Evaluating preservation strategies:
Decision support for preservation planning

Christoph Becker

"Tools and Trends" – The Hague, November 1, 2007

Decision support for preservation planning

- ❏ Provide a workflow and mechanism capable of translating influence parameters into best available strategy
  - ▪ Conduct case studies to identify sets of essential characteristics for different digital objects and requirements for preservation strategies (objective trees)
  - ▪ Developing a decision support software
    - • Plato – Planning Tool
    - • 1st version end of November 2007 (project internal)
    - • 2nd version publicly available, second half of 2008
  - ▪ Systematic procedure for evaluating preservation strategies
- ❏ Based on evaluation of strategies create preservation plan

Agenda

- Methodology
    - Workflow
    - Objective Trees
- Tool support: Plato
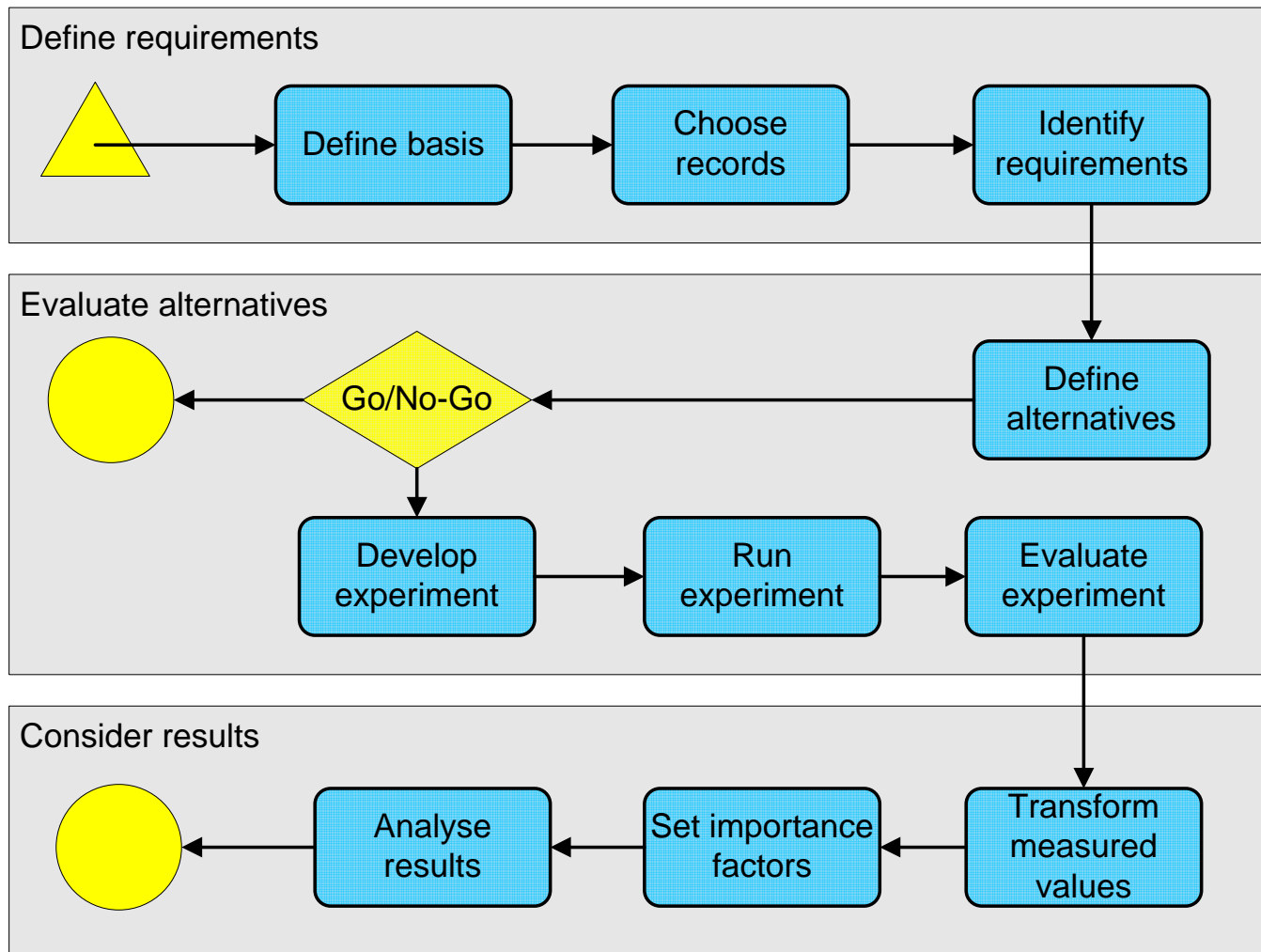- Integration of Planets concepts and services

