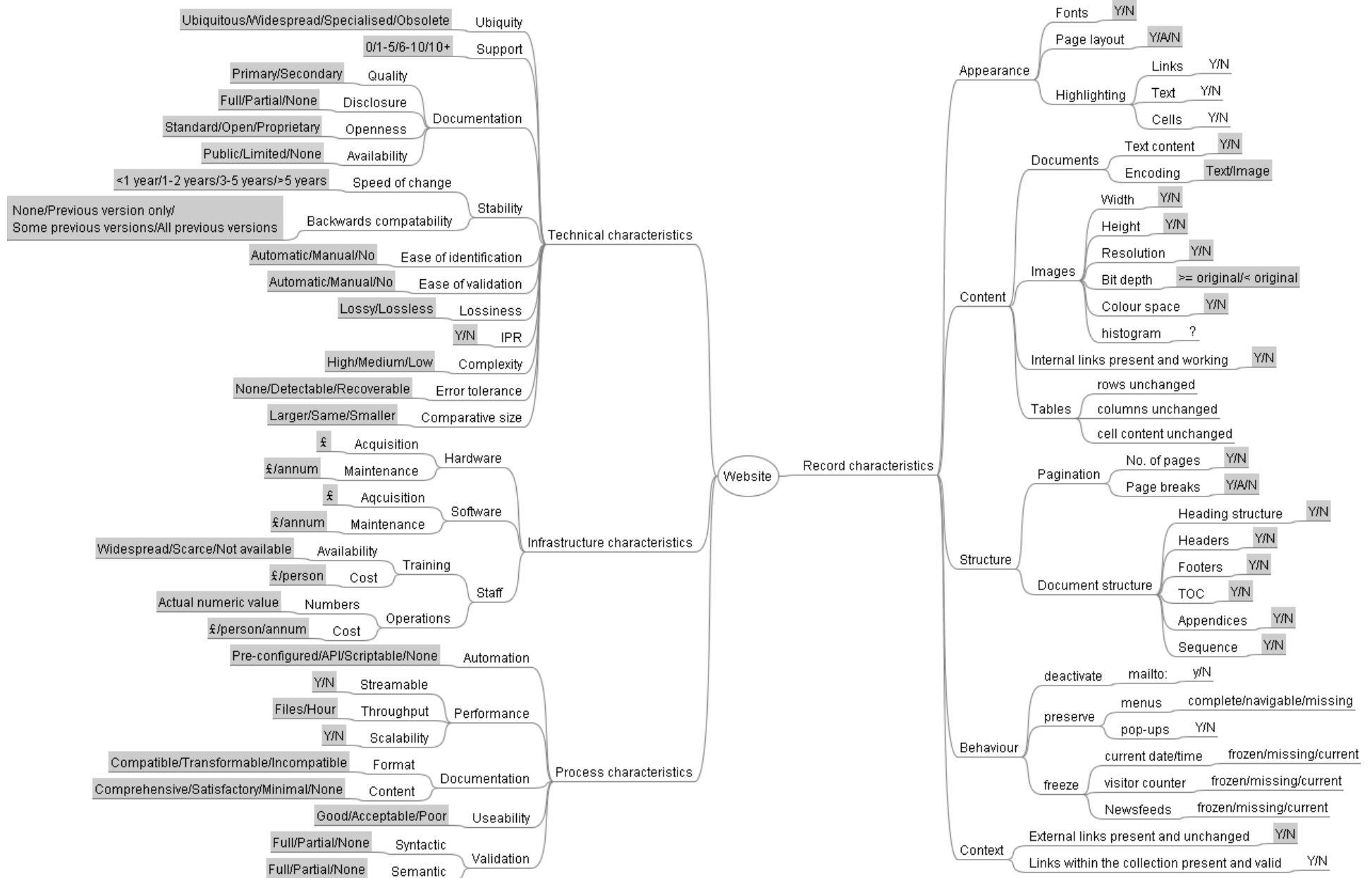


Stakeholders

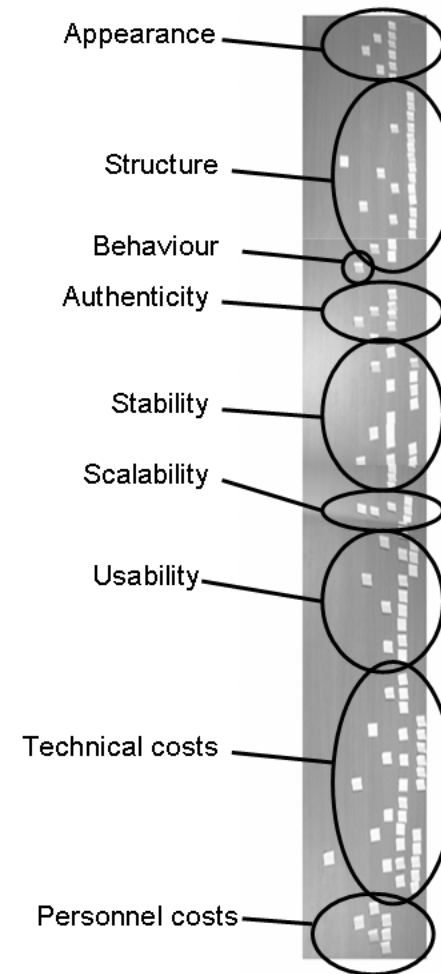
- Input needed from a wide range of persons, depending on the institutional context and the collection



