The Preservation planning workflow
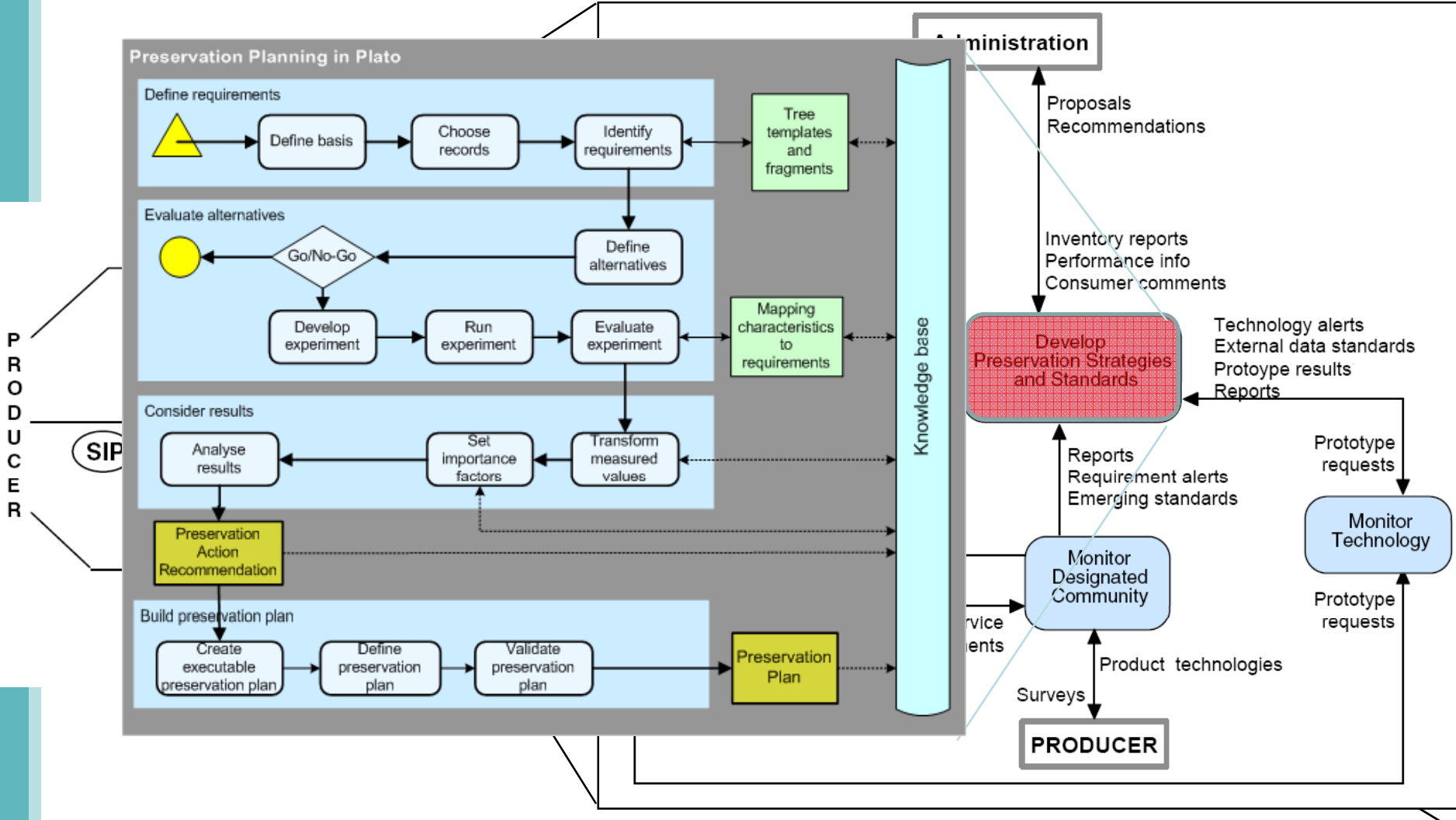
Andreas Rauber

