Trustworthy Preservation Planning with Plato

Andreas Rauber

Department of Software Technology and Interactive Systems

Vienna University of Technology
rauber@ifs.tuwien.ac.at
http://www.ifs.tuwien.ac.at/~andi

Outline

- Introduction
- Why do we need preservation planning?
- Preservation planning and Plato
- Bringing it all together and closing

Why do we need Digital Preservation?

Why do we need Digital Preservation?

Why do we need Digital Preservation?

- Digital Objects require specific environment to be accessible :
  - Files need specific programs
  - Programs need specific operating systems (-versions)
  - Operating systems need specific hardware components
- SW/HW environment is not stable:
  - Files cannot be opened anymore
  - Embedded objects are no longer accessible/linked
  - Programs won‘t run
  - Information in digital form is lost (usually total loss, no degradation)
- Digital Preservation aims at maintaining digital objects authentically usable and accessible for long time periods.

Preservation Planning

Why Preservation Planning?

- Several preservation strategies developed
  - For each strategy: several tools available
    - For each tool: several parameter settings available
- How do you know which one is most suitable?
- What are the needs of your users? Now? In the future?
- Which aspects of an object do you want to preserve?
- What are the requirements?
- How to prove in 10, 20, 50, 100 years, that the decision was correct / acceptable at the time it was made?

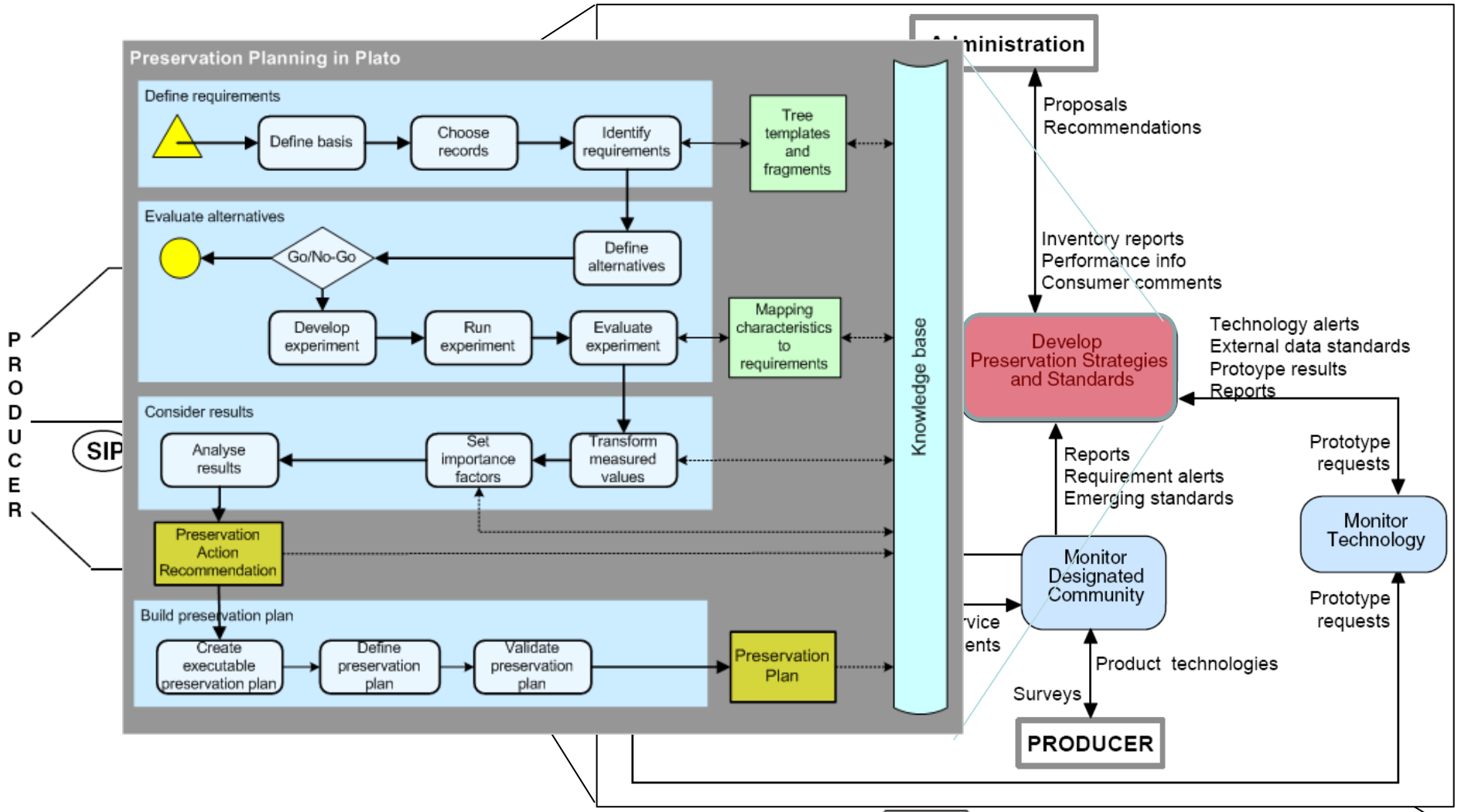
Preservation Planning

- Consistent workflow leading to a preservation plan
- Analyses, which solution to adopt
- Considers
  - preservation policies
  - legal obligations
  - organisational and technical constraints
  - user requirements and preservation goals
- Describes the
  - preservation context
  - evaluated preservation strategies
  - resulting decision including the reasoning
- Repeatable, solid evidence
- Trust and audit

Preservation Planning

- Trust and Audit
- Compliance to best practices, standards
- 3 core initiatives, of which 2 prescriptive
  - RLG- National Archives and Records Administration Digital Repository Certification Task Force:
    Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
  - NESTOR:
    Catalogue of Criteria of Trusted Digital Repositories
  - DCC/DPE:
    DRAMBORA: Digital Repository Audit Method Based on Risk Assessment
- Embedding into OAIS model

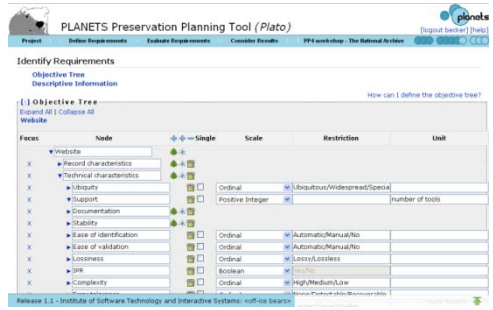
Preservation Planning



Outline

- Introduction
- Why do we need preservation planning?
- Preservation planning and Plato
- Bringing it all together and closing

Preservation Planning with Plato



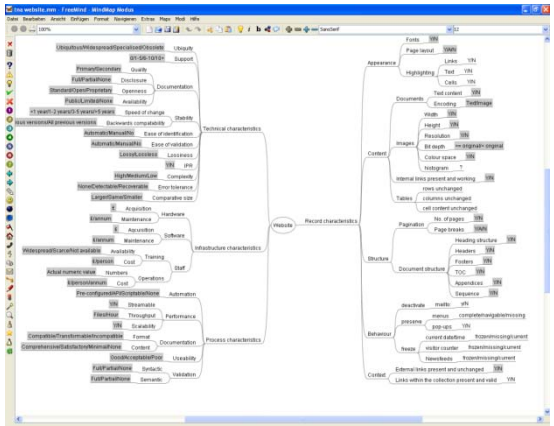
Plato

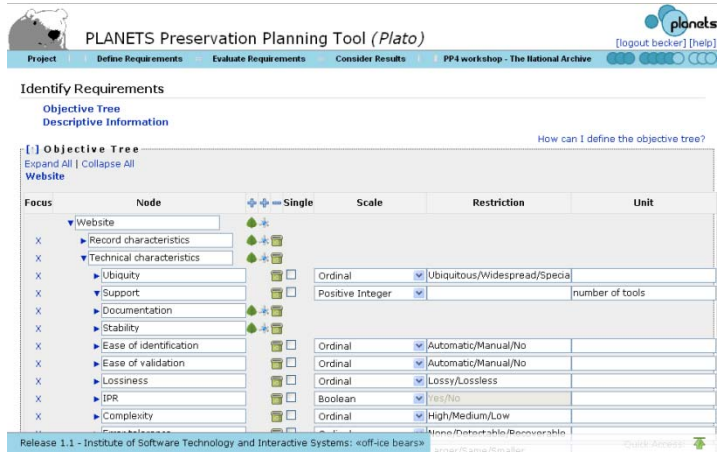
- Preservation Planning Tool
- Reference implementation of planning workflow
- Web-based application, 1st release 2.0 Nov. 12 2008, latest version 2.1 released November 2009
- Documents the process and ensures that all steps are considered
- Automation of the planning process via integration of registries and services
- Knowledge base to support planning
- Creates a preservation plan (XML, PDF)
- http://www.ifs.tuwien.ac.at/dp/plato

Preservation Planning with Plato

Plato

- Assists in analyzing the collection
  - Profiling, analysis of sample objects via Pronom and other services
- Allows creation of objective tree
  - Within application or via import of mindmaps
- Allows the selection of Preservation action tools