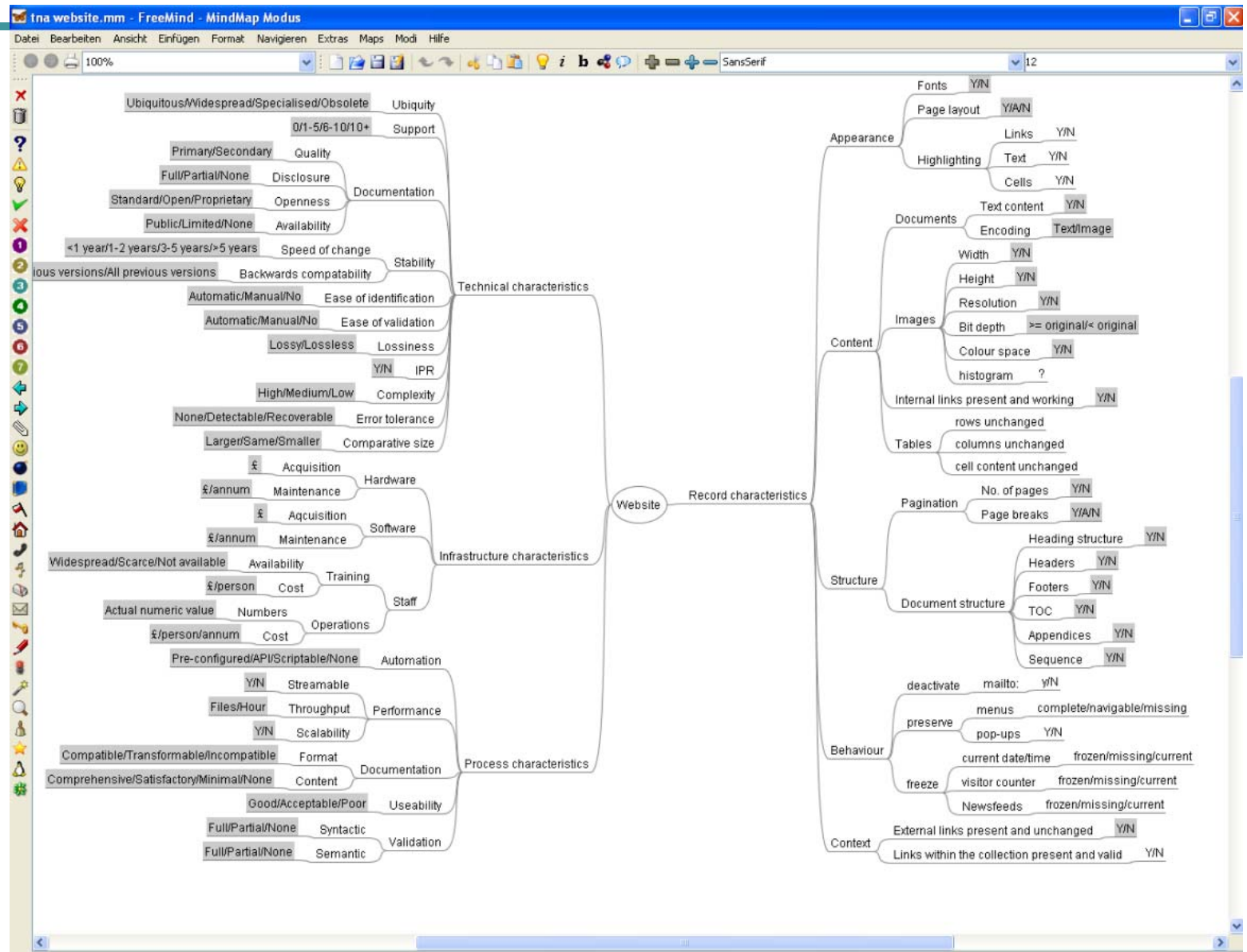
An Objective Tree



Analog...