Vienna University of Technology
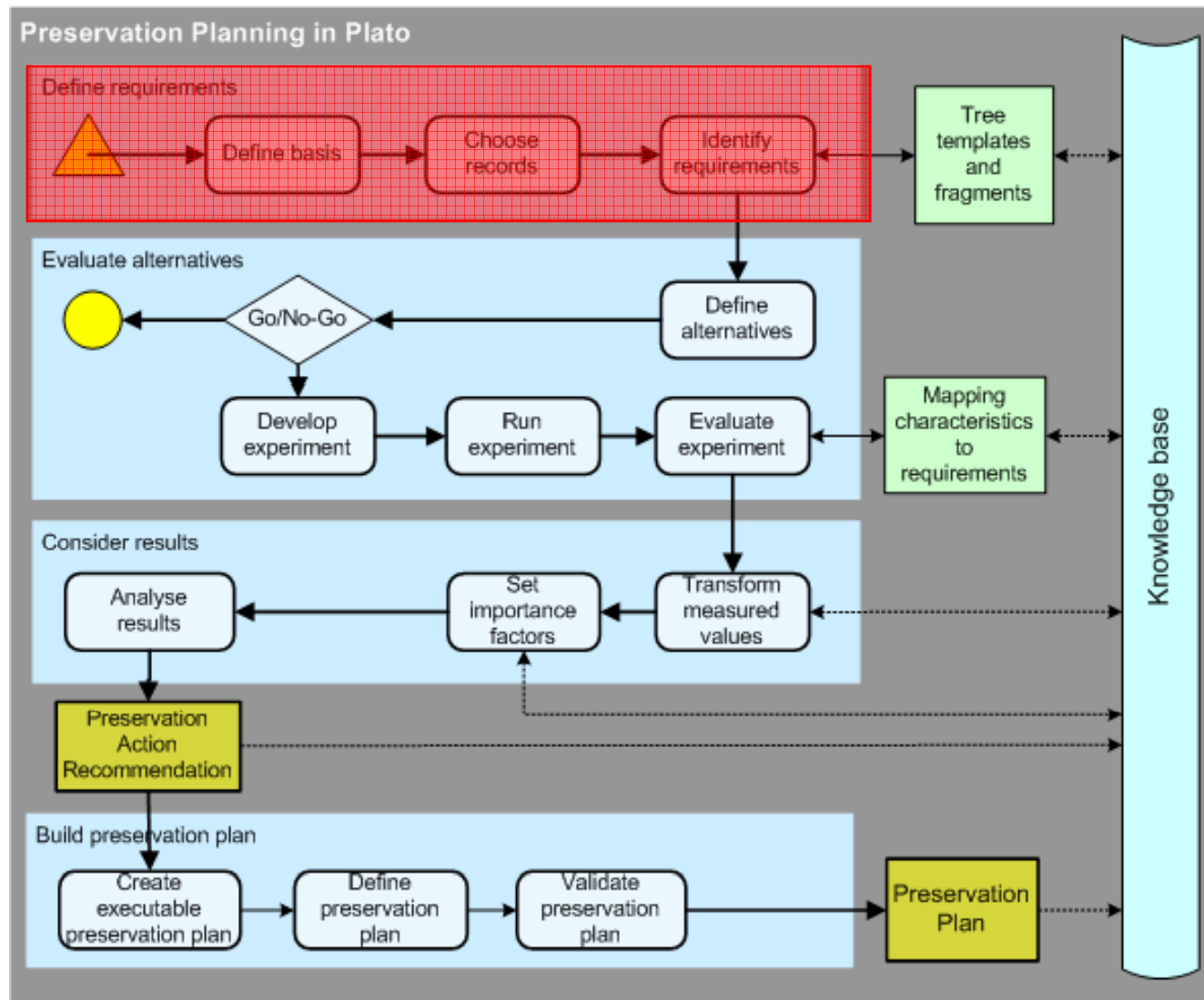www.ifs.tuwien.ac.at/~andi

Motivation

➢ Several preservation strategies developed

➢ How do you know what is most suitable?

  ▪ Right choice depends on the needs (no clear preferences)

➢ How to measure and evaluate the results of each preservation strategy?

➢ What are the requirements?

➢ How to define a controlled and trusted environment and a procedure for applying or testing preservation strategies?
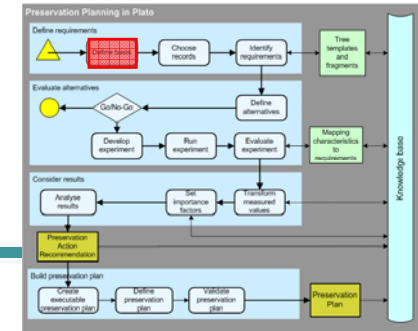
Preservation Planning

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
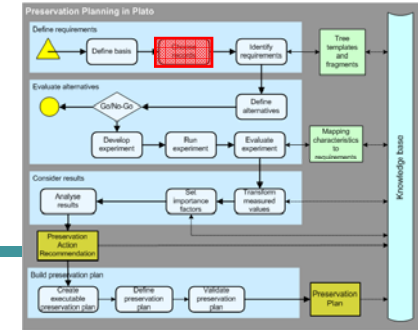
Preservation Plan

Knowledge base

Define basis

➢ What are the objects?

➢ What are the essential characteristics?

  ▪ Content, context, structure, form and behaviour

➢ What are the requirements?

  ▪ Authenticity, reliability, integrity, useability

  ▪ Metadata (for different purposes)

➢ What preservation strategies will be applied and

  evaluated?

Choose objects/records

- ➢ Different object types

  - ▪ Text documents, audio, video, e-mail, multimedia, databases, data sets, ...