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
- ❑ Evaluation of strategies on representative sample content according to specific requirements
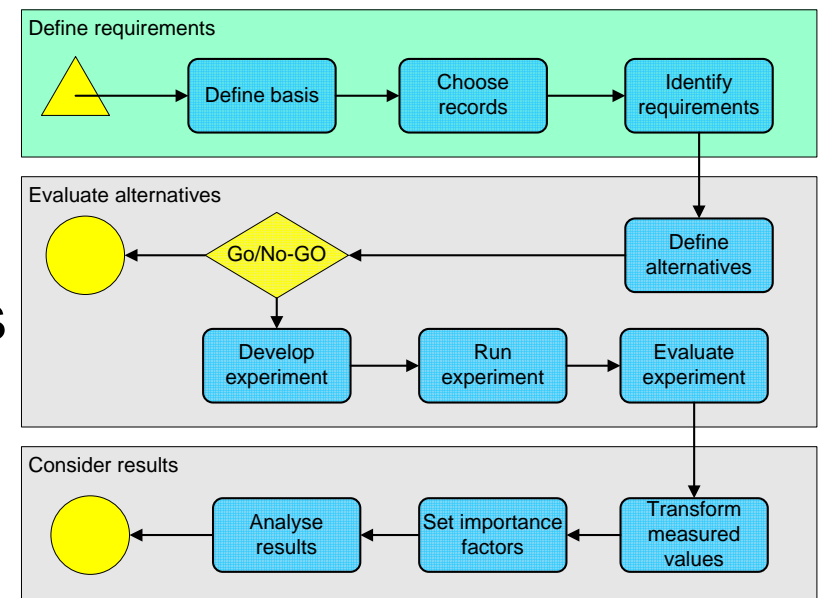
Workflow

Define requirements

Define basis → Choose records → Identify requirements

Evaluate alternatives

Go/No-Go ← Define alternatives

Develop experiment → Run experiment → Evaluate experiment

Consider results

Analyse results ← Set importance factors ← Transform measured values

Phase 1: Define requirements

1. **Define basis**
   – Describe Collection
   – Institutional settings

2. **Choose sample objects/records**
   – Representative for the objects in the collection
   – Right choice of samples is essential