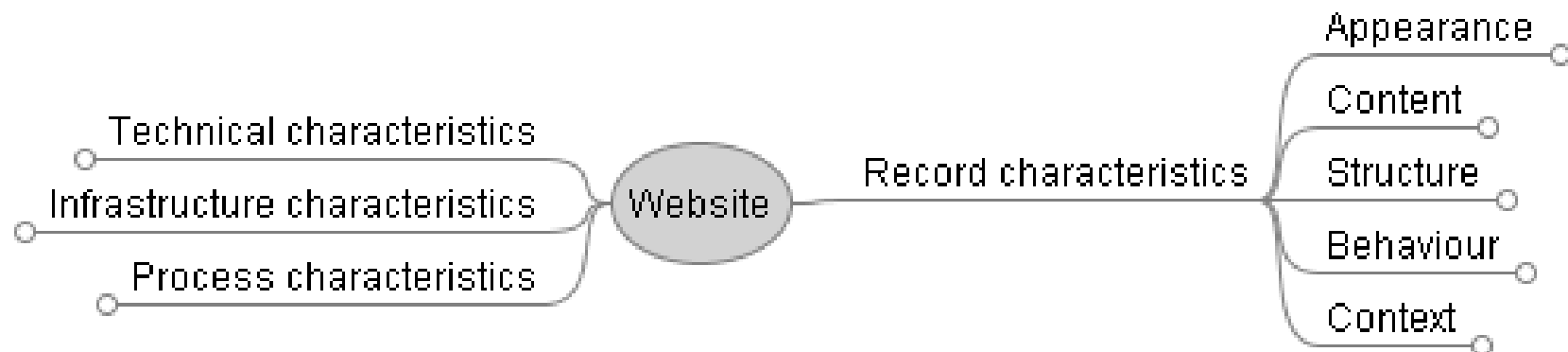


... or born-digital



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

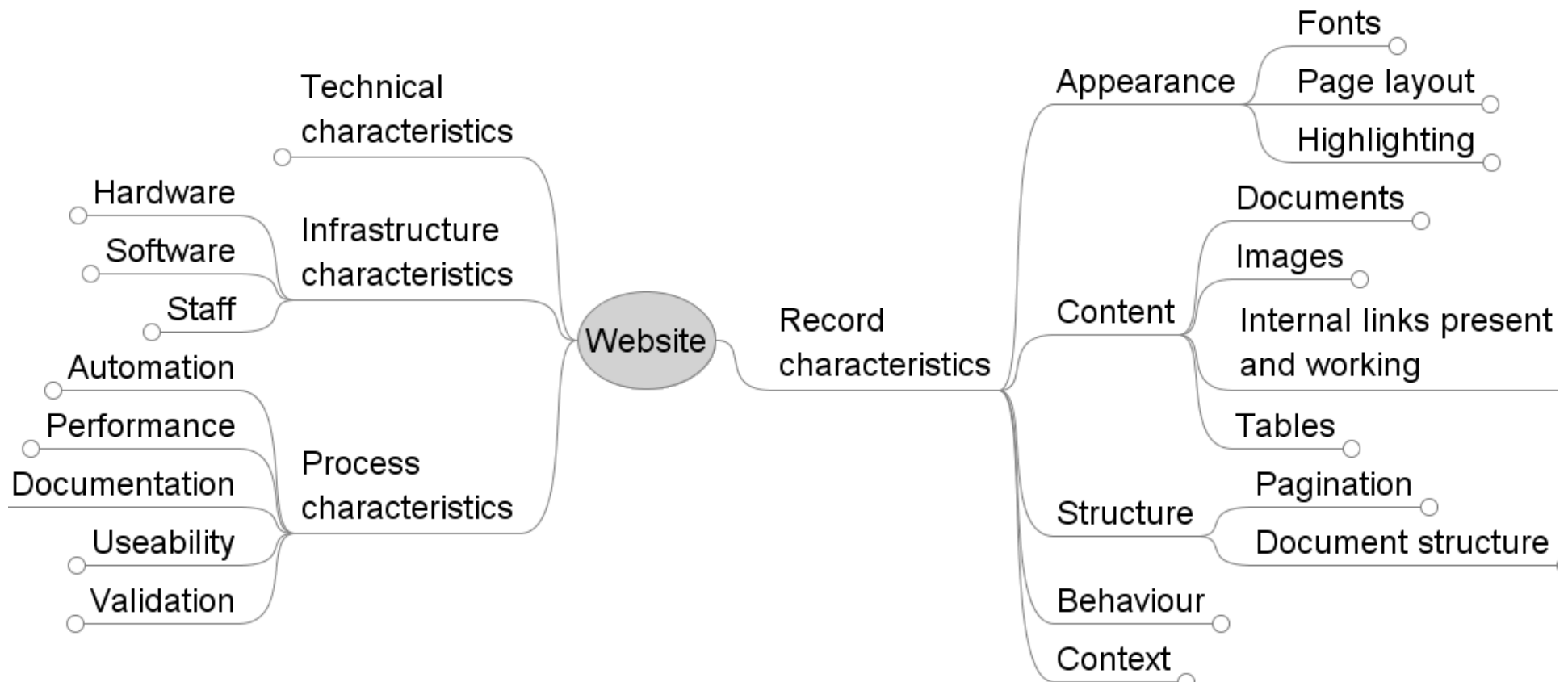


Object characteristics

- 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

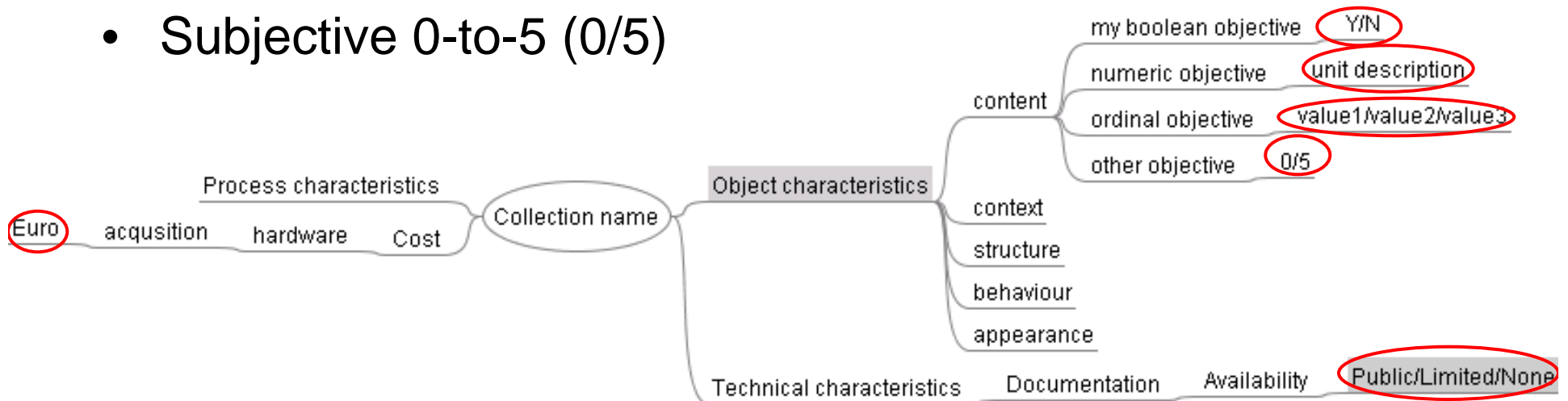
- ❑ Subjective scales where necessary
 - Adoption of file format
 - Amount of (expected) support

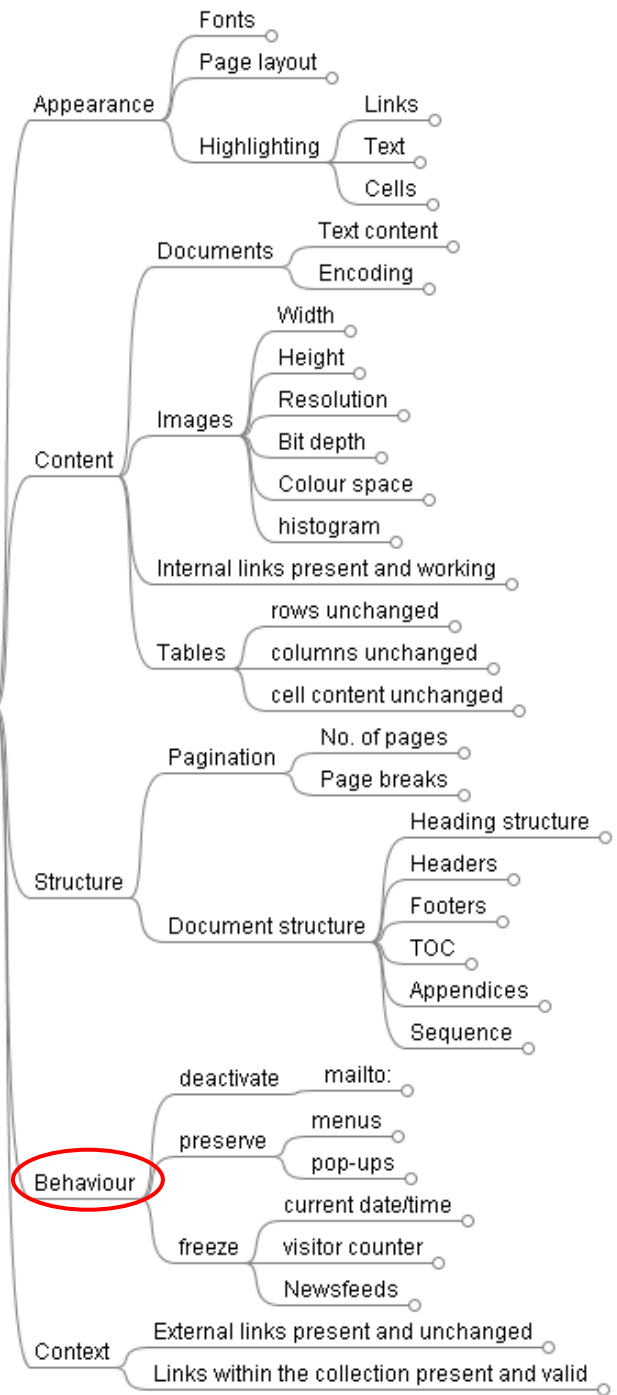
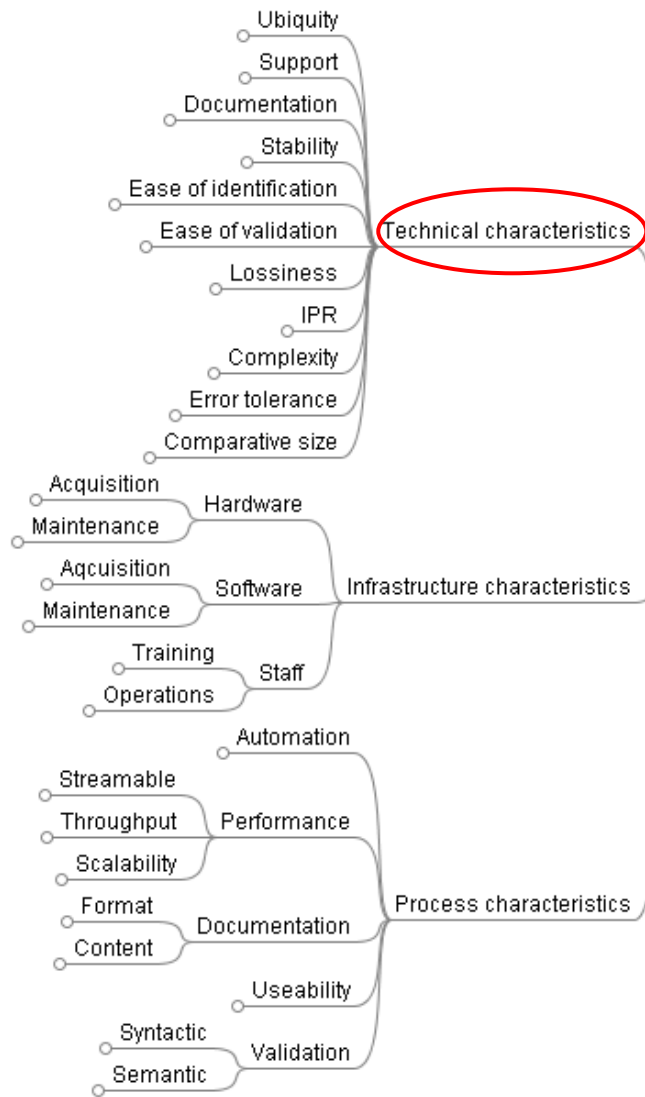
- Quantitative results