- ➢ Distinction between

  - ▪ Physical (technical) object = computer file, and

  - ▪ The intellectual object (e.g. what is shown on the screen)

- ➢ Choice of objects affects the evaluation

Identify requirements
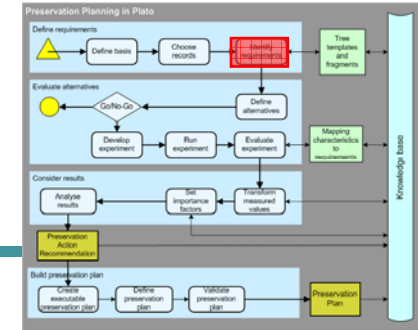
- ➢ Define all relevant goals and characteristics
  (high-level, detail) with respect to a given application domain

- ➢ 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

Identify requirements
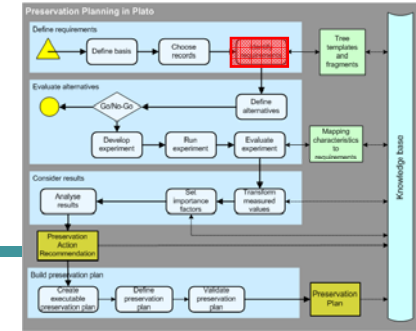
Appearance

Structure

Behaviour

Authenticity

Stability

Scalability

Usability

Technical costs

Personnel costs

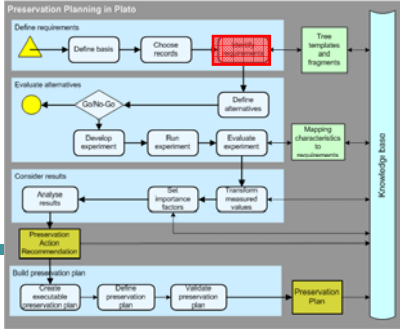Analog…
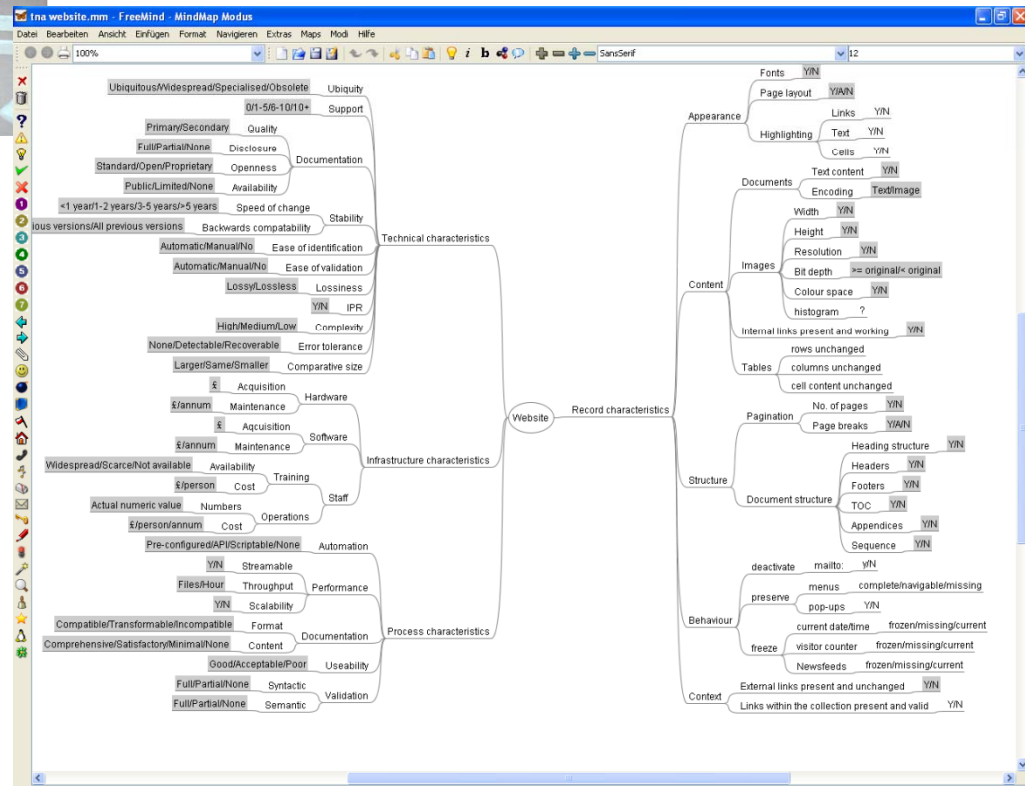
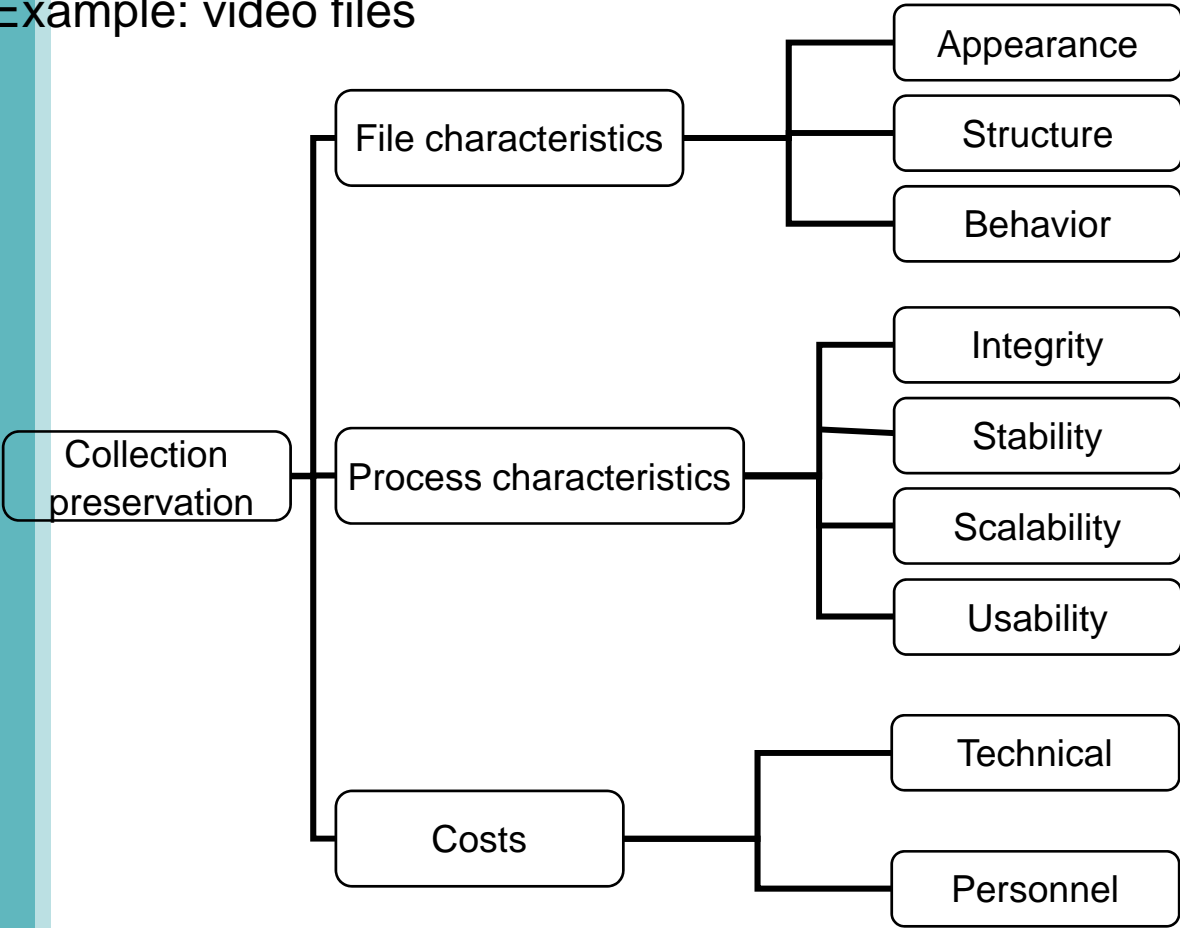… or born-digital