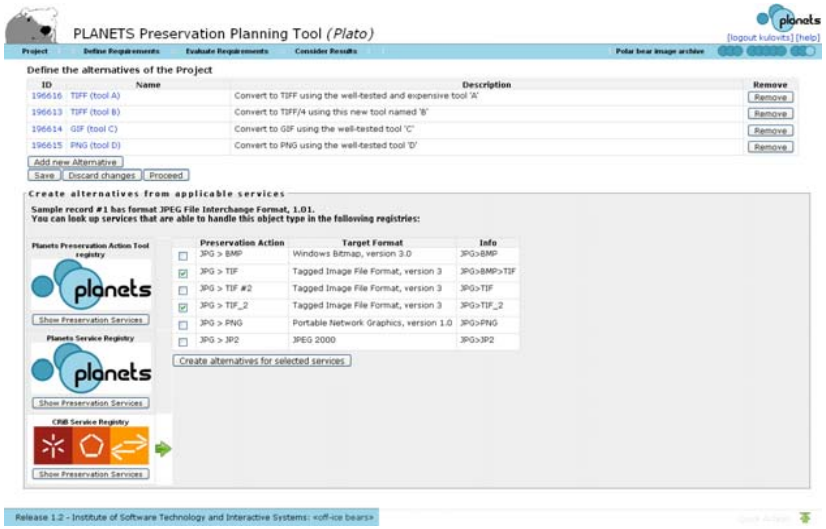


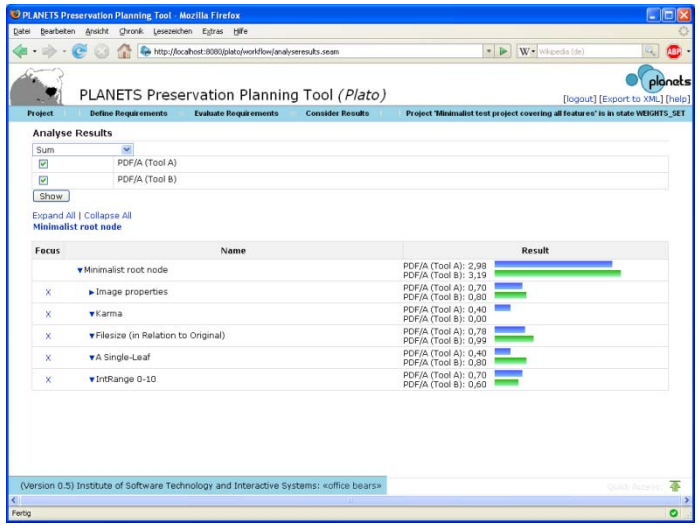


Preservation Planning with Plato

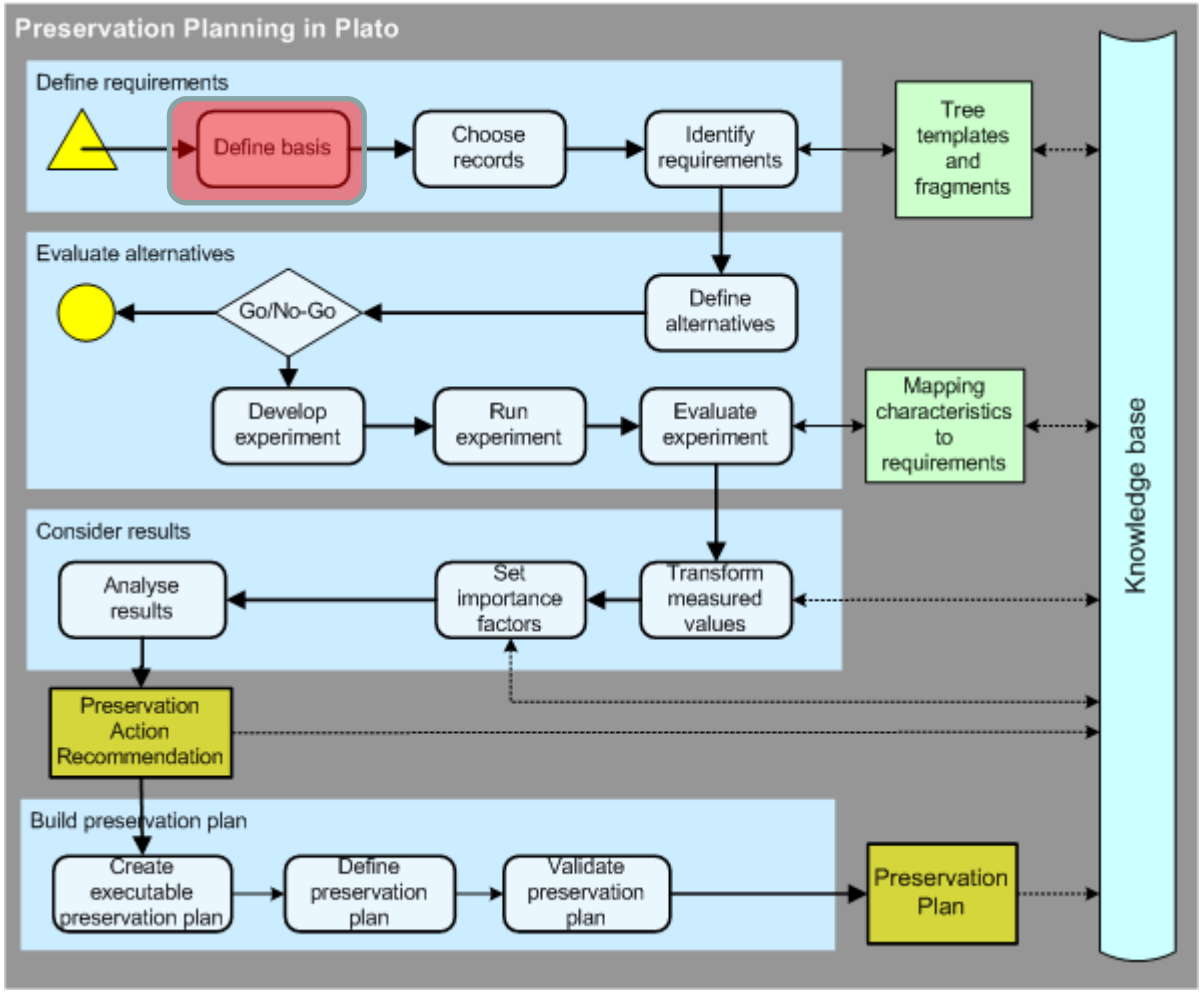
Plato

- Runs experiments and documents results
- Allows definition of transformation rules, weightings
- Performs evaluation, sensitivity analysis,
- Provides recommendation (ranks solutions)

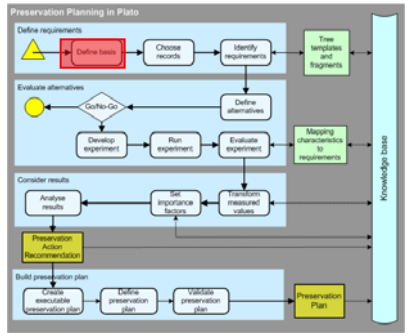




PP Workflow

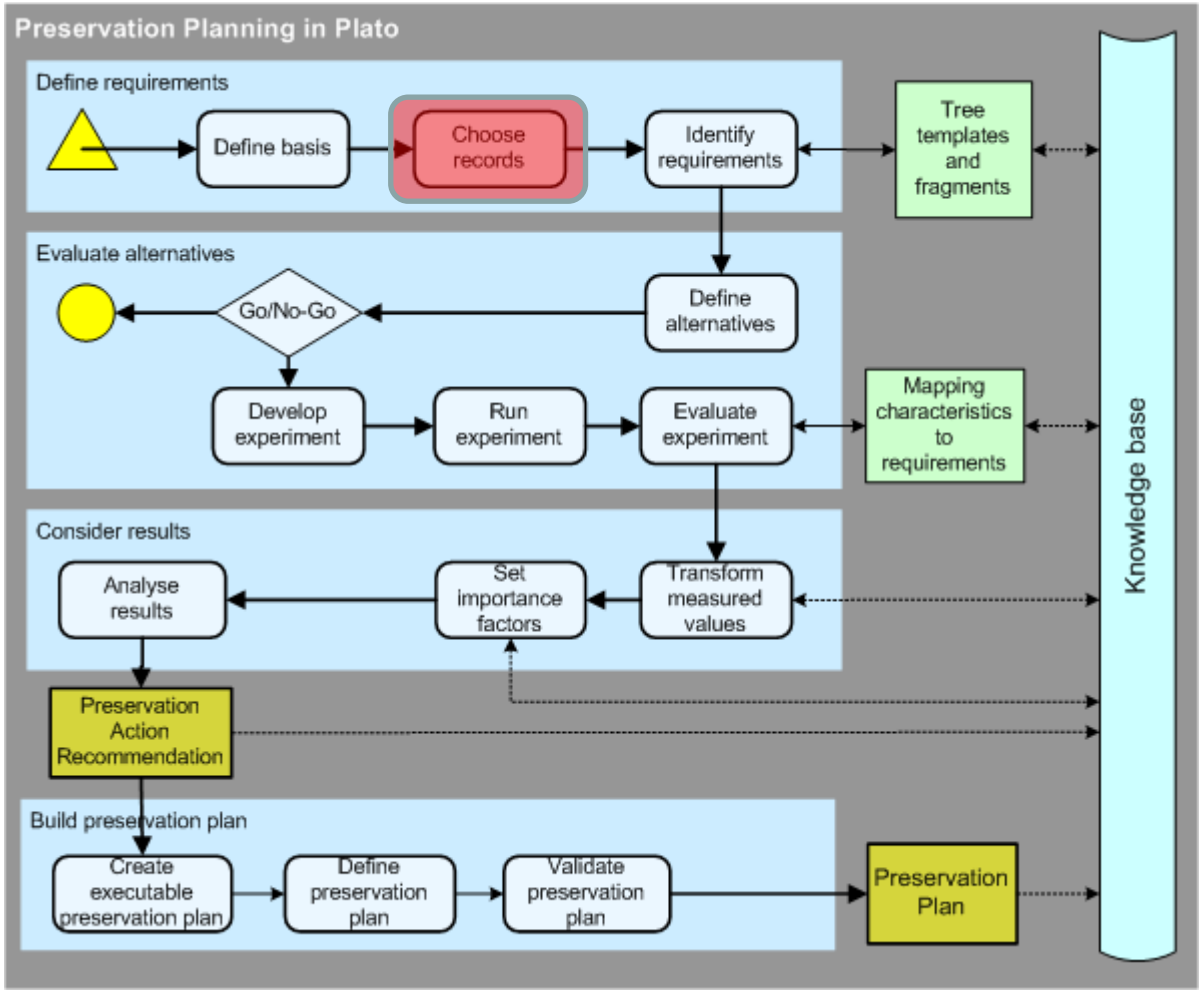

Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results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

Define Basis

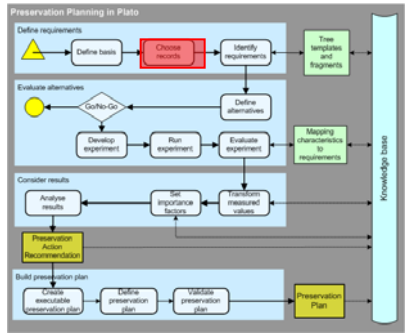


- Basic preservation plan properties
- Describe the context
  - Institutional settings
  - Legal obligations
  - User groups, target community
  - Organisational constraints
- 5 triggers
  - New Collection Alert (NCA)
  - Changed Collection Profile Alert (CPA)
  - Changed Environment Alert (CEA)
  - Changed Objective Alert (COA)
  - Periodic Review Alert (PRA)