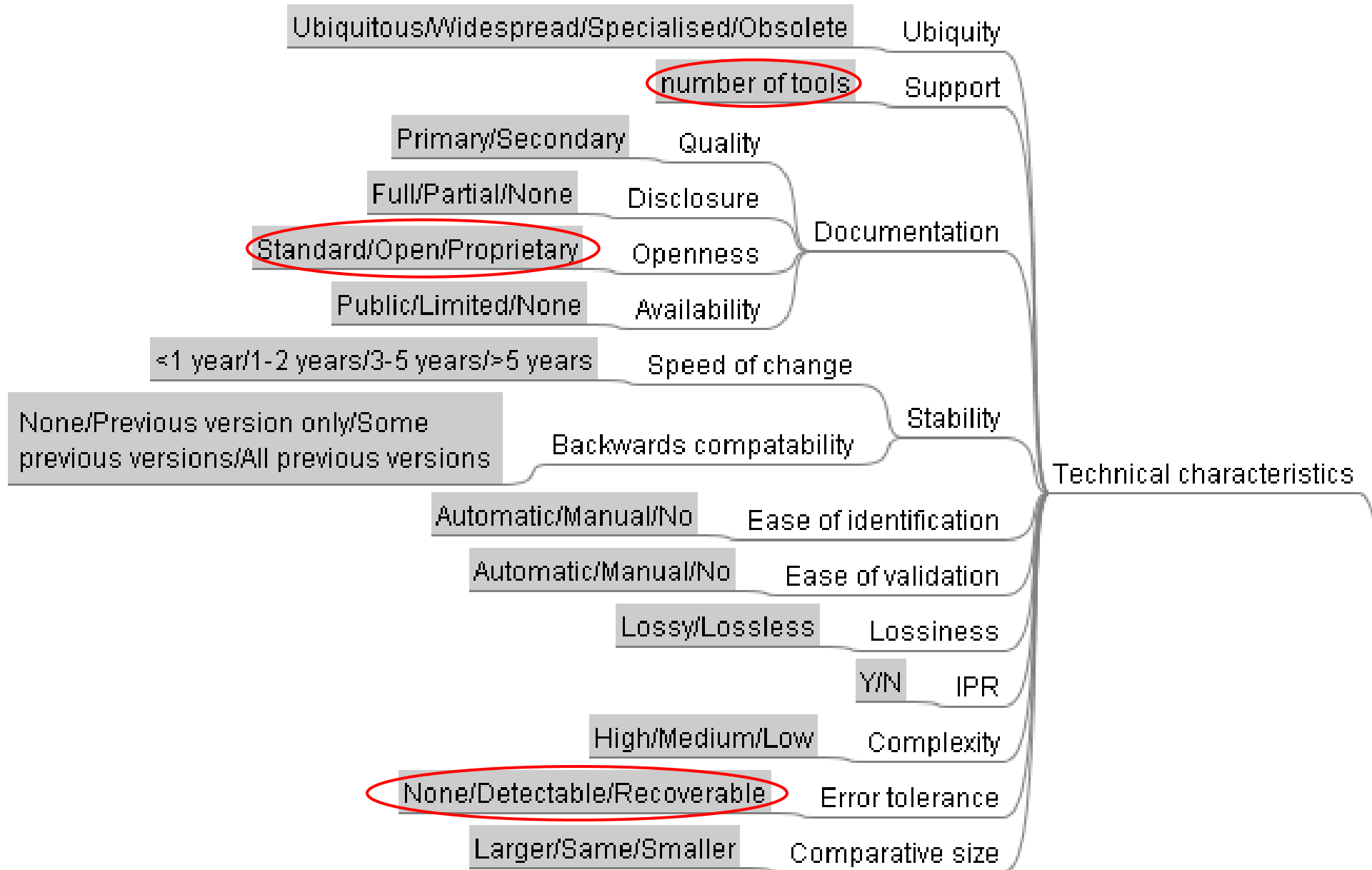
Types of scales

- 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)