Identify requirements

Example: video files

```
Collection
preservation
    ├── File characteristics
    │       ├── Appearance        e.g. Color-proof, Frame rate,, ..
    │       ├── Structure         e.g. Original compression, ..
    │       └── Behavior          e.g. Subtitles, …
    │
    ├── Process characteristics
    │       ├── Integrity         e.g. File format verification
    │       ├── Stability         e.g. Durability
    │       ├── Scalability       e.g. Format scalability
    │       └── Usability         e.g. Complexity, Functionality
    │
    └── Costs
            ├── Technical         e.g. Hardware, Software
            └── Personnel         e.g. Enrolment, Maintenance
```

Assign measurable units
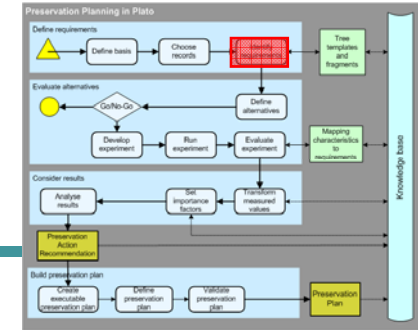
- ❑ Assign measurable effect to each leaf
  - ▪ Ensure that leaf criteria are objectively (and automatically) measurable
    - ▪ Seconds/Euro per object
    - ▪ Bits of color depth

    ...
  - ▪ Subjective scales where necessary
    - ▪ diffusion of file format
    - ▪ amount of (expected) support

    ...
- ❑ No limitations on the use of scale

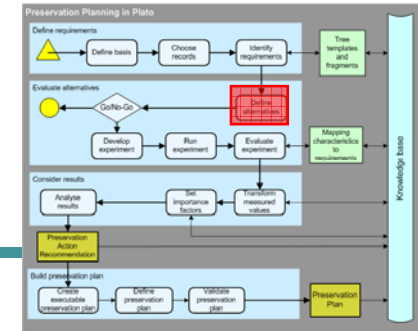# Identify requirements

PP Workflow

Define alternatives
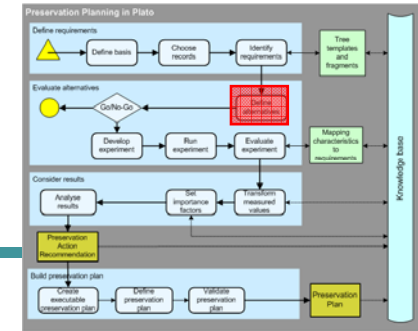
➢ Given the type of objects and requirements, what strategies would be best suitable/are possible?

- Migration
- Emulation
- Both
- Other?

➢ For each alternative precise definition of

- Which tool (OS, version,...)
- Which functions of the tool in which order
- Which parameters

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
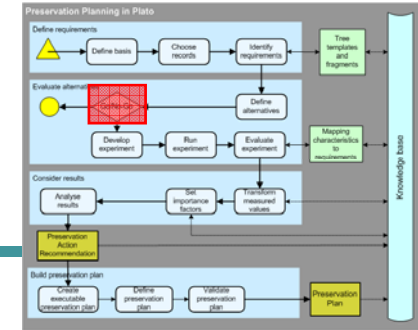
- ➢ Deliberate step for taking a decision whether it will be useful and cost-effective to continue the procedure, given

  - ▪ 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 redressed or not?