3. **Define requirements**

Define requirements

Define basis → Choose records → Identify requirements

Evaluate alternatives

Go/No-GO ← Define alternatives

Develop experiment → Run experiment → Evaluate experiment

Consider results

Analyse results ← Set importance factors ← Transform measured values

Objective Tree

- Identify requirements and goals
- Tree structure
- Top-down or bottom-up
  - Start from high-level goals and break down to specific criteria
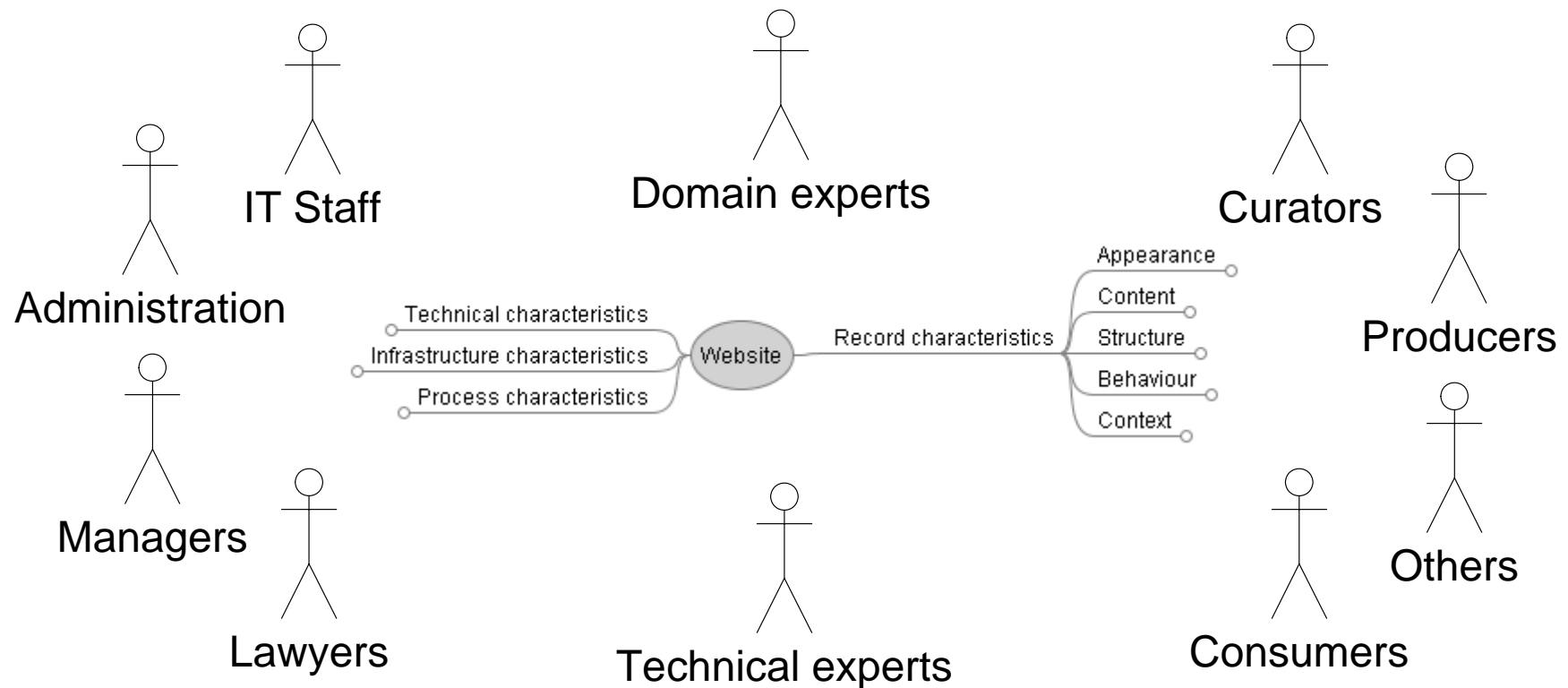 - Collect criteria and organize in tree structure
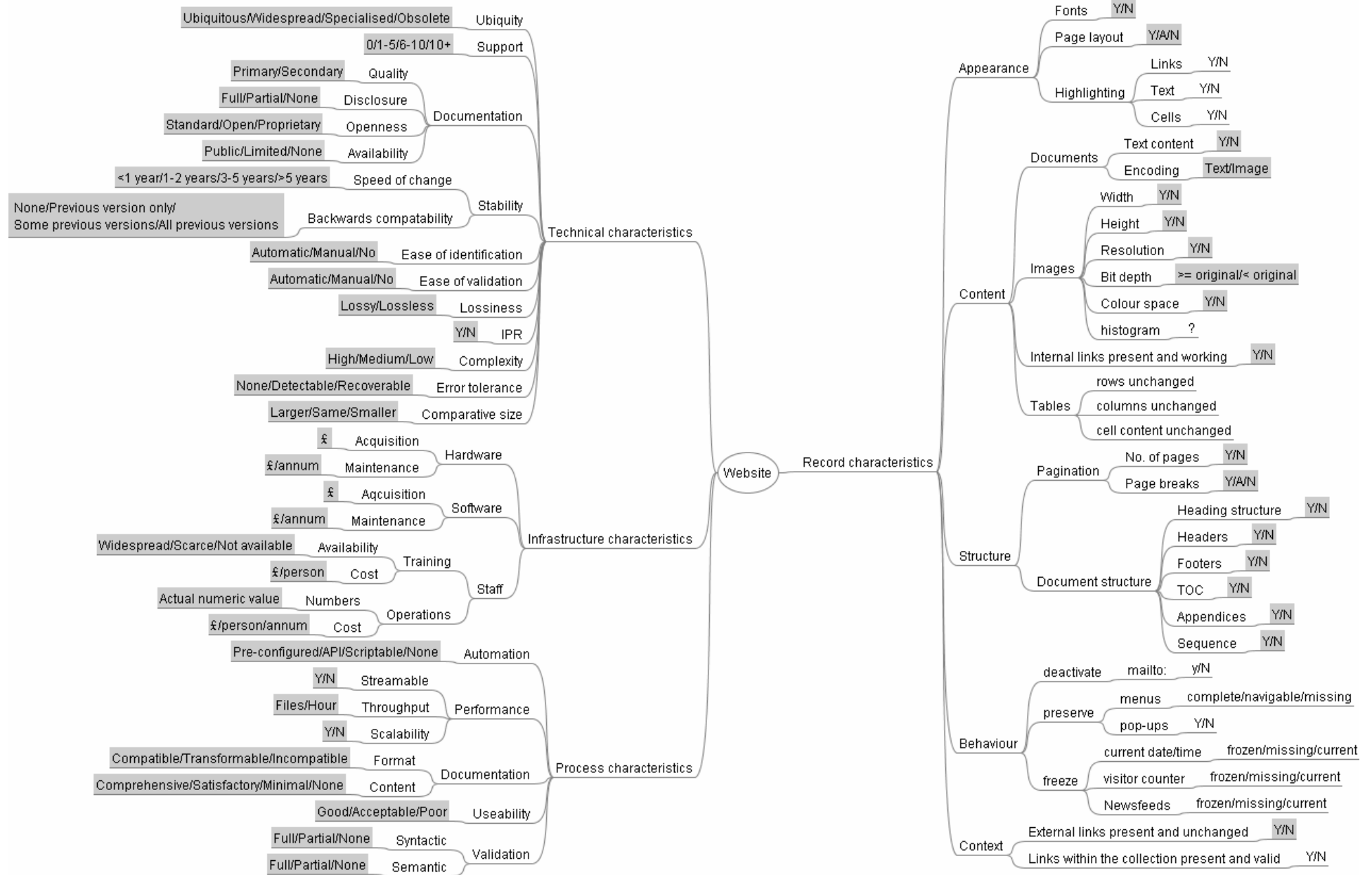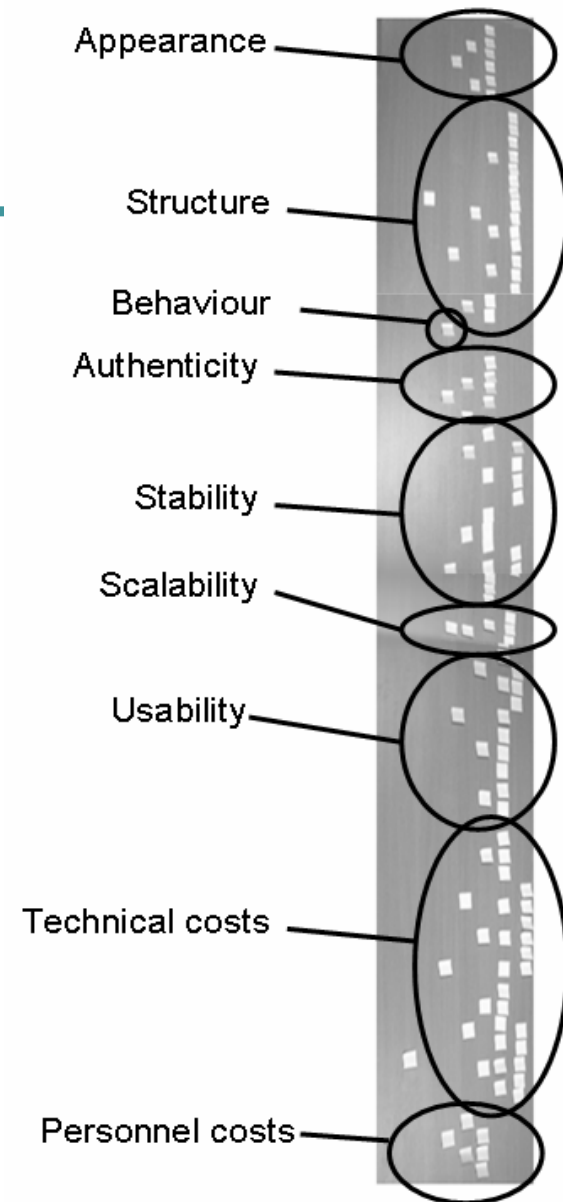
Influence Factors
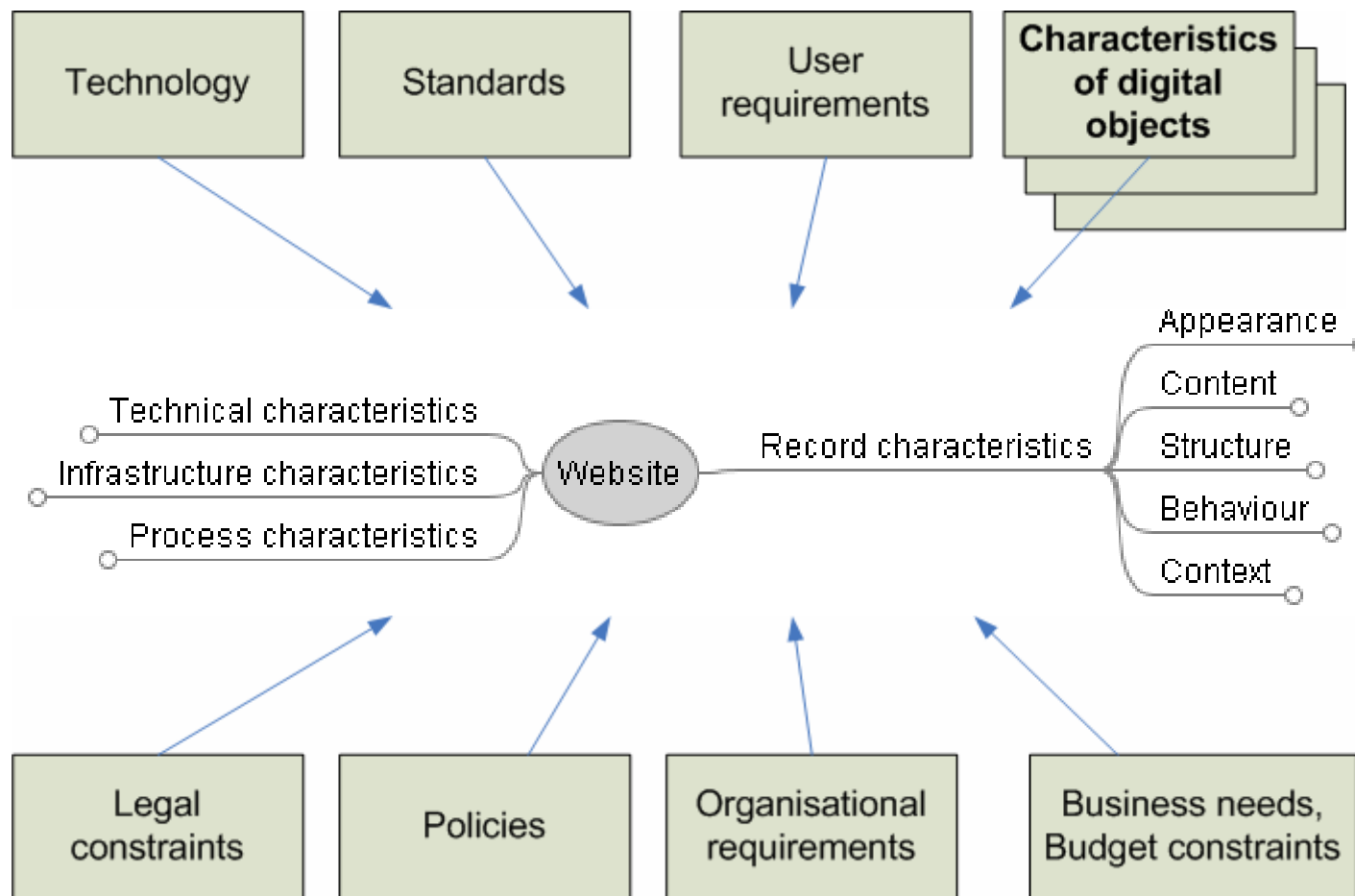
Stakeholders