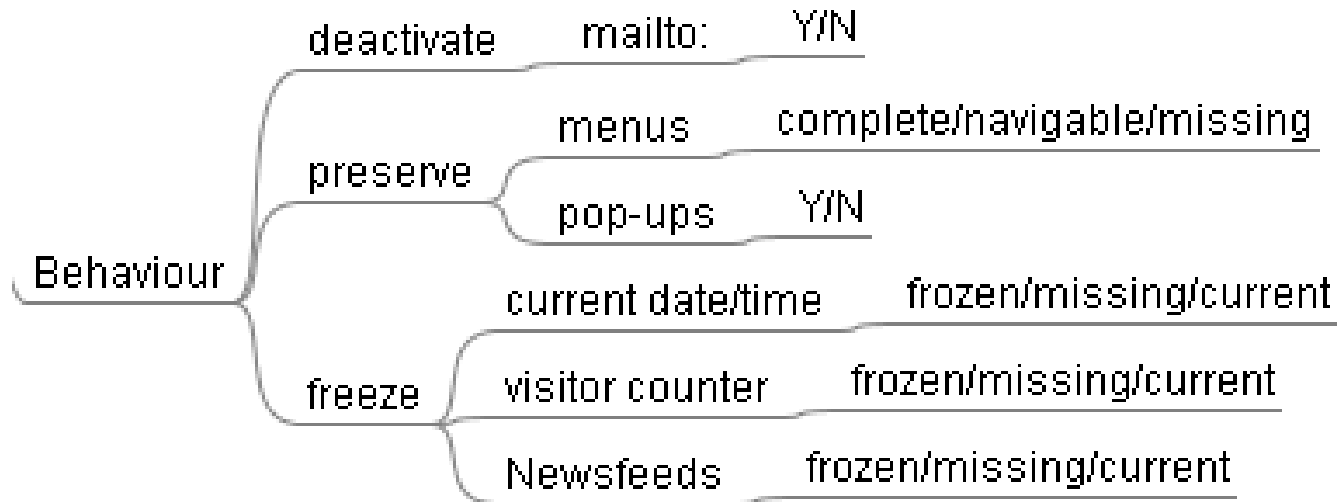




File format characteristics

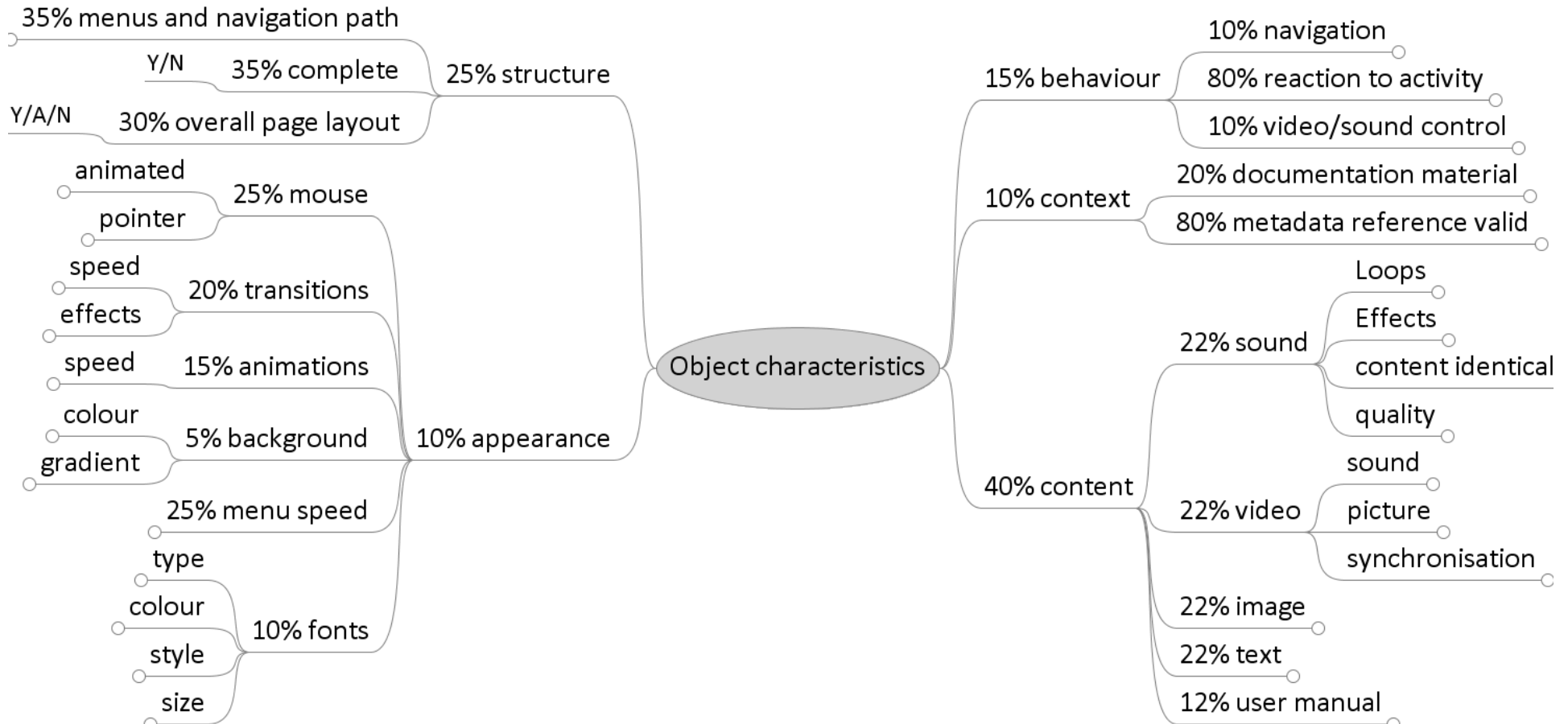


Behaviour



- 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

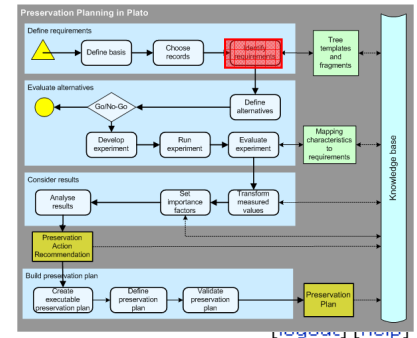
Node	Scale	Restriction
▼ Object characteristics		
▼ behaviour		
▶ navigation	Ordinal <input checked="" type="checkbox"/>	interactive and integrated/navigatable/none
▼ reaction to activity		
▼ mouse		
▶ position	Boolean <input type="checkbox"/>	
▶ clicks	Boolean <input type="checkbox"/>	
▶ keyboard	Boolean <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