Develop experiment
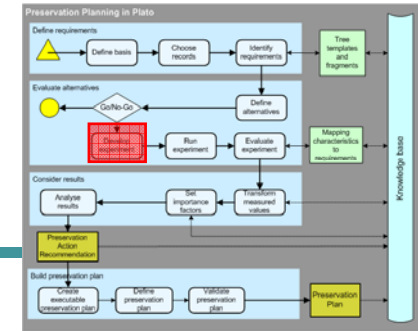
- ➤ Formulate for each evaluation or experiment or preservation process detailed

  - Development plan

    - steps to build and test software components

    - procedures and preparation

    - parameter settings for integrating preservation services

  - Test plan (mechanisms how to)

 - Evaluation/experiment plan (workflow/sequence of activities)

Run experiment
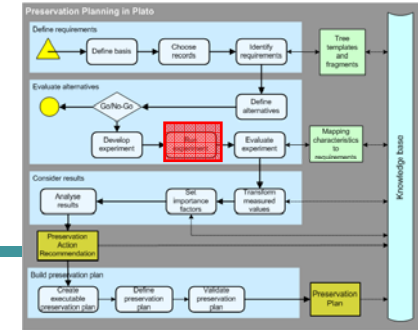
➢ Before conducting an evaluation or running an experiment, the experiment process as designed has to be tested

▪ It may lead to re-design or even termination of the evaluation/ experiment process

➢ The results will be evaluated in the next stage

➢ The whole process needs to be documented

Evaluate experiment
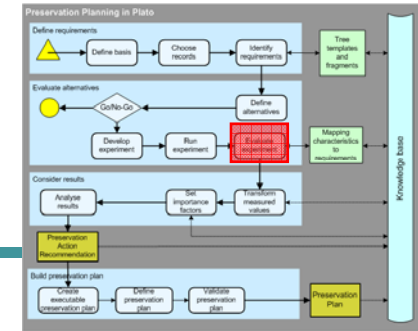
➢Evaluate the outcome of each alternative for each leaf of the objective tree