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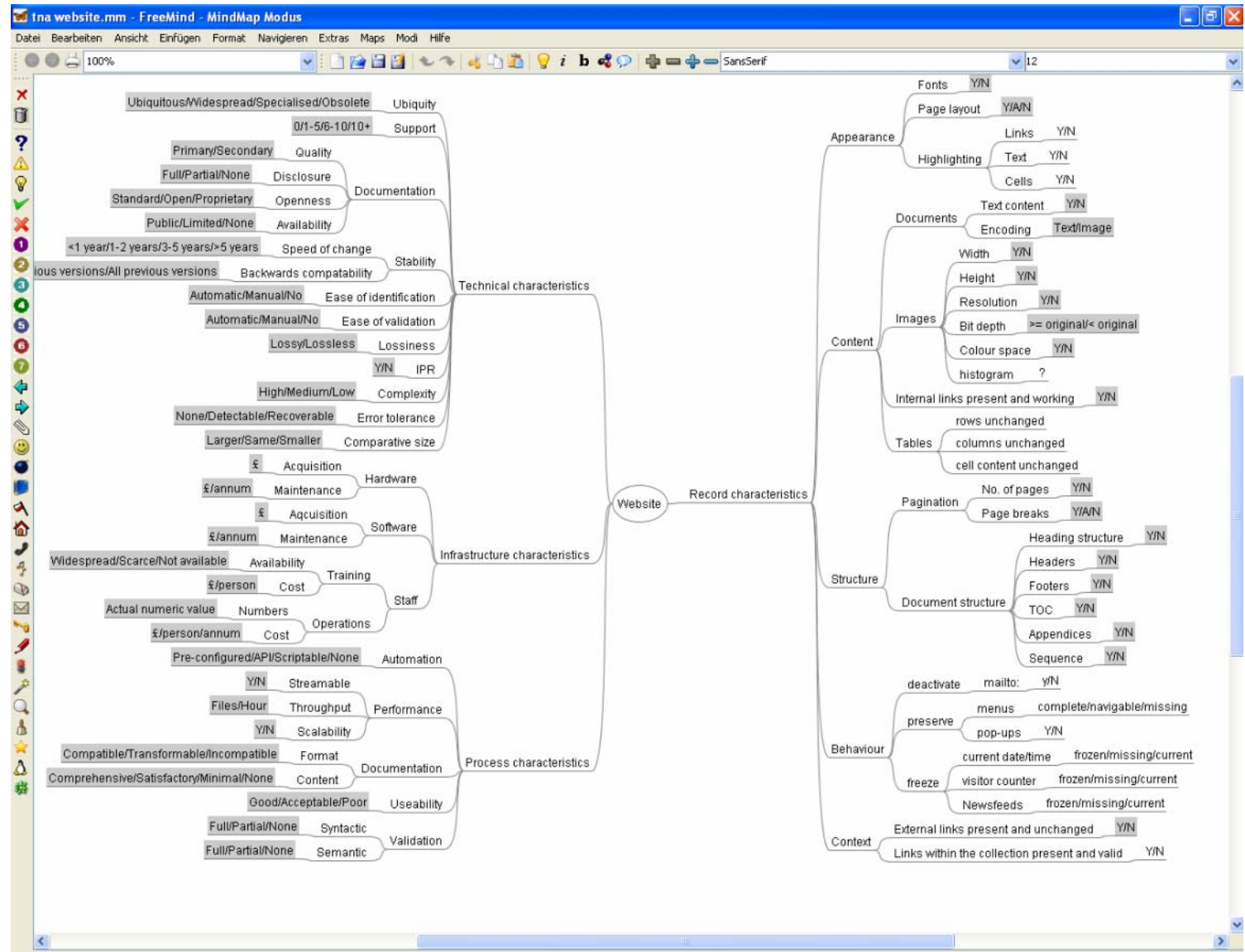 depending on the institutional context and the collection

