Plato: A Service Oriented Decision Support System for Preservation Planning

Christoph Becker, Hannes Kulovits, Andreas Rauber, Hans Hofman

ACM/IEEE Joint Conference on Digital Libraries (JCDL 2008)

Pittsburgh, PA, USA. June 16-20, 2008

Outline

- Digital Preservation and Preservation Planning
  - Evaluation of potential actions
- Plato: The Planets Preservation Planning Tool
  - Underlying methodology and workflow
  - Service discovery and integration
    - Characterisation
    - Preservation action
  - State of development and roadmap
- Current and future work

The Longevity of Digital Objects

- We create, shape, and exchange information digitally
- Digital objects need technical environment to “function”
- Heterogeneity and complexity of formats and environments and the speed of technological change make long-term access a challenge

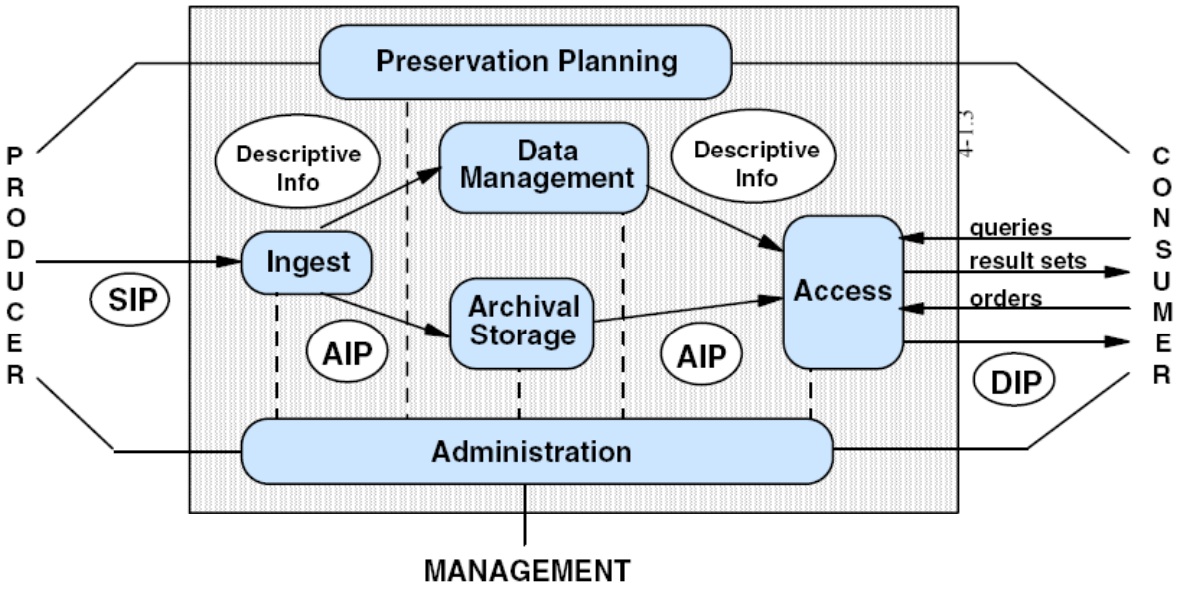
- Dominant types of preservation actions:
  - Migration
  - Emulation

Evaluating preservation strategies

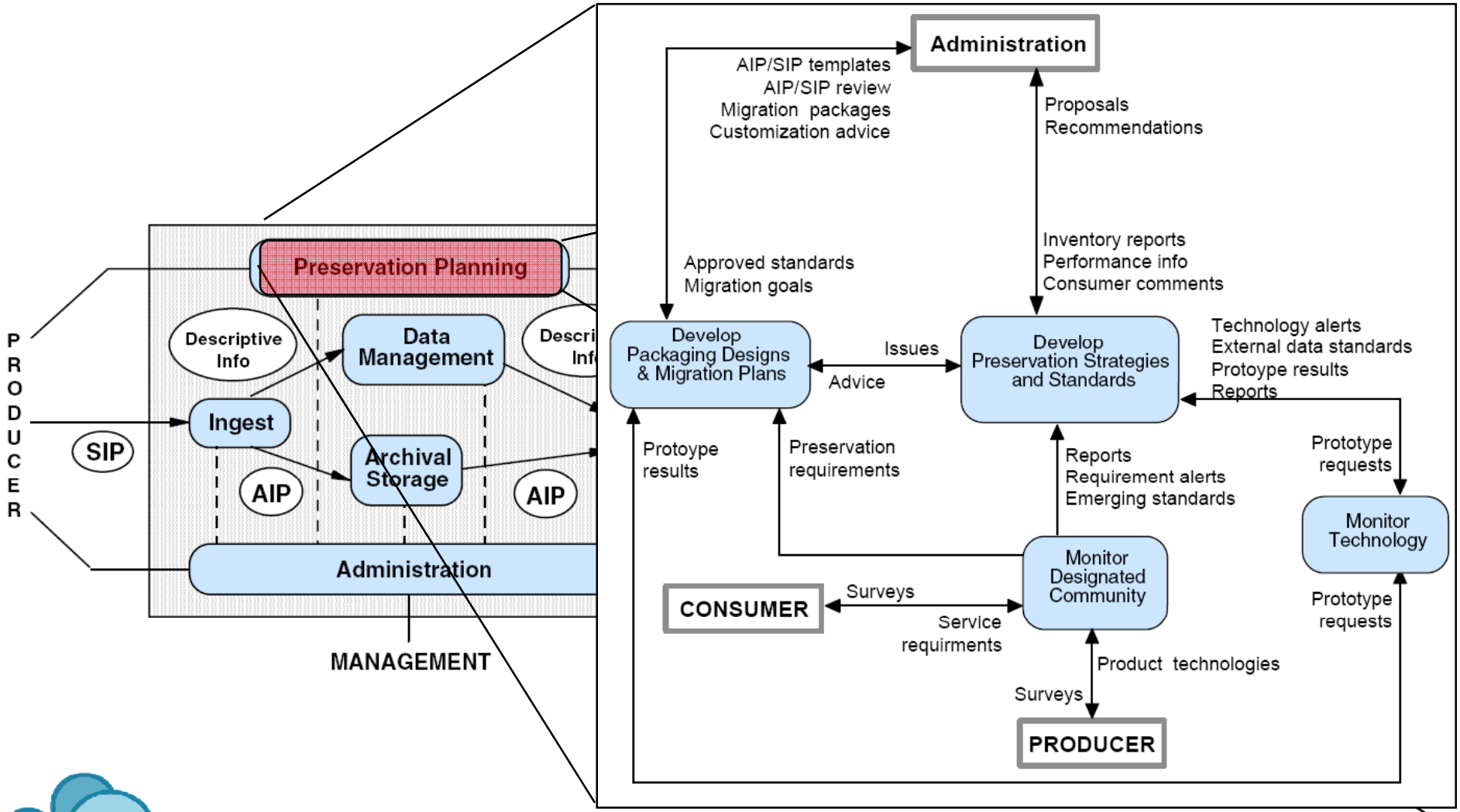
- Variety of solutions and tools exist
- Each strategy has unique strengths and weaknesses
- Requirements vary across settings
- Decision on which solution to adopt is complex
- Documentation and accountability is essential