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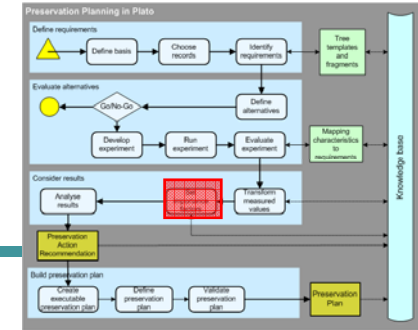
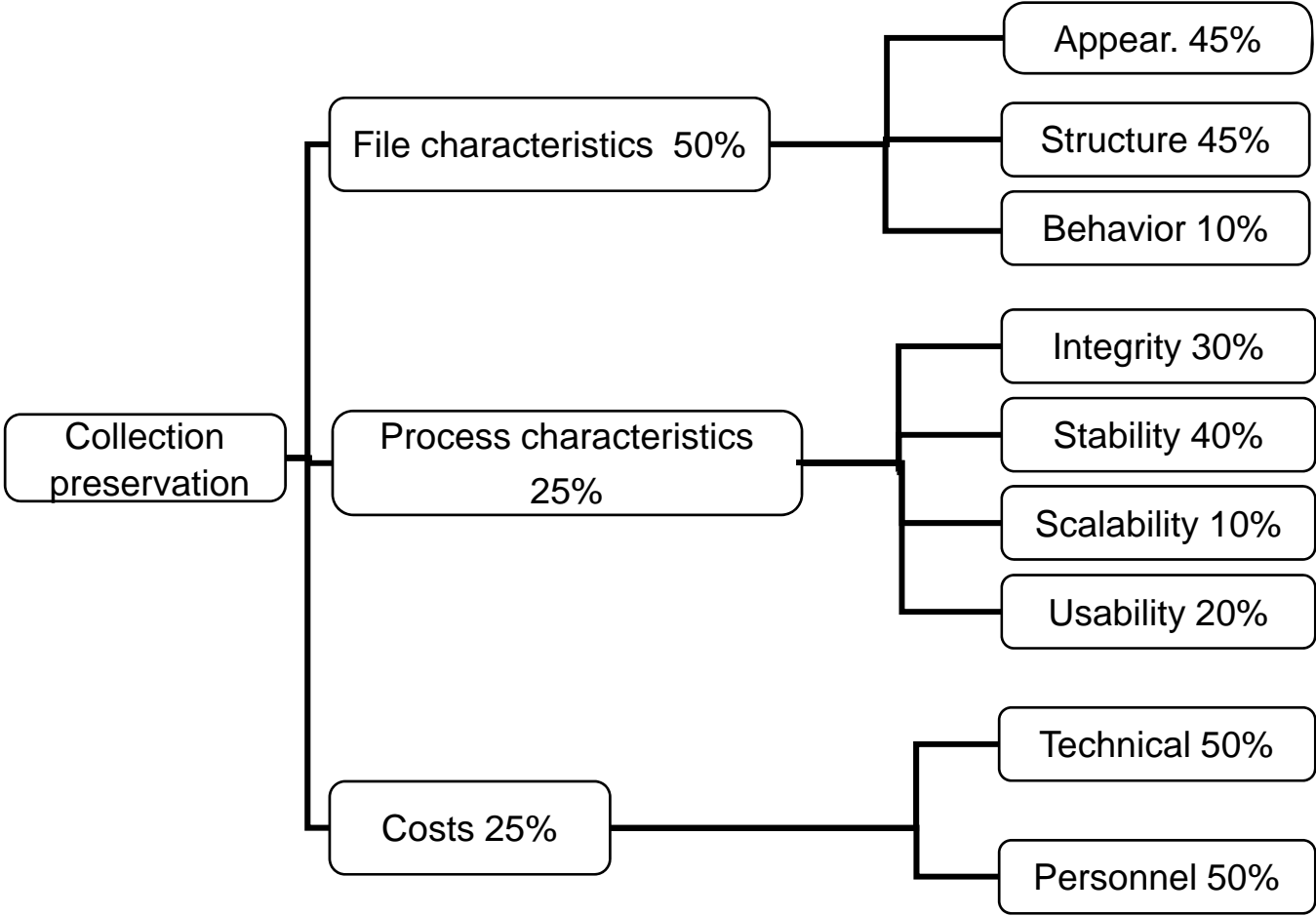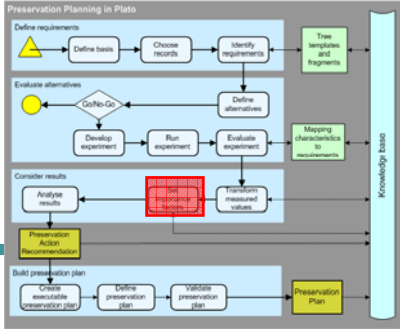# PP Workflow
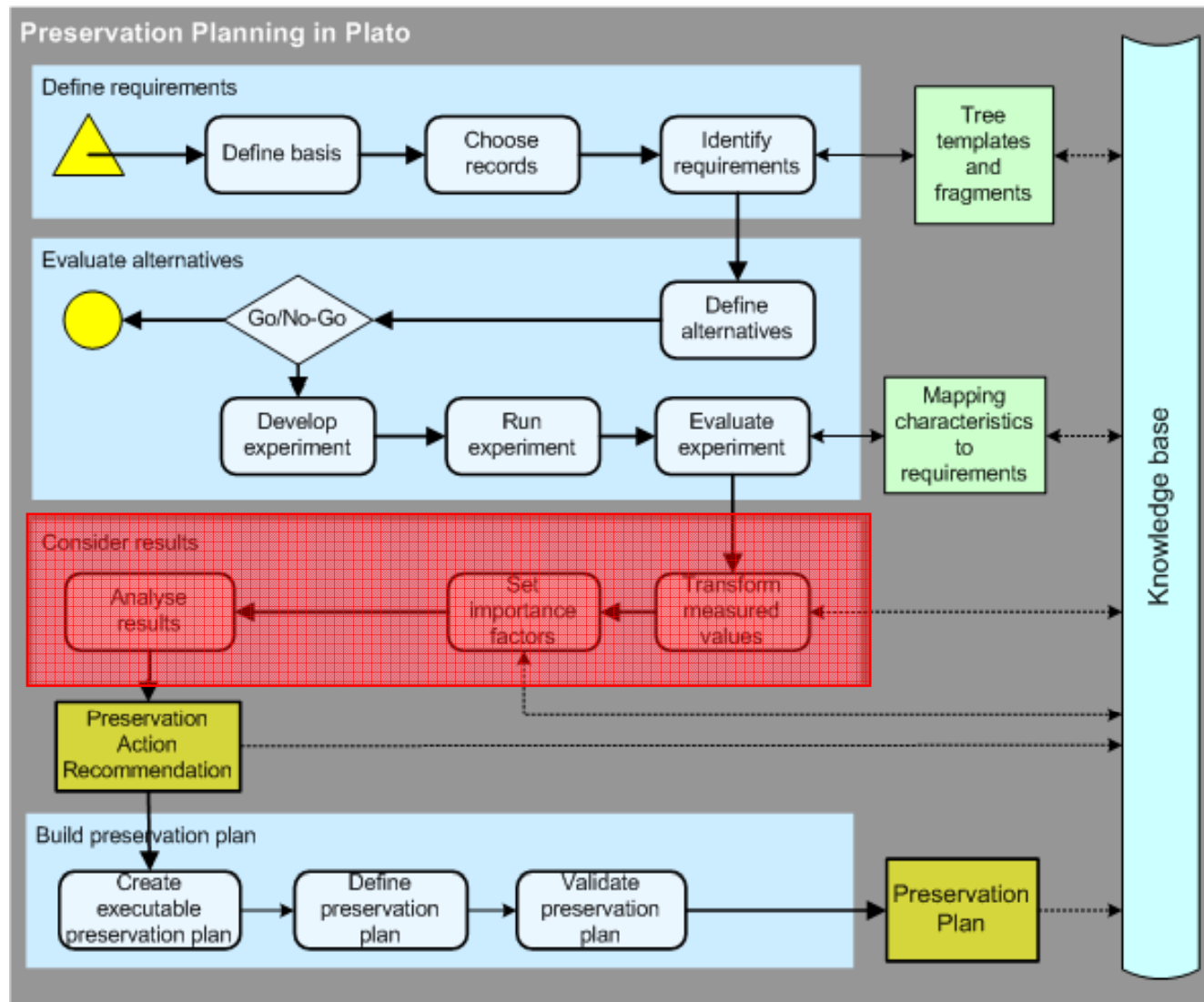
Transform measured values
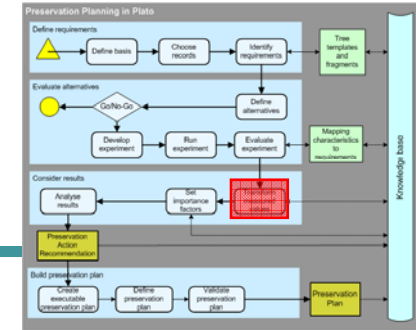
- ❏ Measures come in seconds, euro, bits, goodness values,…
- ❏ Need to make them comparable
- ❏ Transform measured values to uniform scale
- ❏ Transformation tables for each leaf criterion
- ❏ Linear transformation, logarithmic, special scale
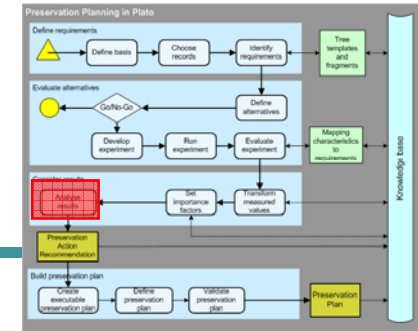- ❏ Scale 1-5 plus "not-acceptable"