PP Workflow

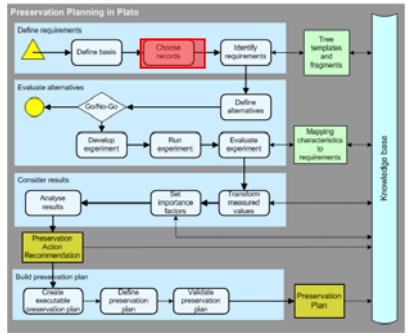

Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results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

Choose Sample Objects



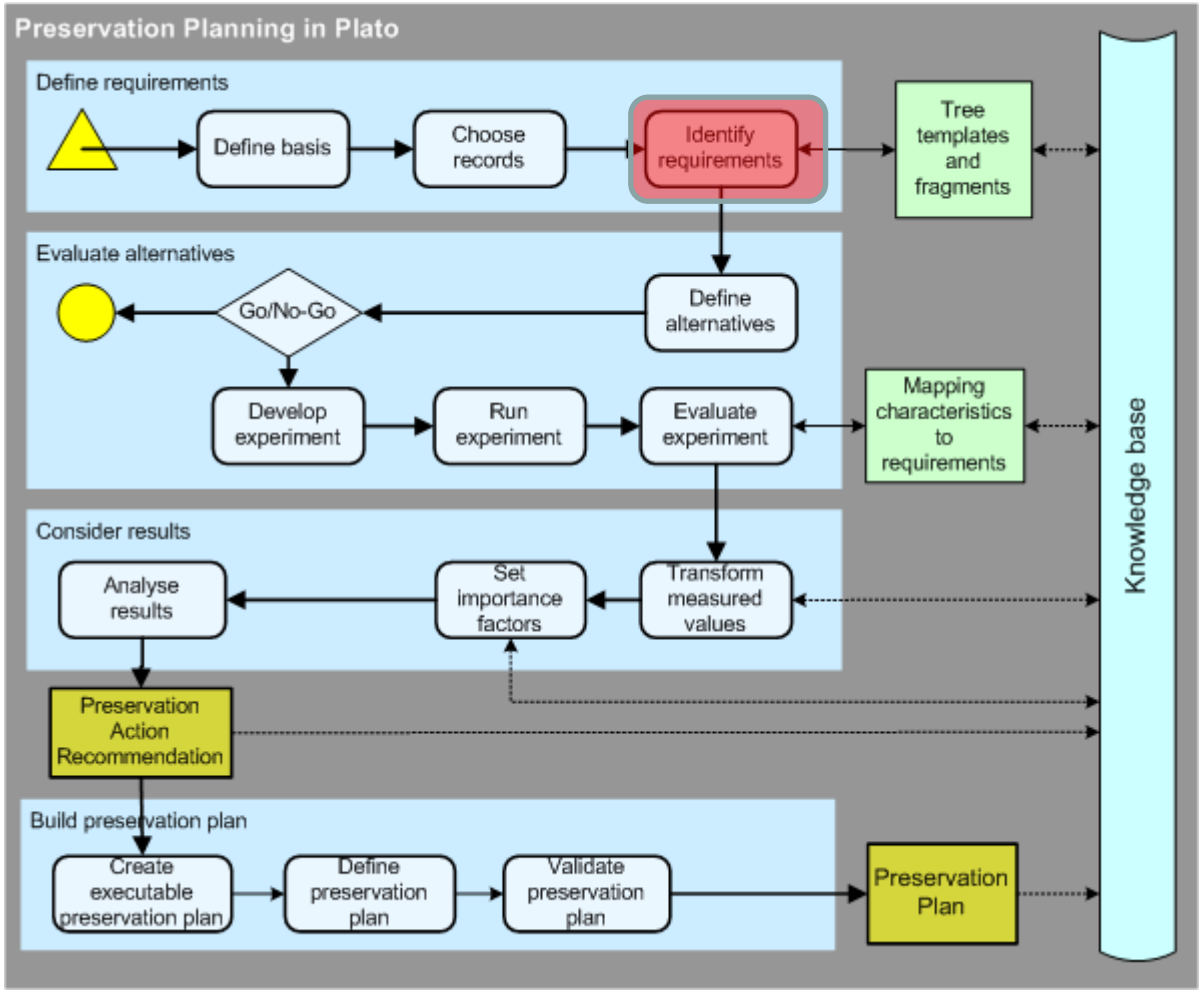
- Identify consistent (sub-)collections
  - Homogeneous type of objects (format, use)
  - To be handled with a specific (set of) tools
- Describe the collection
  - What types of objects?
  - How many?
  - Which format(s)?
- Selection
  - Representative for the objects in the collection
  - Right choice of sample is essential
  - They should cover all essential features and characteristics of the collection in question
  - As few as possible, as many as needed
  - Often between 3 – 10

Choose Sample Objects

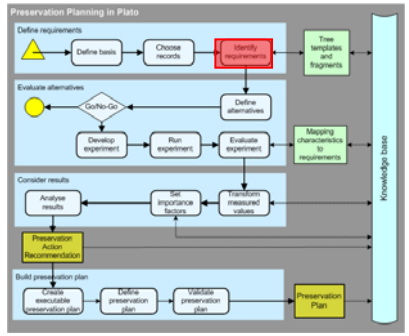


- Stratification – all essential groups of digital objects should be chosen according to their relevance
- Possible stratification strategies
  - File type
  - Size
  - Content (e.g. document with lots of images, including macros)
  - Time (objects from different periods of times)
- File Format Identification
  - DROID
  - PRONOM

PP Workflow

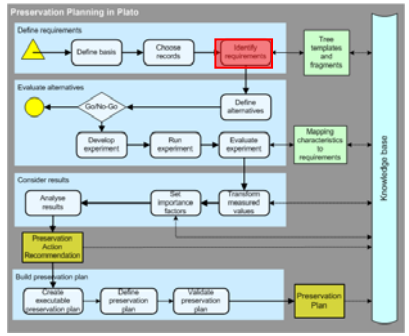

Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results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

Identify Requirements

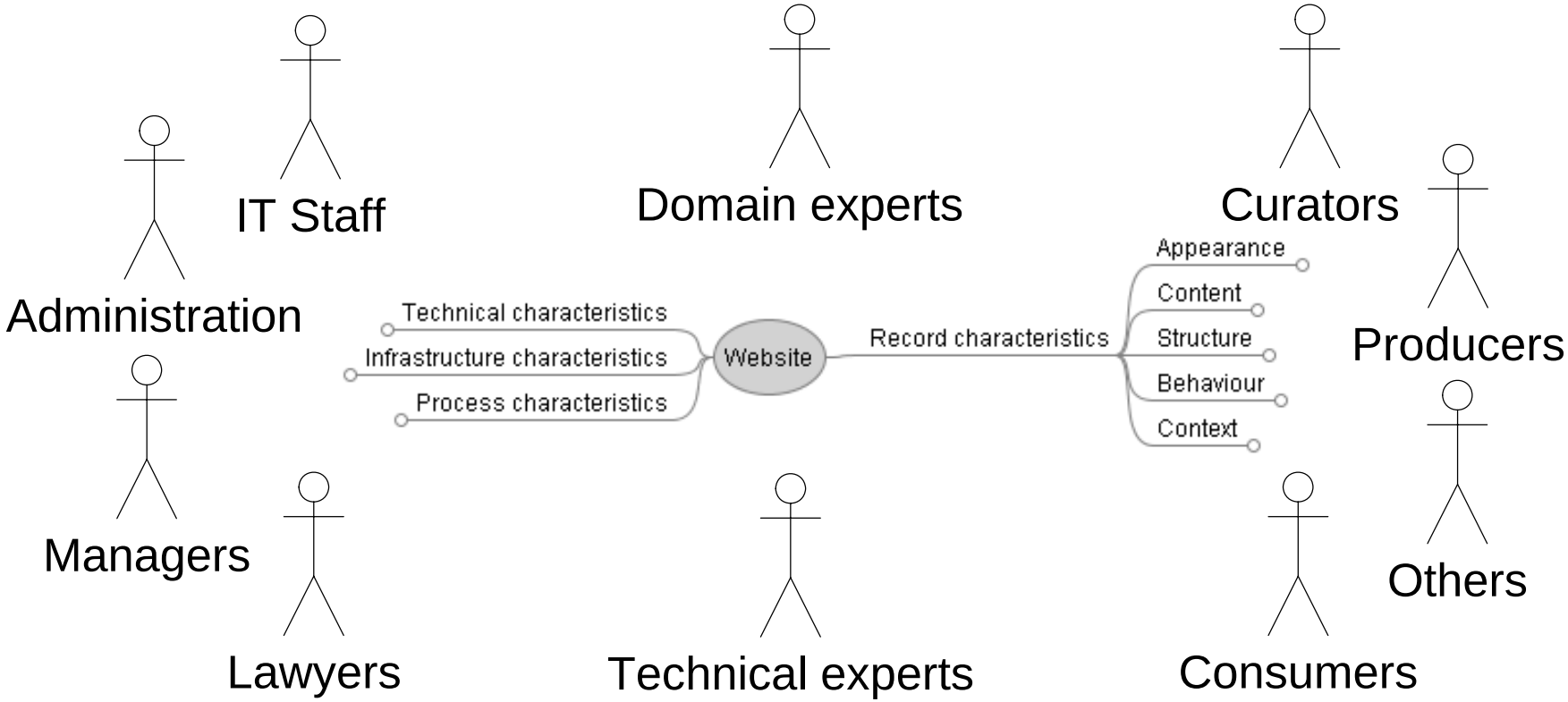


- Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain
- Put the requirements in relation to each other
  → Tree structure
- Top-down or bottom-up
  - Start from high-level goals and break down to specific criteria
  - Collect criteria and organize in tree structure