IT Staff

Domain experts

Curators

Administration

Appearance
Content
Technical characteristics
Structure
Infrastructure characteristics
Website
Record characteristics
Behaviour
Process characteristics
Context

Producers

Managers

Others

Lawyers

Technical experts

Consumers

The Objective Tree

Ubiquitous/Widespread/Specialised/Obsolete — Ubiquity

0/1-5/6-10/10+ — Support

Primary/Secondary — Quality

Full/Partial/None — Disclosure

Standard/Open/Proprietary — Openness — Documentation

Public/Limited/None — Availability

<1 year/1-2 years/3-5 years/>5 years — Speed of change

None/Previous version only/ Some previous versions/All previous versions — Backwards compatability — Stability

Automatic/Manual/No — Ease of identification

Automatic/Manual/No — Ease of validation

Lossy/Lossless — Lossiness

Y/N — IPR

High/Medium/Low — Complexity

None/Detectable/Recoverable — Error tolerance

Larger/Same/Smaller — Comparative size

Technical characteristics

£ — Acquisition

£/annum — Maintenance — Hardware

£ — Aqcuisition

£/annum — Maintenance — Software

Widespread/Scarce/Not available — Availability

£/person — Cost — Training

Actual numeric value — Numbers — Staff

£/person/annum — Cost — Operations

Infrastructure characteristics

Pre-configured/API/Scriptable/None — Automation

Y/N — Streamable

Files/Hour — Throughput — Performance

Y/N — Scalability

Compatible/Transformable/Incompatible — Format

Comprehensive/Satisfactory/Minimal/None — Content — Documentation

Good/Acceptable/Poor — Useability

Full/Partial/None — Syntactic

Full/Partial/None — Semantic — Validation

Process characteristics

Website — Record characteristics

Appearance

Fonts — Y/N

Page layout — Y/A/N

Links — Y/N

Text — Y/N — Highlighting

Cells — Y/N

Content

Text content — Y/N — Documents

Encoding — Text/Image

Width — Y/N

Height — Y/N

Resolution — Y/N

Bit depth — >= original/< original — Images

Colour space — Y/N

histogram — ?

Internal links present and working — Y/N

rows unchanged

columns unchanged — Tables

cell content unchanged

Structure

No. of pages — Y/N — Pagination

Page breaks — Y/A/N

Heading structure — Y/N

Headers — Y/N

Footers — Y/N

TOC — Y/N — Document structure

Appendices — Y/N

Sequence — Y/N

Behaviour

deactivate — mailto: — y/N

menus — complete/navigable/missing — preserve

pop-ups — Y/N

current date/time — frozen/missing/current

visitor counter — frozen/missing/current — freeze

Newsfeeds — frozen/missing/current

Context

External links present and unchanged — Y/N

Links within the collection present and valid — Y/N

Analog…

Appearance

Structure

Behaviour

Authenticity

Stability

Scalability

Usability

Technical costs

Personnel costs

… or born-digital

Importing objective trees

Objective Tree

Assigning Scales

Phase 2: Evaluate Alternatives

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

PLANETS Preservation Planning Tool *(Plato)*

| Project | Define Requirements | Evaluate Requirements | Consider Results | | Project 'PP4 workshop - The National Archive' is in state EXPERIMENT_PERFORMED |

Evaluate Experiment

Expand All | Collapse All
Website > **Record characteristics**

Focus	Node
	▼Record characteristics
X	▶Appearance
X	▶Content
X	▶Structure
X	▼Behaviour
X	▶deactivate
X	▶preserve
X	▶freeze
X	▶Context

deactivate > mailto:

Alternative	first	second
solutionA	Yes ▾	No ▾
solutionB	Yes ▾	Yes ▾

preserve > menus

Alternative	first	second
solutionA	complete ▾	complete ▾
solutionB	navigable ▾	missing ▾

preserve > pop-ups

Alternative	first	second
solutionA	Yes ▾	Yes ▾
solutionB	No ▾	Yes ▾

freeze > current date/time

Alternative	first	second
solutionA	frozen ▾	frozen ▾
solutionB	missing ▾	frozen ▾

freeze > visitor counter

Alternative	first	second
solutionA	missing ▾	frozen ▾
solutionB	current ▾	current ▾

freeze > Newsfeeds

Alternative	first	second
solutionA	frozen ▾	frozen ▾
solutionB	frozen ▾	frozen ▾

Save   Proceed

PLANETS Preservation Planning Tool *(Plato)*

planets

| Project | Define Requirements | Evaluate Requirements | Consider Results | Project 'PP4 workshop - The National Archive' is in state EXPERIMENT_PERFORMED |

Evaluate Experiment

Expand All | Collapse All
Website

Focus	Node
	▼Website
X	▼Record characteristics
X	►Appearance
X	►Content
X	►Structure
X	▼Behaviour
X	►deactivate
X	►preserve
X	►freeze
X	►Context
X	►Technical characteristics
X	►Infrastructure characteristics
X	►Process characteristics

Process characteristics > Automation

Alternative	Single result
solutionA	Pre-configured ▼
solutionB	API ▼

Performance > Streamable

Alternative	Single result
solutionA	Yes ▼
solutionB	Yes ▼

Performance > Throughput

Alternative	first	Unit	second	Unit
solutionA	0.0	files per hour	0.0	files per hour
solutionB	0.0	files per hour	0.0	files per hour

Performance > Scalability

Alternative	Single result
solutionA	No ▼
solutionB	Yes ▼

Documentation > Format

Alternative	Single result
solutionA	Compatible ▼
solutionB	Incompatible ▼

Documentation > Content

Alternative	first	second
solutionA	Comprehensive ▼	Comprehensive ▼
solutionB	Satisfactory ▼	Satisfactory ▼