Set importance factors

❑ Definition which criteria are more important

❑ Depends on individual preferences and requirements

❑ Adaptation for each implementation

❑ High influence on the final ranking

❑ Aggregation of weights

Set importance factors


```
                                    ┌─────────────────┐
                                    │  Appear. 45%    │
                                    └─────────────────┘
            ┌──────────────────────┐┌─────────────────┐
            │ File characteristics │││  Structure 45%  │
            │        50%           │└─────────────────┘
            └──────────────────────┘┌─────────────────┐
                                    │  Behavior 10%   │
                                    └─────────────────┘

                                    ┌─────────────────┐
                                    │  Integrity 30%  │
                                    └─────────────────┘
                                    ┌─────────────────┐
            ┌──────────────────────┐│  Stability 40%  │
┌──────────┐│ Process characteristics└─────────────────┘
│Collection││        25%           │┌─────────────────┐
│preservation└─────────────────────┘│ Scalability 10% │
└──────────┘                         └─────────────────┘
                                    ┌─────────────────┐
                                    │  Usability 20%  │
                                    └─────────────────┘

                                    ┌─────────────────┐
                                    │  Technical 50%  │
            ┌──────────────────────┐└─────────────────┘
            │     Costs 25%        │┌─────────────────┐
            └──────────────────────┘│  Personnel 50%  │
                                    └─────────────────┘
```

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

- Creates performance values for each alternative on each of the sub-criteria identified

Analyse results

Analyse results

> Single performance value for each alternative to rank the alternatives

> Single performance values for each alternative for each sub-set of criteria to identify the best combination of alternatives

> Sensitivity Analysis: Analysis of the influence of small changes in the weight on the final value

> Basis for making Informed, well-documented, repeatable, accountable decisions

Analyse results
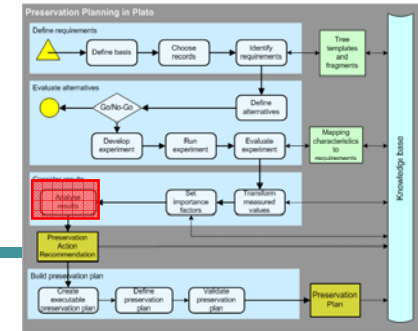
- ❑ Rank alternatives according to overall utility value at root

- ❑ Performance of each alternative
  - ▪ overall
  - ▪ for each sub-criterion (branch)

- ❑ Allows performance measurement of combinations of strategies

- ❑ Final sensitivity analysis against minor fluctuations in
  - ▪ measured values
  - ▪ importance factors

Consider results
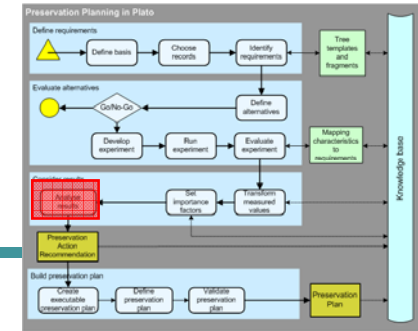
> The review of the results may help to refine

- The evaluation process/procedure

- The preservation planning environment itself

- The evaluation metrics

- Understanding of the essential characteristics of the objects,

- and identify further evaluations, experiments

> The review should take into account all previous work done in the preservation planning environment

> The review should look at both the technical and intellectual aspects of digital objects