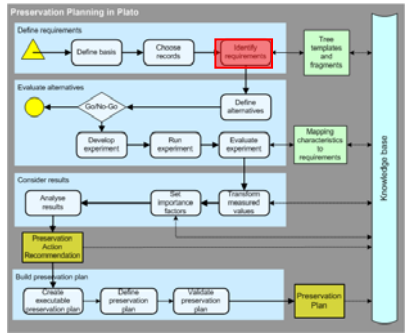
Identify Requirements



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

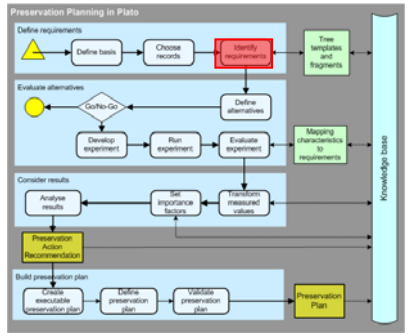


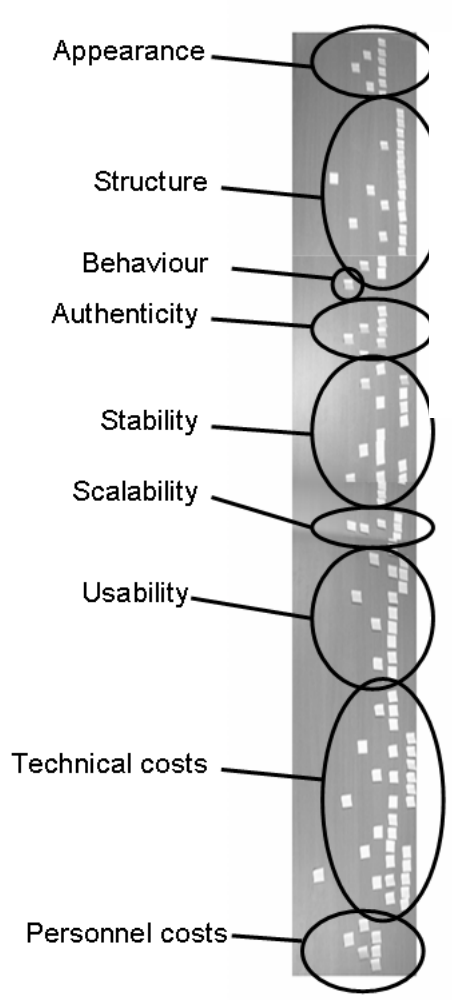
Identify requirements

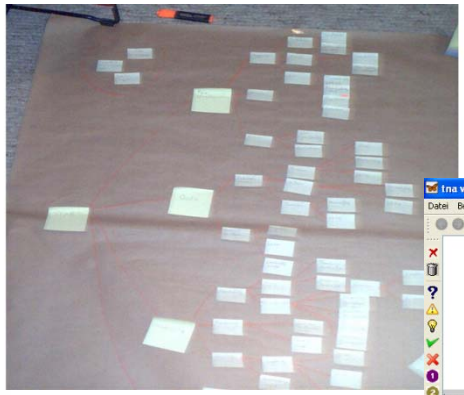


- Core step in the process
- Define all relevant goals and characteristics (high-level, detail) with respect to given application domain
- Usually four major groups
  - Object characteristics (content, metadata,…)
  - Record characteristics (context, relations,…)
  - Process characteristics (scalability, error-detection,…)
  - Costs (set-up, per object, HW/SW; personnel,…)

Identify requirements

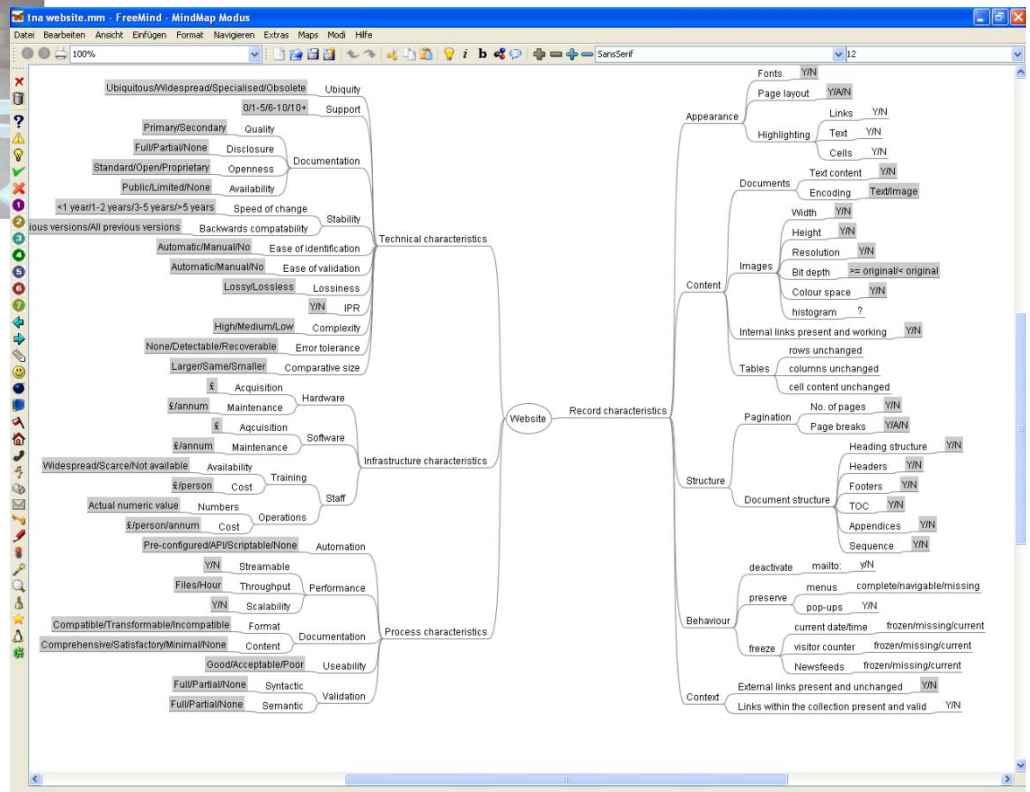




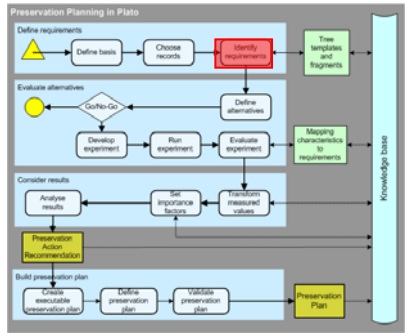


analogue…

… or digital



Identify requirements



- Creation within PLATO with Tree-Editor

PLANETS Preservation Planning Tool *(Plato)*

[logout becker] [help]

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

Identify Requirements

Objective Tree
Descriptive Information

How can I define the objective tree?

[↑] Objective Tree

Expand All | Collapse All

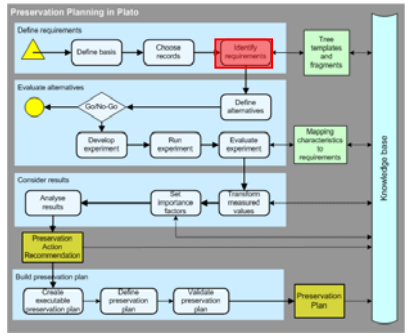
Website

Focus	Node	Single	Scale	Restriction	Unit
	Website				
X	Record characteristics				
X	Technical characteristics				
X	Ubiquity		Ordinal	Ubiquitous/Widespread/Specia	
X	Support		Positive Integer		number of tools
X	Documentation				
X	Stability				
X	Ease of identification		Ordinal	Automatic/Manual/No	
X	Ease of validation		Ordinal	Automatic/Manual/No	
X	Lossiness		Ordinal	Lossy/Lossless	
X	IPR		Boolean	Yes/No	
X	Complexity		Ordinal	High/Medium/Low	
				None/Detectable/Recoverable	

Release 1.1 - Institute of Software Technology and Interactive Systems: «off-ice bears»

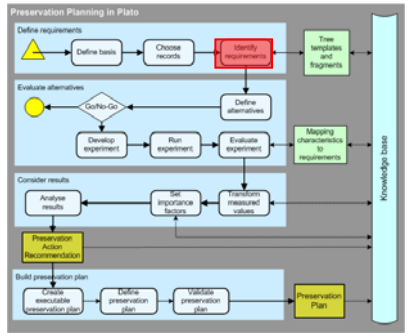
Quick Access:

Identify requirements



- Assign measurable unit to each leaf criterion
  - As far as possible automatically measurable
    - seconds / Euro per object
    - colour depth in bits
    - ...
  - Subjective measurement units where necessary
    - diffusion of file format
    - amount of expected support
    - ...
- No limitations on the type of scale used

Identify requirements



- Creation within PLATO with Tree-Editor

PLANETS Preservation Planning Tool *(Plato)*

[logout becker] [help]

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

Identify Requirements

Objective Tree
Descriptive Information

How can I define the objective tree?