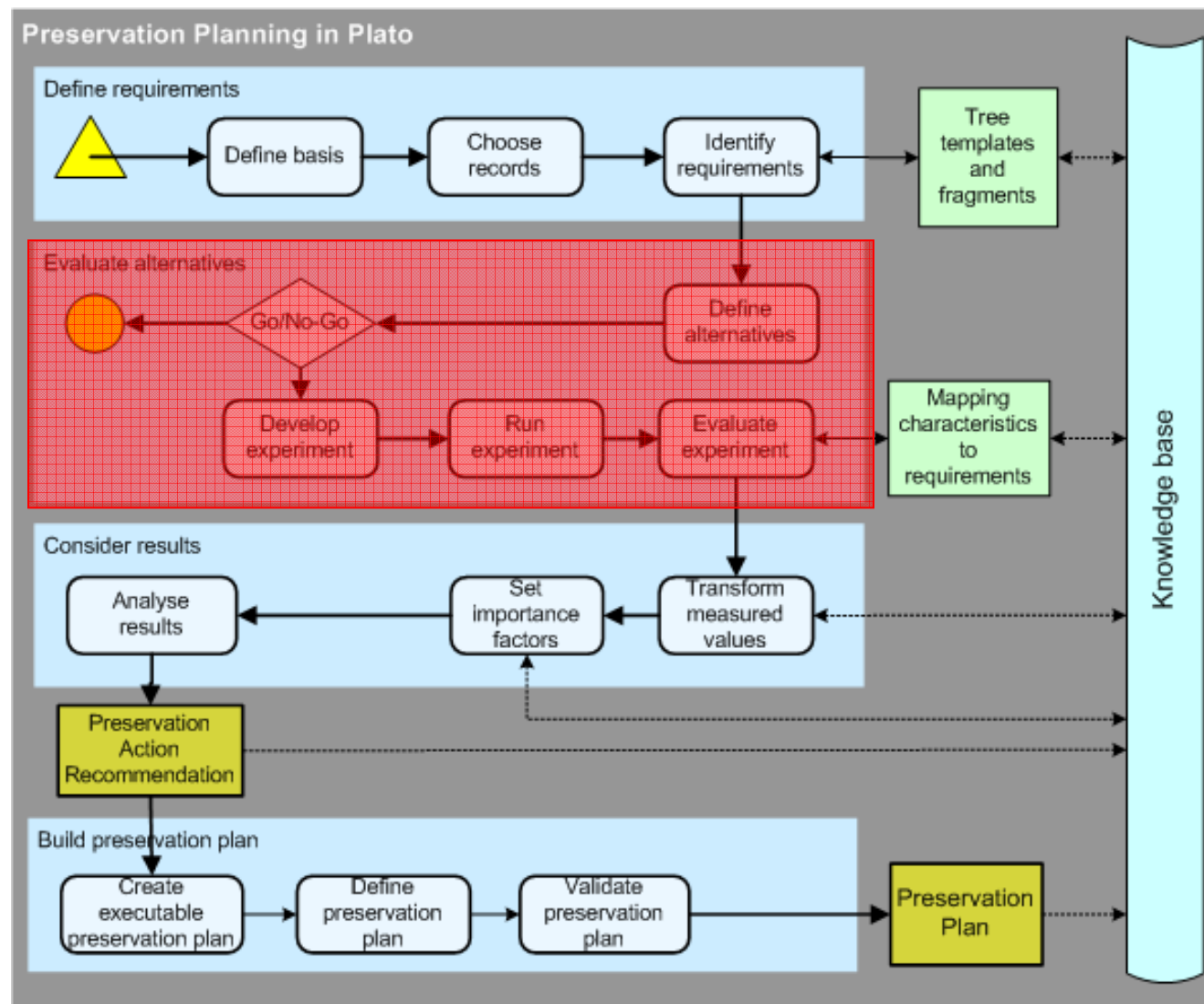
Project | Define Requirements | Evaluate Requirements | Consider Results | Loaded project: PP4 workshop - The National Archive

Identify Requirements

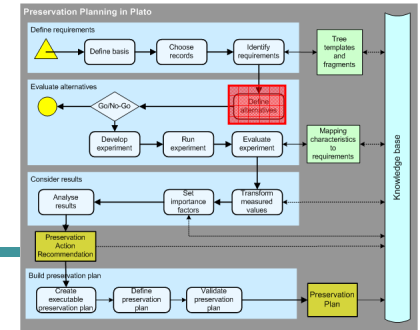
[Expand All](#) | [Collapse All](#)
[Website](#)

Focus	Node	+	+	-	Single	Scale	Restriction	Unit
	Website							
X	Record characteristics							
X	Appearance							
X	Content							
X	Structure							
X	Behaviour							
X	deactivate							
X	mailto:					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	

PP Workflow

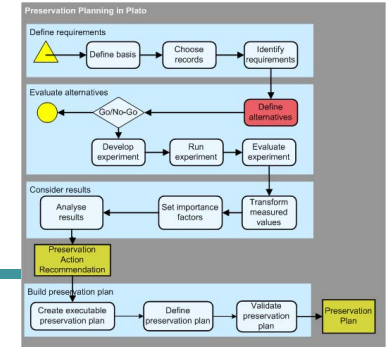


Define alternatives



- 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

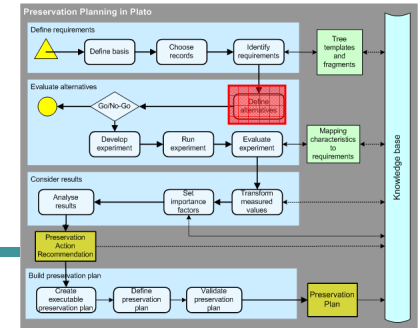


Show Preservation Services

	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

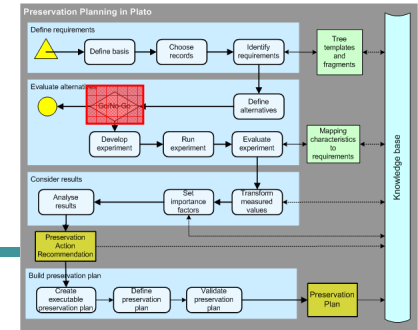
Create alternatives for selected services