PP Workflow
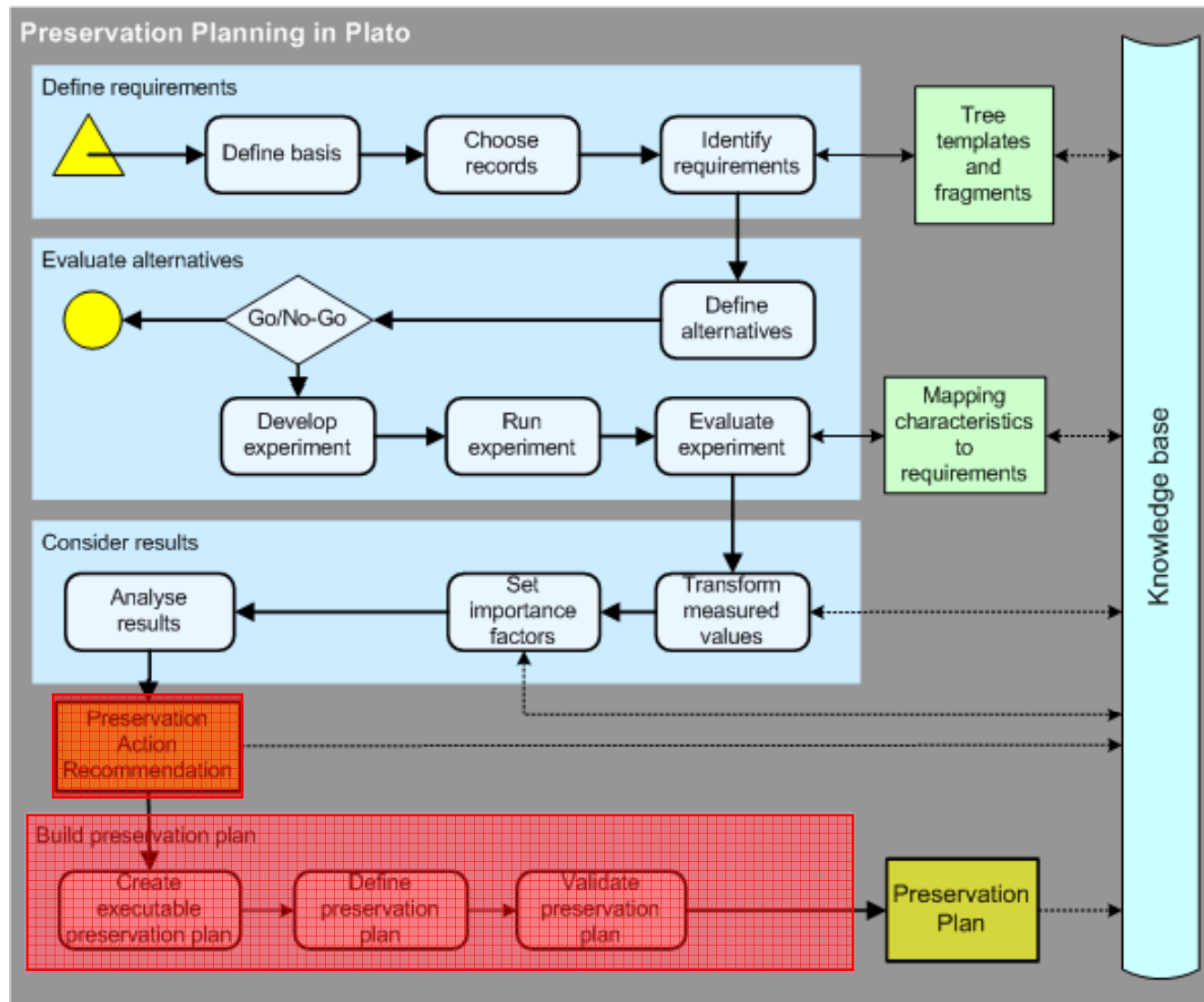
Build Preservation Plan
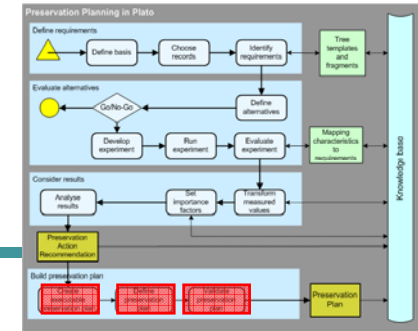
- ➢ Create executable elements of preservation plan
  - ▪ Sequence of preservation actions to call, parameters, …
  - ▪ Automatic steps + manual interventions where required
  - ▪ Automatic verification of results during deployment

- ➢ Define preservation plan
  - ▪ Create PP based on evidence produced during the PP process
  - ▪ Verify completeness of PP

- ➢ Seek approval and validation of PP
  - ▪ Management activity according to OAIS
  - ▪ Sign and deploy

Conclusions

- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation for informed and accountable decisions
- Set of templates to assist institutions
- Generic workflow that can easily be integrated in different institutional settings
- **Plato**:
  Tool support to perform solid, well-documented analyses
- Provides basic preservation plan

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

Preservation Planning Workflow

Thank you very much for your attention

www.planets-project.eu

rauber@ifs.tuwien.ac.at
www.ifs.tuwien.ac.at/~rauber