[↑] Objective Tree

Expand All | Collapse All

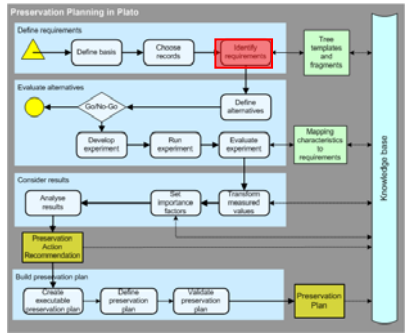
Website

Focus	Node	Single	Scale	Restriction	Unit
	Website				
X	Record characteristics				
X	Technical characteristics				
X	Ubiquity	☐	Ordinal	Ubiquitous/Widespread/Specia	
X	Support	☐	Positive Integer		number of tools
X	Documentation				
X	Stability				
X	Ease of identification	☐	Ordinal	Automatic/Manual/No	
X	Ease of validation	☐	Ordinal	Automatic/Manual/No	
X	Lossiness	☐	Ordinal	Lossy/Lossless	
X	IPR	☐	Boolean	Yes/No	
X	Complexity	☐	Ordinal	High/Medium/Low	

Release 1.1 - Institute of Software Technology and Interactive Systems: «off-ice bears»

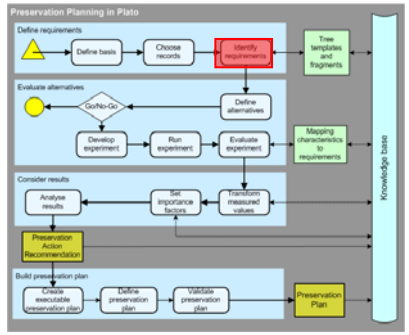
Quick Access:

Identify Requirements: Example

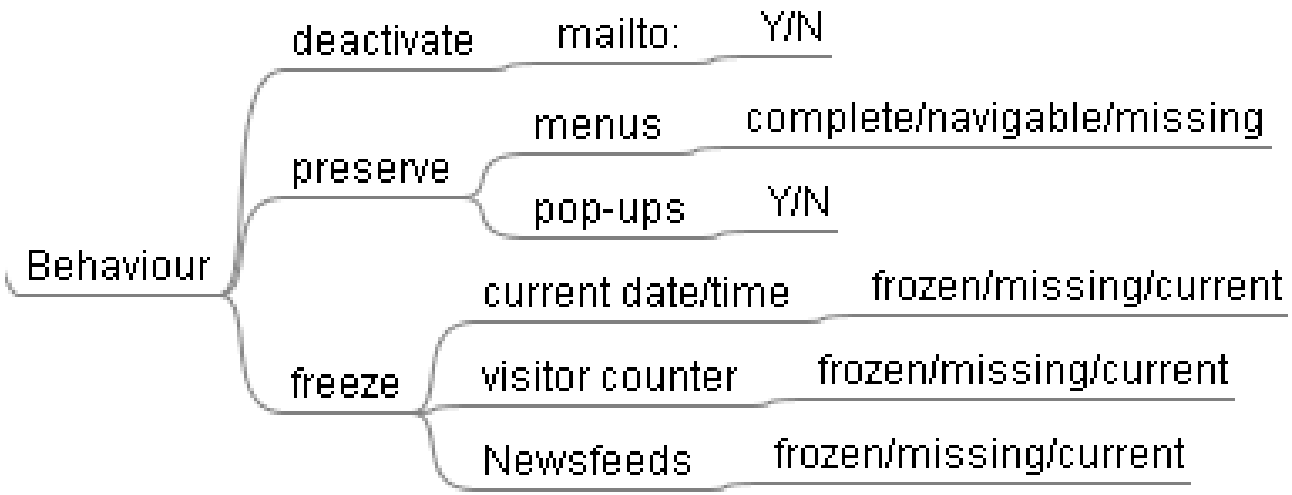


- Technical characteristics
  - Ubiquity: Ubiquitous/Widespread/Specialised/Obsolete
  - Support: number of tools
  - Documentation
    - Quality: Primary/Secondary
    - Disclosure: Full/Partial/None
    - Openness: Standard/Open/Proprietary
    - Availability: Public/Limited/None
  - Stability
    - Speed of change: <1 year/1-2 years/3-5 years/>5 years
    - Backwards compatability: None/Previous version only/Some previous versions/All previous versions
  - Ease of identification: Automatic/Manual/No
  - Ease of validation: Automatic/Manual/No
  - Lossiness: Lossy/Lossless
  - IPR: Y/N
  - Complexity: High/Medium/Low
  - Error tolerance: None/Detectable/Recoverable
  - Comparative size: Larger/Same/Smaller

Identify Requirements: Example

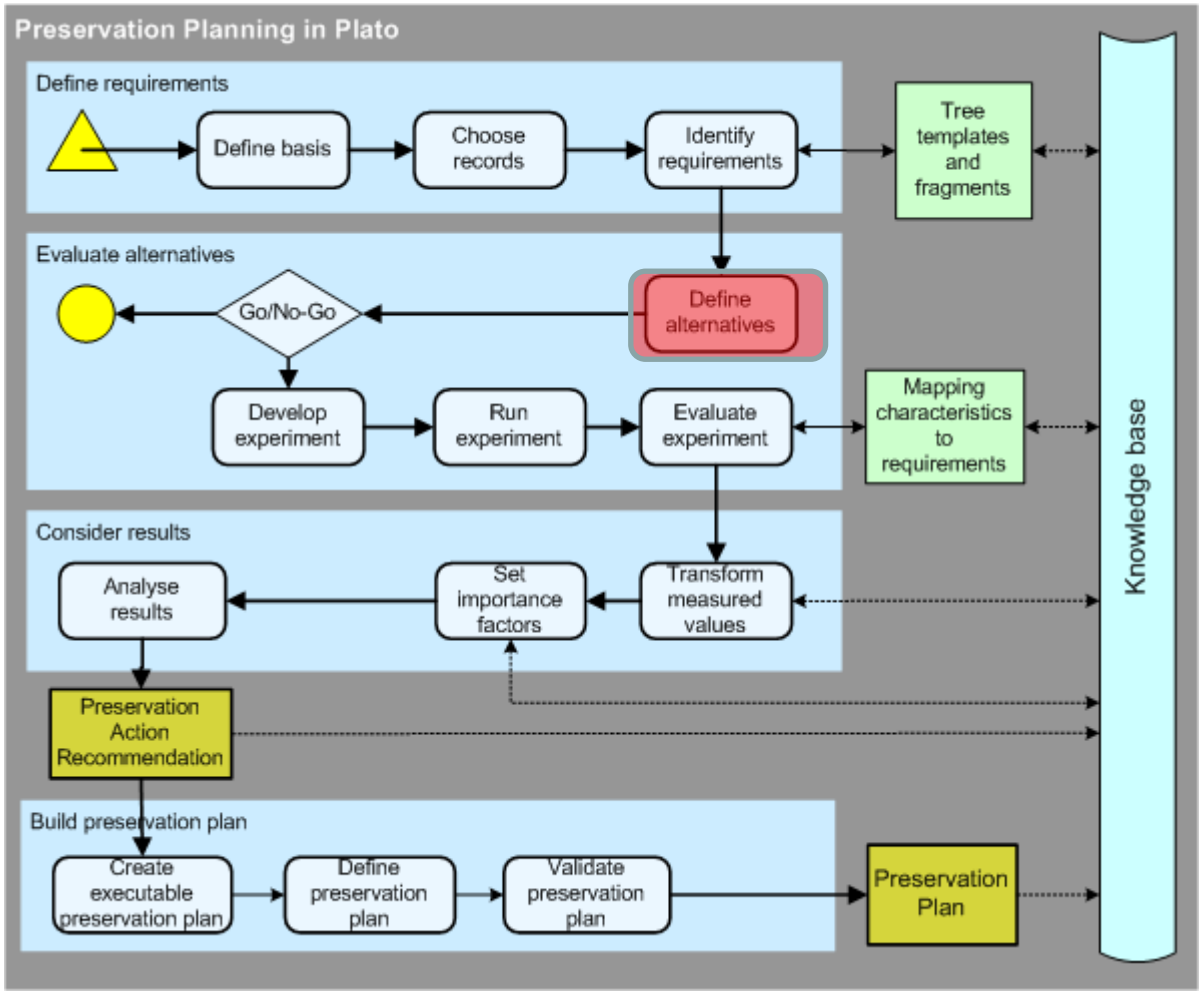


Behaviour

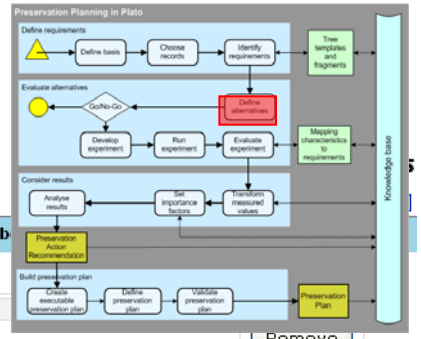


- Visitor counter and similar functionalities can be
  - Frozen at harvesting time
  - Omitted
  - Remain operational, i.e. the counter will be increased upon archival calls
    (is this desired? count? demonstrate functionality?)

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
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan
Knowledge base

Define Alternatives



PLANETS Preservation Planning Tool *(Plato)*

Project | Define Requirements ::: Evaluate Requirements ::: Consider Results | Polar b

Define the alternatives of the Project

ID	Name	Description	
196616	TIFF (tool A)	Convert to TIFF using the well-tested and expensive tool 'A'	Remove
196613	TIFF (tool B)	Convert to TIFF/4 using this new tool named 'B'	Remove
196614	GIF (tool C)	Convert to GIF using the well-tested tool 'C'	Remove
196615	PNG (tool D)	Convert to PNG using the well-tested tool 'D'	Remove