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



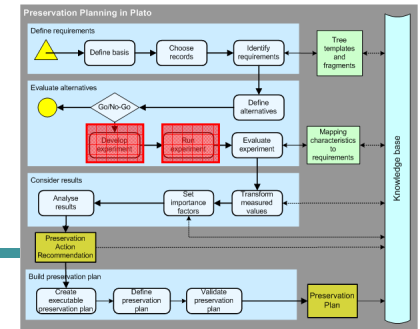
- Deliberate step for taking a decision whether it will be useful and cost-effective to continue the procedure, given
 - The resources to be spent (people, money)
 - The availability of tools and solutions,
 - The expected result(s).
- Review of the experiment/ evaluation process design so far
 - Is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

Develop and run experiment

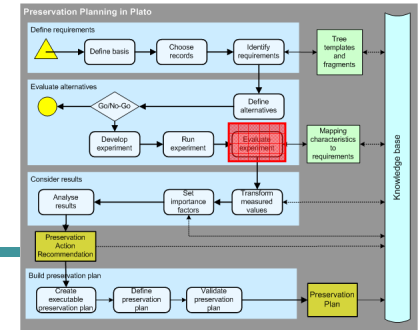
➤ Formulate for each experiment detailed

- 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

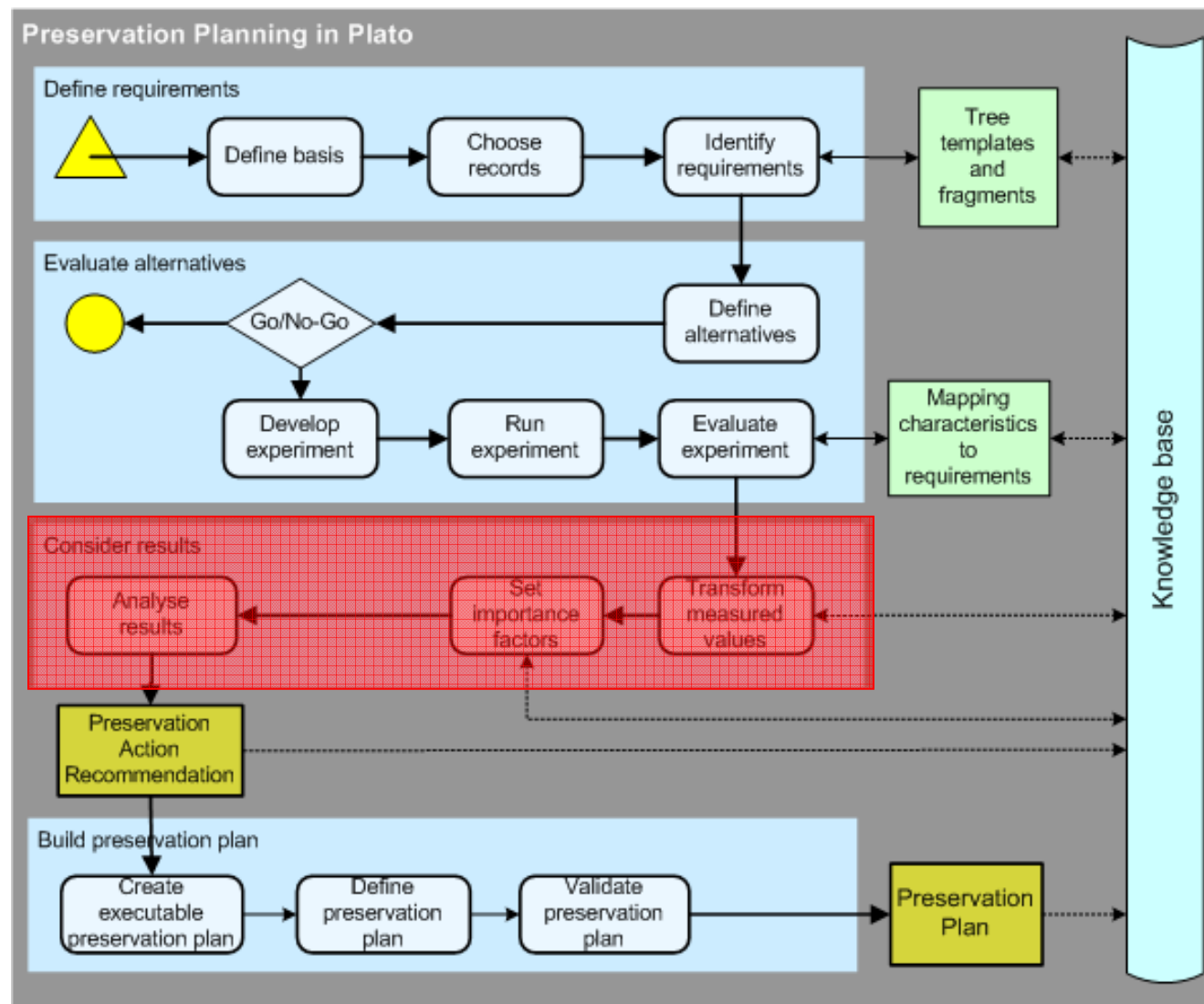


Evaluate experiment

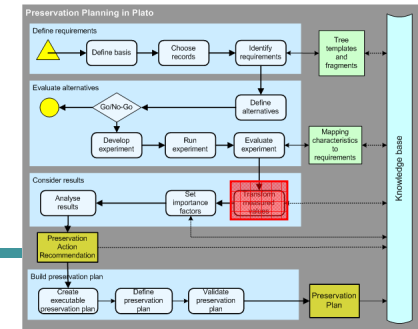


- Evaluate the outcome of each alternative for each leaf of the objective tree
- The evaluation will identify
 - Need for repeating the process
 - Unexpected (or undesired) results
- Includes both technical and intellectual aspects
- Evaluation may include comparing the results of more than one experiment/evaluation.

PP Workflow



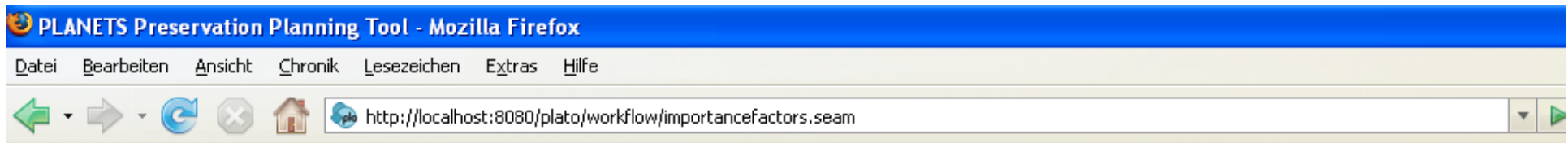
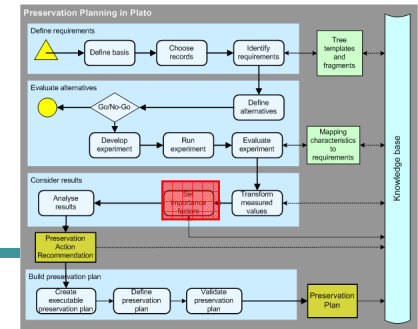
Transform measured values