Process characteristics > Useability

Alternative	Single result

Phase 3:
Consider Results
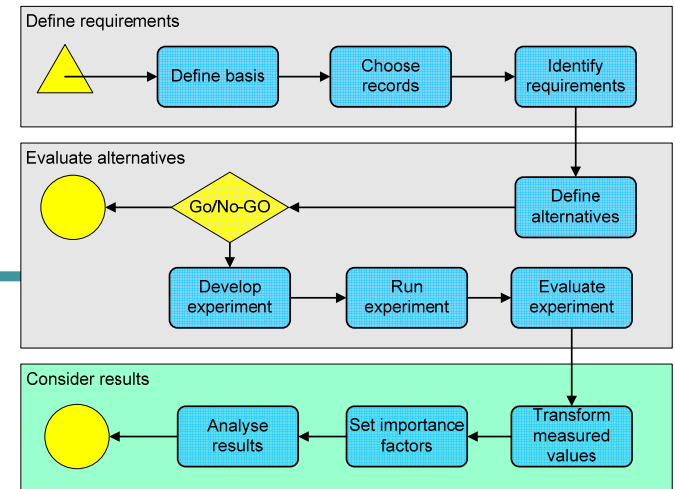
9. Transform measured values to a unified scale to make them comparable

10. Set importance factors to model the relative importance of siblings in each branch

11. Analyse results

Transform measured values
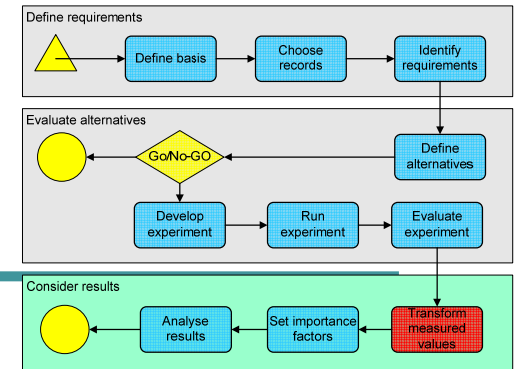
- Measures come in seconds, euro, bits, goodness values,…
- Need to make them comparable
- Transform measured values to uniform scale
- Transformation tables for each leaf criterion
- Scale 0-5 (0 is *unacceptable*)

Tool support

Set importance factors

- Branches are weighted equally by default

- Not all leaf criteria are equally important

- Adjust relative importance of all siblings in a branch

- Weights are propagated down the tree to the leaves

Balancing weights

Analyse Results

- Aggregate values
  - Multiply the transformed measured values in the leaf nodes with the leaf weights
  - Sum up the transformed weighted values over all branches of the tree
  - Performance values for each alternative
- Rank alternatives according to overall performance value at root
- Performance of each alternative
  - overall
  - for each sub-criterion (branch)
- Comparison of different alternatives
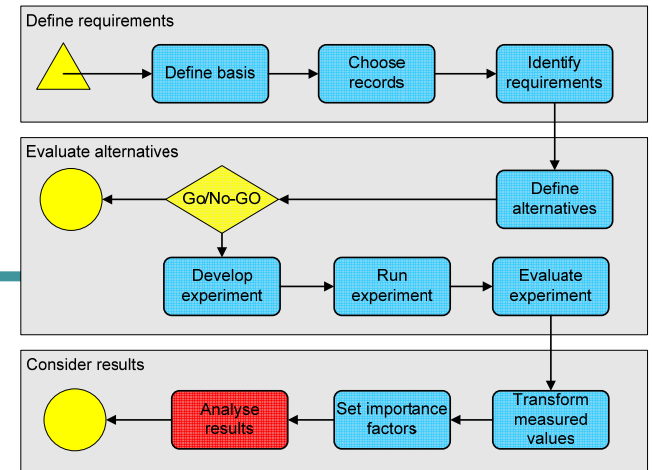
PLANETS Preservation Planning Tool *(Plato)*

[logout] [Export to XML] [help]

Project | **Define Requirements** | **Evaluate Requirements** | **Consider Results** | **Project 'Minimalist test project covering all features' is in state WEIGHTS_SET**

Analyse Results

Sum ▾

☑	PDF/A (Tool A)
☑	PDF/A (Tool B)

[Show]

Expand All | Collapse All

Minimalist root node

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

(Version 0.5) Institute of Software Technology and Interactive Systems: «office bears»   Quick Access:

Fertig

PLANETS Preservation Planning Tool *(Plato)*

Project | **Define Requirements** | **Evaluate Requirements** | **Consider Results** | **Project 'Minimalist test project covering all features' is in state WEIGHTS_SET**