Add new Alternative

Save | Discard changes | Proceed

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

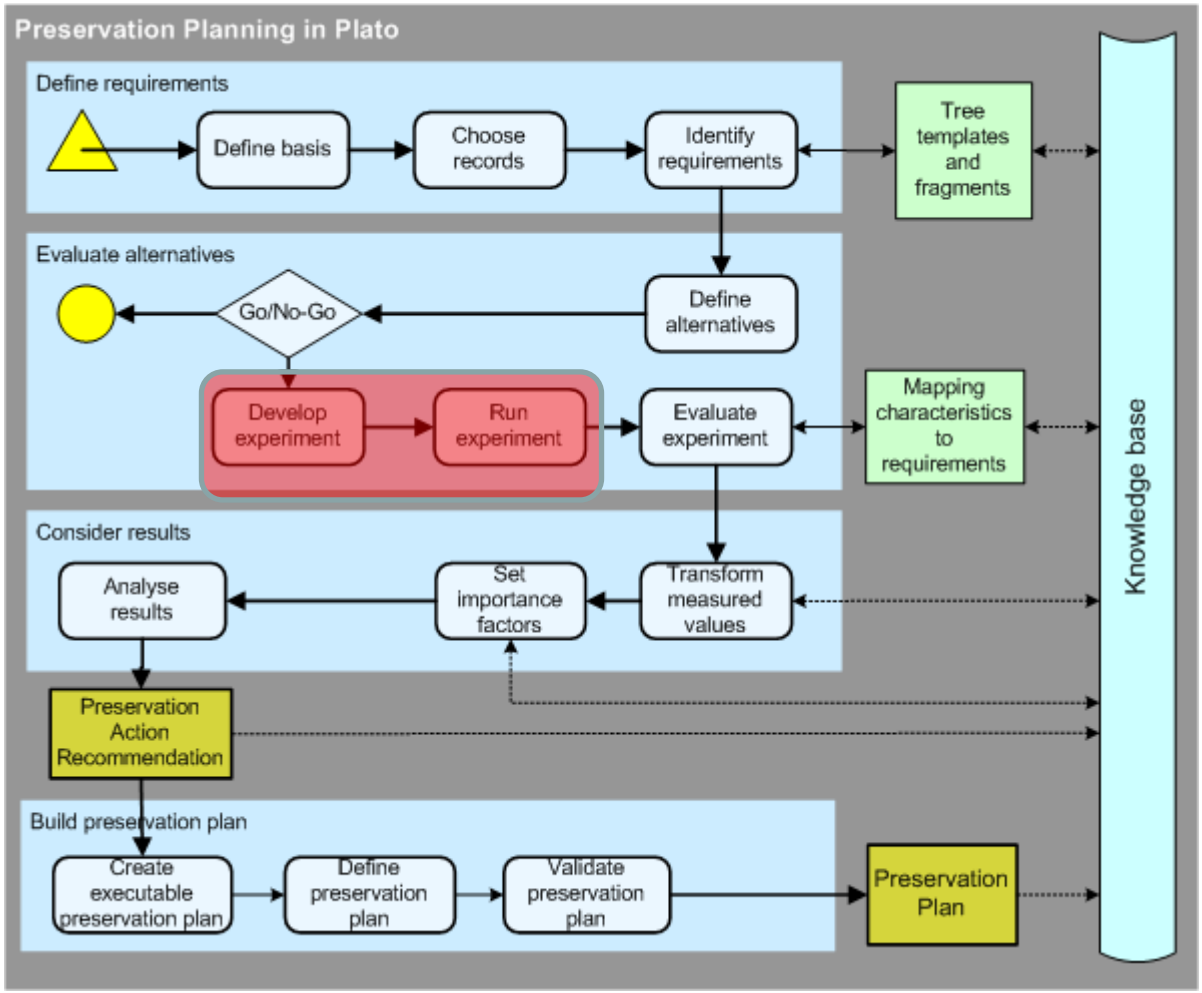
CRiB Service Registry

Show Preservation Services

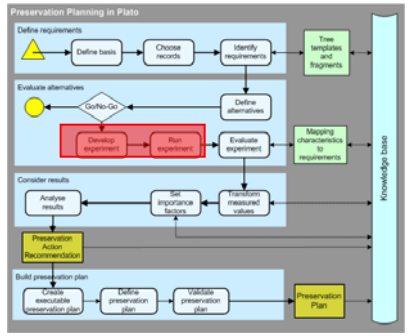
	Preservation Action	Target Format	Info
☐	JPG > BMP	Windows Bitmap, version 3.0	JPG>BMP
☑	JPG > TIF	Tagged Image File Format, version 3	JPG>BMP>TIF
☐	JPG > TIF #2	Tagged Image File Format, version 3	JPG>TIF
☑	JPG > TIF_2	Tagged Image File Format, version 3	JPG>TIF_2
☐	JPG > PNG	Portable Network Graphics,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

Create alternatives for selected services

PP Workflow

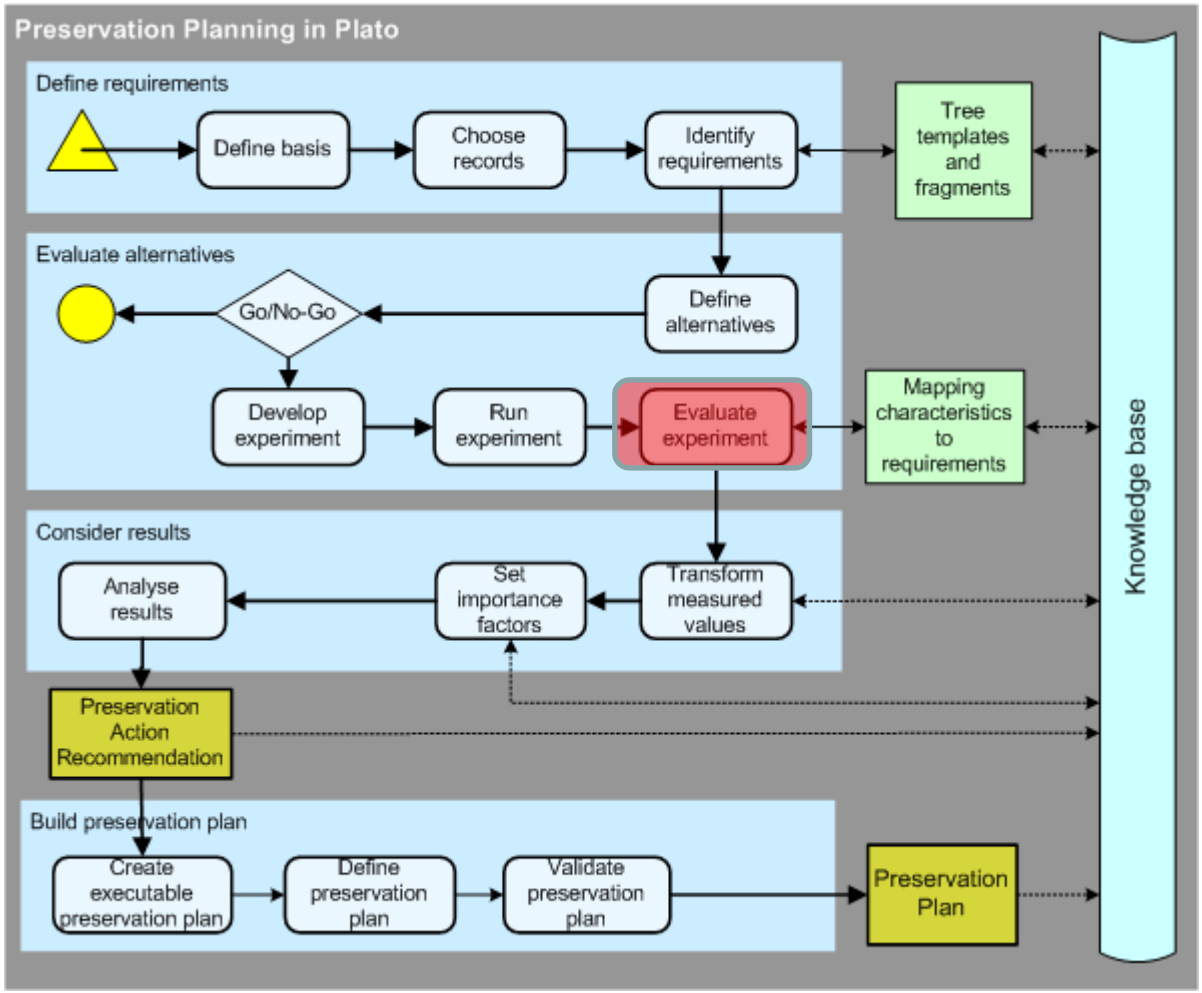

Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results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

Develop and Run Experiment

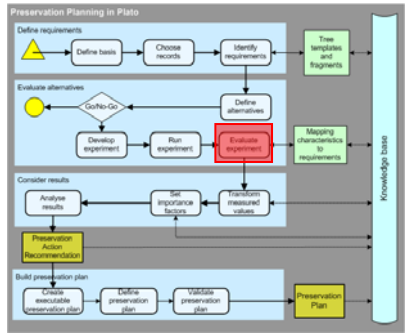


- Call the migration tools / emulators
- Convert the sample objects / open in emulators
- Take a look at the results.
  each of the sample object in various versions
- Measure time / memory needed to migrate/open
- Measure program output, error messages, …

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
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan
Knowledge base

Evaluate Experiment



PLANETS Preservation Planning Tool *(Plato)*

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

Evaluate Experiment

Expand All | Collapse All

Polar bear image preservation

Focus	Node
	▼Polar bear image preservation
X	▼Process
X	▼Complexity
X	▼Cost
X	▼Image properties
X	▼Bits of colour depth
X	▼Technical characteristics
X	▼Official standard
X	▼Filesize (in Relation to Original)

Comments:

Process > Complexity

Alternative	Single result		Comments
TIFF (tool A)	Simple	*	
TIFF (tool B)	Simple	*	
GIF (tool C)	Complex	*	
PNG (tool D)	Medium	*	