- ❑ 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

- Adjust relative importance of all siblings in a branch



PLANETS Preservation Planning Tool (*Plato*)

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

Set Importance Factors

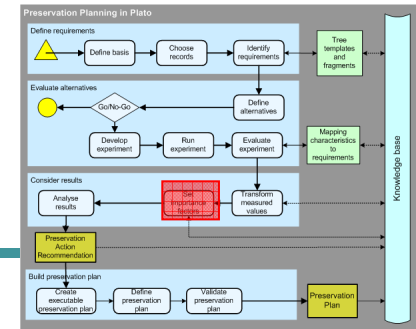
Balance weights automatically ☒

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

Object characteristics

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

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

Analyse results



PLANETS Preservation Planning Tool (*Plato*)



[\[logout admin\]](#) [\[help\]](#)

Project | Define Requirements | Evaluate Requirements | Consider Results | Minimalist test project in state #11

Analyse Results

Aggregation method:

Select	Alternative
<input checked="" type="checkbox"/>	PDF/A ToolA
<input checked="" type="checkbox"/>	PDF/A ToolB

Show

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

Minimalist root node

Focus	Name	Result
	▼ Minimalist root node	PDF/A ToolA: 2,98 PDF/A ToolB: 3,19
X	► Image properties	PDF/A ToolA: 0,70 PDF/A ToolB: 0,80
X	▼ Karma	PDF/A ToolA: 0,40 PDF/A ToolB: 0,00
X	▼ Filesize (in Relation to Original)	PDF/A ToolA: 0,78 PDF/A ToolB: 0,99
X	▼ A Single-Leaf	PDF/A ToolA: 0,40 PDF/A ToolB: 0,80
X	▼ IntRange 0-10	PDF/A ToolA: 0,70 PDF/A ToolB: 0,60

Recommendation

Recommendation:

Reasoning: Tool B has very bad Karma

Save recommendation

Generate final report

Questions?

becker@ifs.tuwien.ac.at

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

www.planets-project.eu



Outline

- ❑ 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



Practice time! (after lunch)

□ 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

- 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

□ Part 3 (optional, if time permits)

- Walk through the demo project in Plato (online)



Scenarios

- ❑ A: Word documents
in a governmental archive
- ❑ B: Word documents
in an enterprise archive
- ❑ C: Word documents
in an eLearning environment
- ❑ D: PowerPoint presentations
in an eLearning environment



Groups



- ❑ 1: A – Word docs in government archive
- ❑ 2: B – Word docs in enterprise archive
- ❑ 3: C – Word docs in eLearning environment
- ❑ 4: D – Powerpoint presentations in eLearning environment

- ❑ 5: A – Word docs in government archive
- ❑ 6: B – Word docs in enterprise archive
- ❑ 7: C – Word docs in eLearning environment
- ❑ 8: D – Powerpoint presentations in eLearning environment

- ❑ 9: A – Word docs in government archive
- ❑ 10: B – Word docs in enterprise archive
- ❑ 11: C – Word docs in eLearning environment
- ❑ 12: D – Powerpoint presentations in eLearning environment



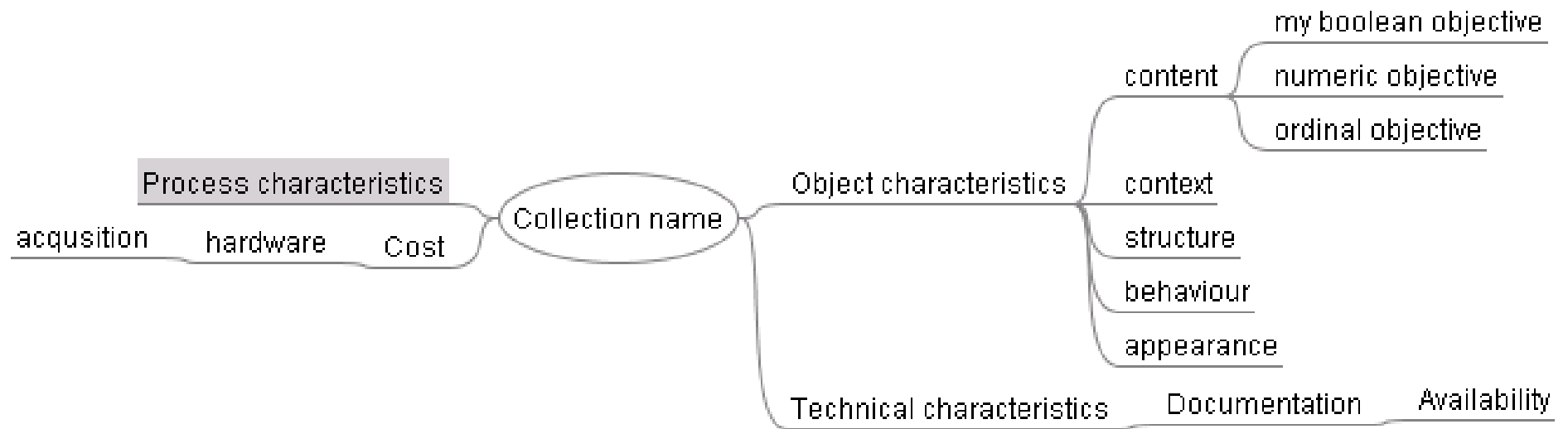
How to construct the tree

- With the open-source mind-mapping tool Freemind
 - USB stick with file, default mindmap and sample files
 - Java required
 - Freemind is installed in 20 seconds
- With post-it notes



Tree template

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



Questions?

becker@ifs.tuwien.ac.at

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

www.planets-project.eu



Groups



- ❑ 1: A – Word docs in government archive
- ❑ 2: B – Word docs in enterprise archive
- ❑ 3: C – Word docs in eLearning environment
- ❑ 4: D – Powerpoint presentations in eLearning environment

- ❑ 5: A – Word docs in government archive
- ❑ 6: B – Word docs in enterprise archive
- ❑ 7: C – Word docs in eLearning environment
- ❑ 8: D – Powerpoint presentations in eLearning environment

- ❑ 9: A – Word docs in government archive
- ❑ 10: B – Word docs in enterprise archive
- ❑ 11: C – Word docs in eLearning environment
- ❑ 12: D – Powerpoint presentations in eLearning environment