Analyse Results

Multiplication ▾

| ☑ | PDF/A (Tool A) |
| ☑ | PDF/A (Tool B) |

Show

Expand All | Collapse All
Minimalist root node

| Focus	Name	Result
	▾ Minimalist root node	PDF/A (Tool A): 2,86 PDF/A (Tool B): 0,00
X	▸ Image properties	PDF/A (Tool A): 1,28 PDF/A (Tool B): 1,32
X	▾ Karma	PDF/A (Tool A): 1,15 PDF/A (Tool B): 0,00
X	▾ Filesize (in Relation to Original)	PDF/A (Tool A): 1,31 PDF/A (Tool B): 1,38
X	▾ A Single-Leaf	PDF/A (Tool A): 1,15 PDF/A (Tool B): 1,32
X	▾ IntRange 0-10	PDF/A (Tool A): 1,28 PDF/A (Tool B): 1,25

The technical side

- Java Enterprise application
- Planets Application Server based on JBoss 4.0.5
- JBoss Seam 1.2.1
- Java Server Faces, Facelets
  - AJAX-enabled component libraries
    - Apache Trinidad
    - JBoss RichFaces, AJAX4JSF
  - EJB 3 (Hibernate)
  - Database: Apache Derby (exchangeable)
  - XML export and import

Integrating Planets concepts and services
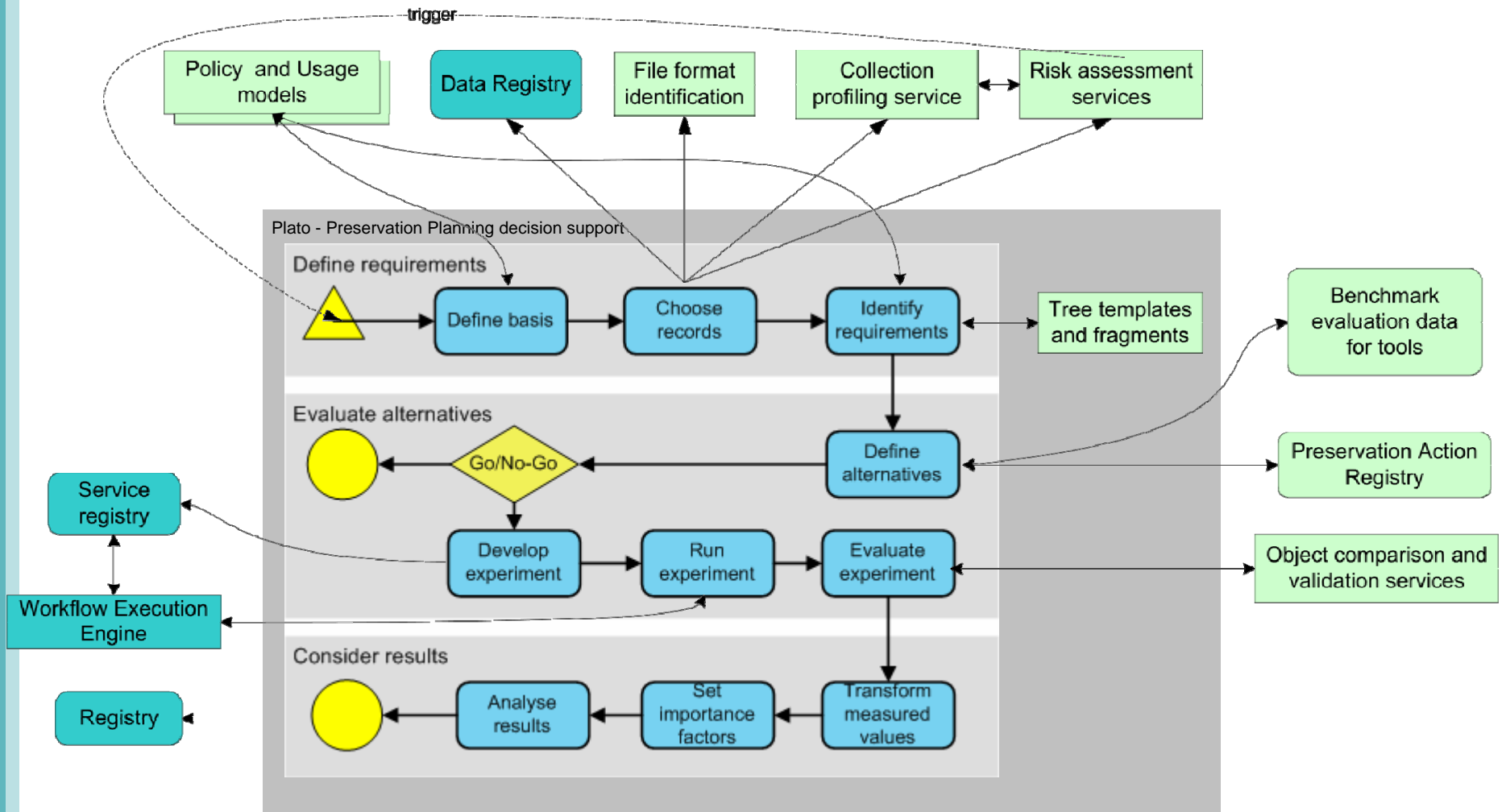
Summary

- ...
- ...
- ...

Thank you very much for your attention.

becker@ifs.tuwien.ac.at
www.planets-project.eu