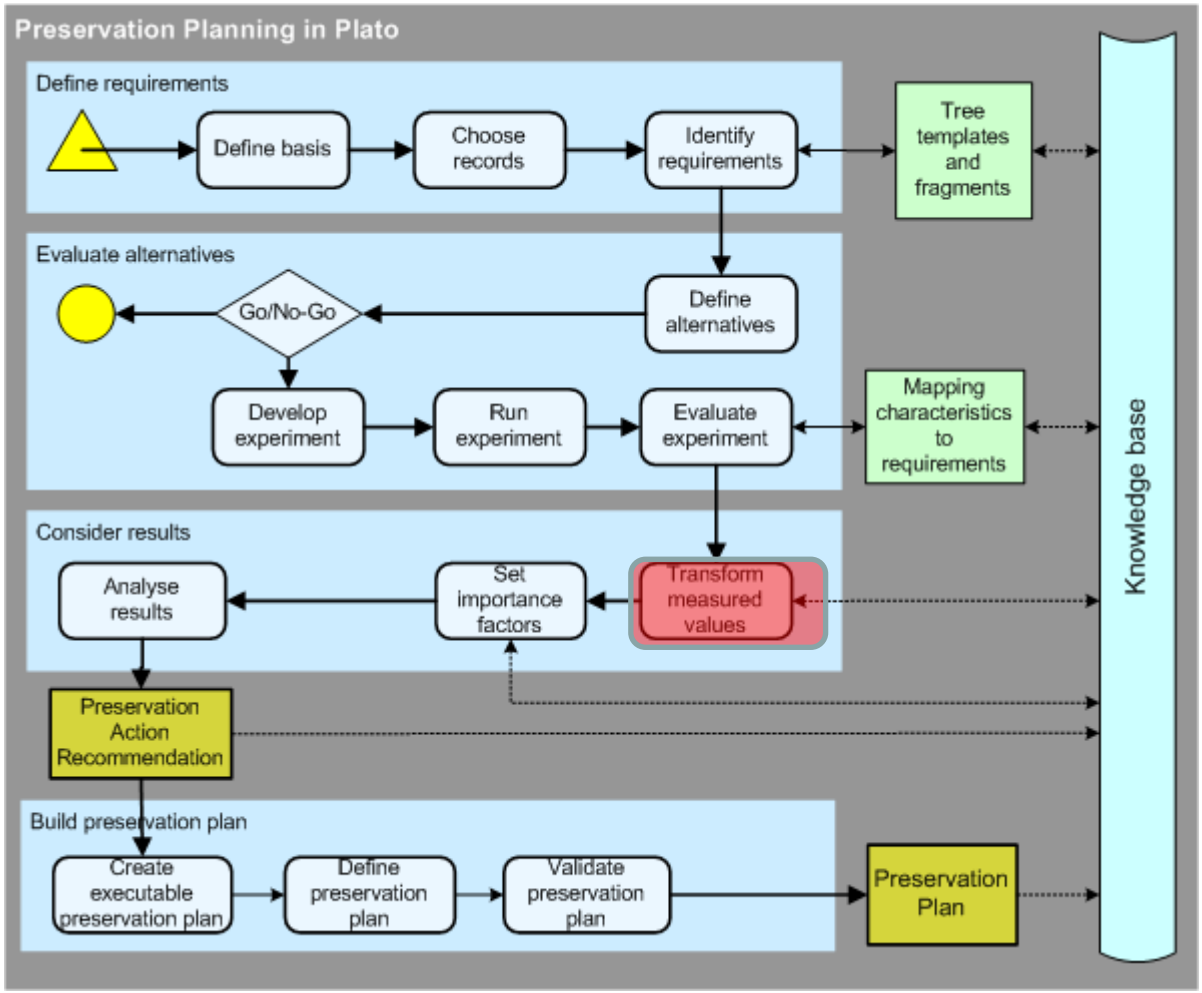
Process > Cost

Alternative			Comments
TIFF (tool A)	173	*	
	152	*	
TIFF (tool B)	100	*	
	88	*	
GIF (tool C)	140	*	
	128	*	
PNG (tool D)	79	*	
	80	*	

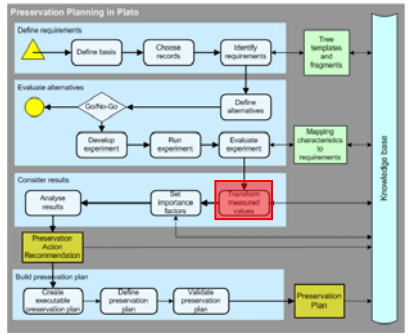
Image properties > Bits of colour depth

Alternative	Single result		Comments
TIFF (tool A)	32	bit *	
TIFF (tool B)	32	bit *	
GIF (tool C)	32	bit *	
PNG (tool D)	32	bit *	

PP Workflow

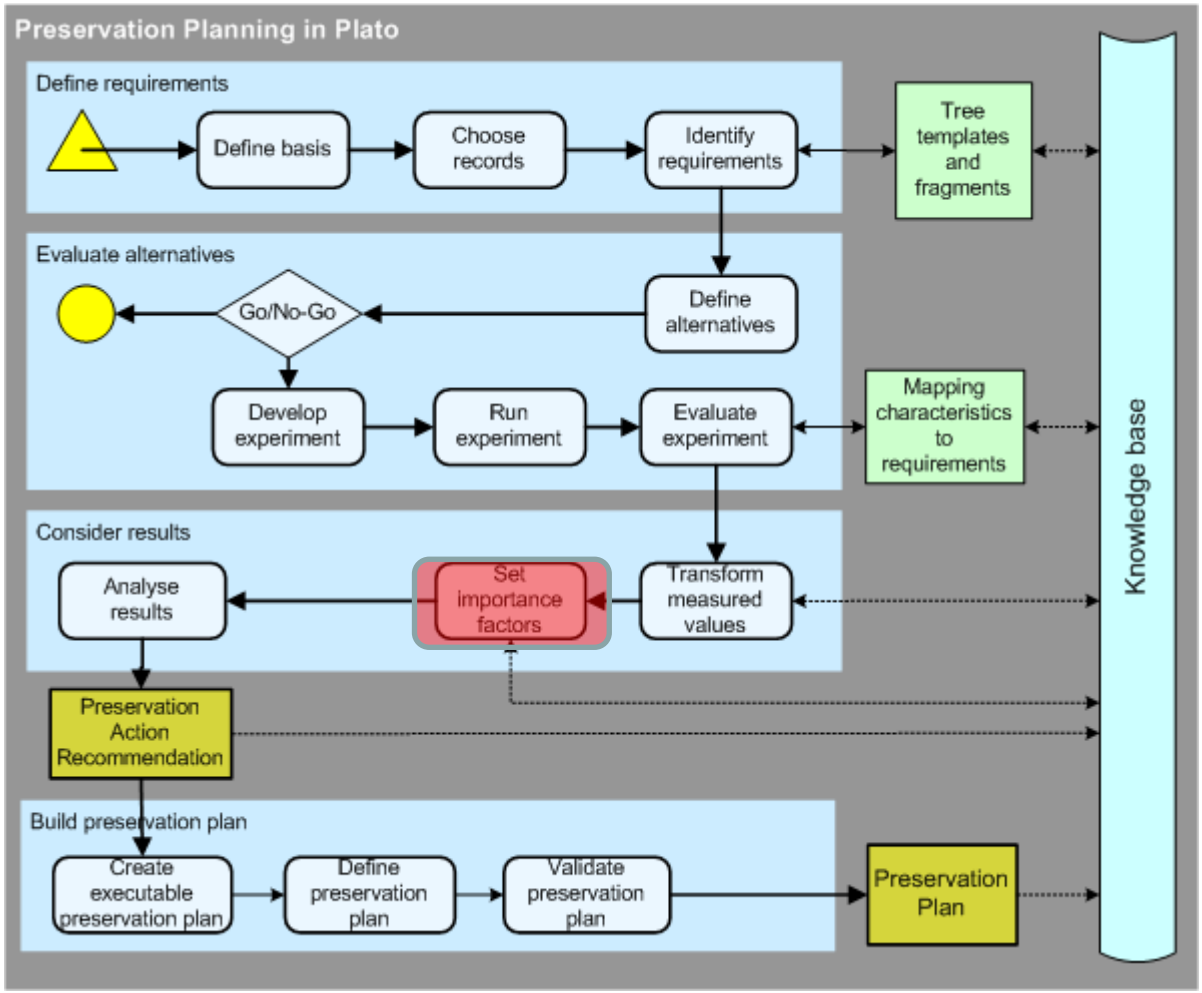

Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results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

Transform measured values

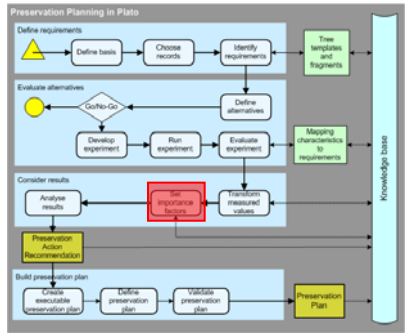


- Measures come in seconds, euro, bits, goodness values,…
- Need to make them comparable
- Transform measured values to uniform scale
- Transformation tables for each leaf criterion
- Linear transformation, logarithmic, special scale
- Scale 1-5 plus "not-acceptable"

PP Workflow

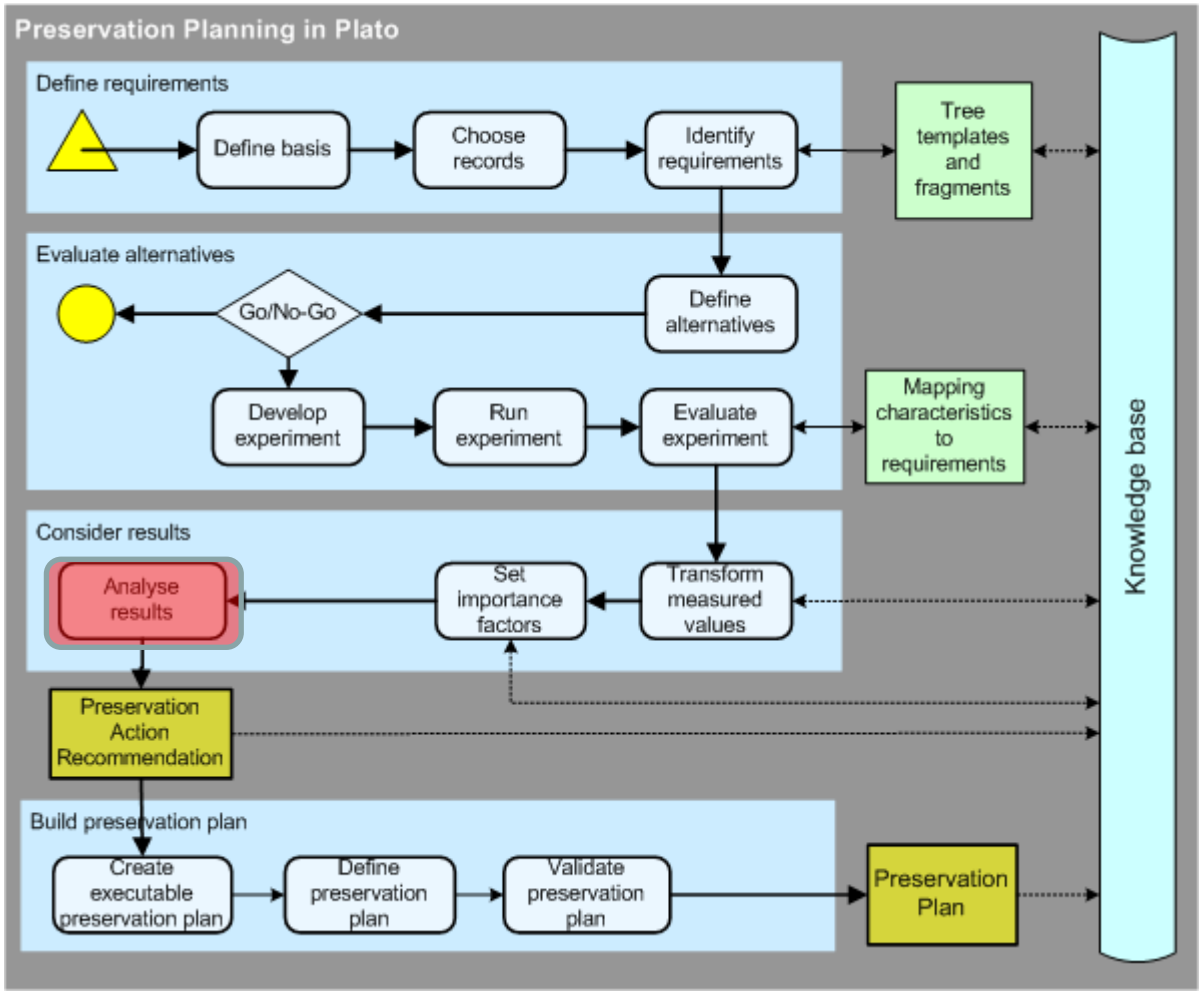

Preservation Planning in Plato
Define requirements
Define basis
Choose records
Identify requirements
Tree templates and fragments
Evaluate alternatives
Go/No-Go
Define alternatives
Develop experiment
Run experiment
Evaluate experiment
Mapping characteristics to requirements
Consider results
Analyse results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

Set Importance Factors

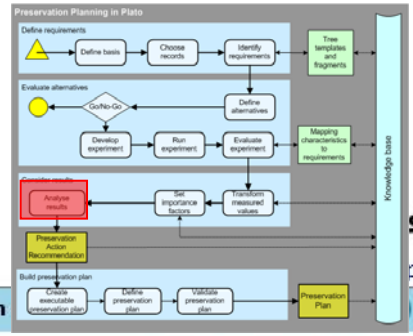


- Not all leaf criteria are equally important
- By default, weights are distributed equally
- Adjust relative importance of all siblings in a branch
- Weights are propagated down the tree to the leaves

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
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan
Knowledge base

Analyse Results



PLANETS Preservation Planning Tool *(Plato)*

Project | | Define Requirements :: Evaluate Requirements :: Consider Results | | | Minimalist test project in

Analyse Results

How do the aggregation mechanisms work?

Aggregation method: Multiplication

Select	Alternative
☑	PDF/A ToolA
☑	PDF/A ToolB

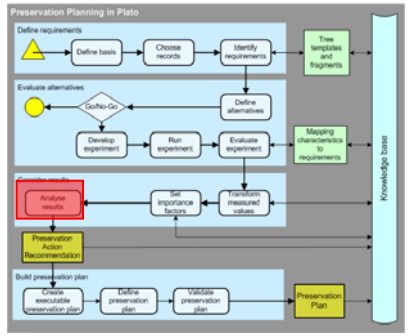
Show

Expand All | Collapse All

Minimalist root node

Focus	Name	Result
	▼Minimalist root node	PDF/A ToolA:2,86 PDF/A ToolB:0,00
X	▼Image properties	PDF/A ToolA:1,28 PDF/A ToolB:1,32
X	▼Amount of Pixel	PDF/A ToolA:3,50 PDF/A ToolB:4,00
X	▼Karma	PDF/A ToolA:1,15 PDF/A ToolB:0,00
X	▼Filesize (in Relation to Original)	PDF/A ToolA:1,31 PDF/A ToolB:1,38
X	▼A Single-Leaf	PDF/A ToolA:1,15 PDF/A ToolB:1,32
X	▼IntRange 0-10	PDF/A ToolA:1,28 PDF/A ToolB:1,25

Analyse results

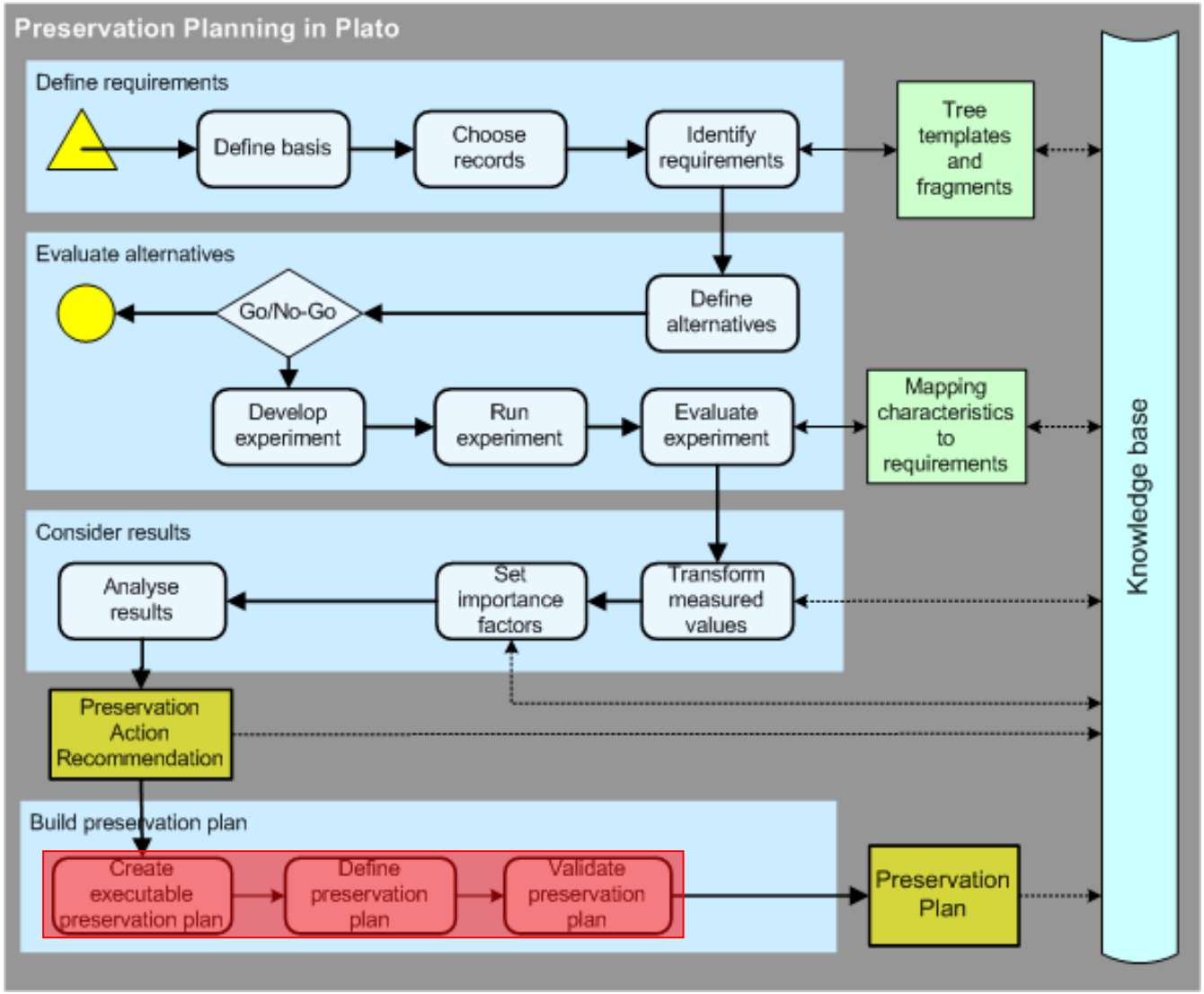


Example: Electronic documents

Alternative	Total Score Weighted Sum	Total Score Weighted Multiplication
PDF/A (Adobe Acrobat 7 prof.)	4.52	**4.31**
PDF (unchanged)	**4.53**	0.00
TIFF (Document Converter 4.1)	4.26	3.93
EPS (Adobe Acrobat 7 prof.)	4.22	3.99
JPEG 2000 (Adobe Acrobat 7 prof.)	4.17	3.77
RTF (Adobe Acrobat 7 prof.)	3.43	0.00
RTF (ConvertDoc 4.1)	3.38	0.00
TXT (Adobe Acrobat 7 prof.)	3.28	0.00

- Deactivation of scripting and security are knock-out criterium (PDF)
- RTF is weak in *Appearance* and *Structure*
- Plain text doesn't satisfy several minimum requirements

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
Knowledge base
Preservation Action Recommendation
Build preservation plan
Create executable preservation plan
Define preservation plan
Validate preservation plan
Preservation Plan

Outline

- Introduction
- Why do we need preservation planning?
- Preservation planning and Plato
- Bringing it all together and closing

Digital Preservation

What is a preservation plan?

- 10 Sections
  - Identification
  - Status
  - Description of Institutional Setting
  - Description of Collection
  - Requirements for Preservation
  - Evidence for Preservation Strategy
  - Cost
  - Trigger for Re-evaluation
  - Roles and Responsibilities
  - Preservation Action Plan

Preservation Plan Template

Preservation Planning with Plato

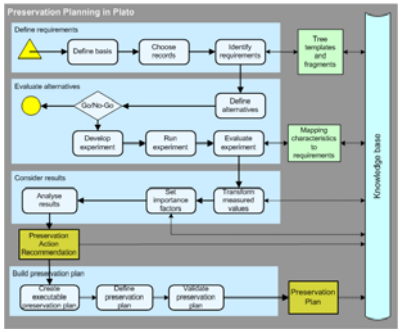
What we have now:

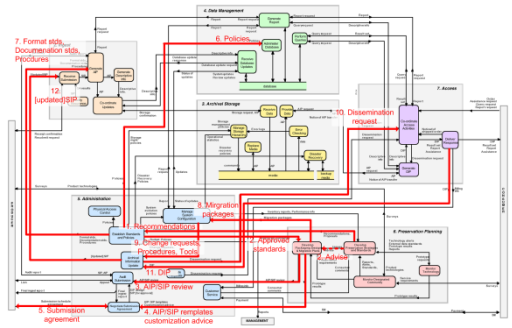
- Basic Preservation Plan:
  - *PDF: Preservation Plan.pdf*
  - *XML: Preservation Plan.xml*
- That was developed in a solid, repeatable and documented process
- That is optimal for the needs of a given institution and for the data at hand

Conclusions

- Physical preservation ensures longevity of resources
- Simple risk analysis reporting
- Preservation Planning to ensure “optimal” preservation
- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation
- Basis for well-informed, accountable decisions
- Follows recommendations of TRAC and nestor
- **Plato**:
  - Tool support to perform solid, well-documented analyses
  - Creates core preservation plan

Thank you!





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