- Preservation planning assists in decision making
- Evaluating preservation strategies on representative samples according to specific requirements and criteria

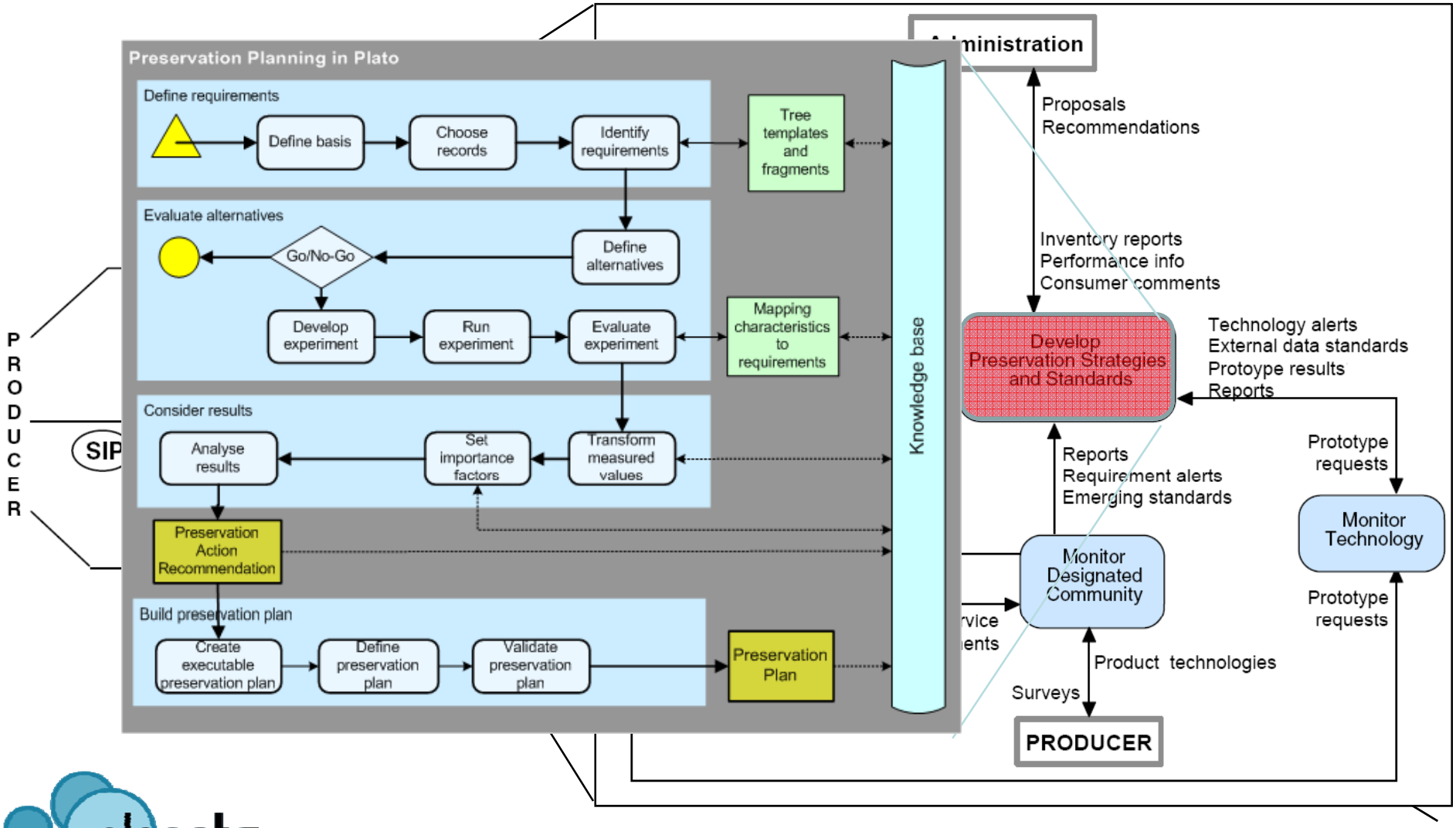
Preservation Planning



Preservation Planning

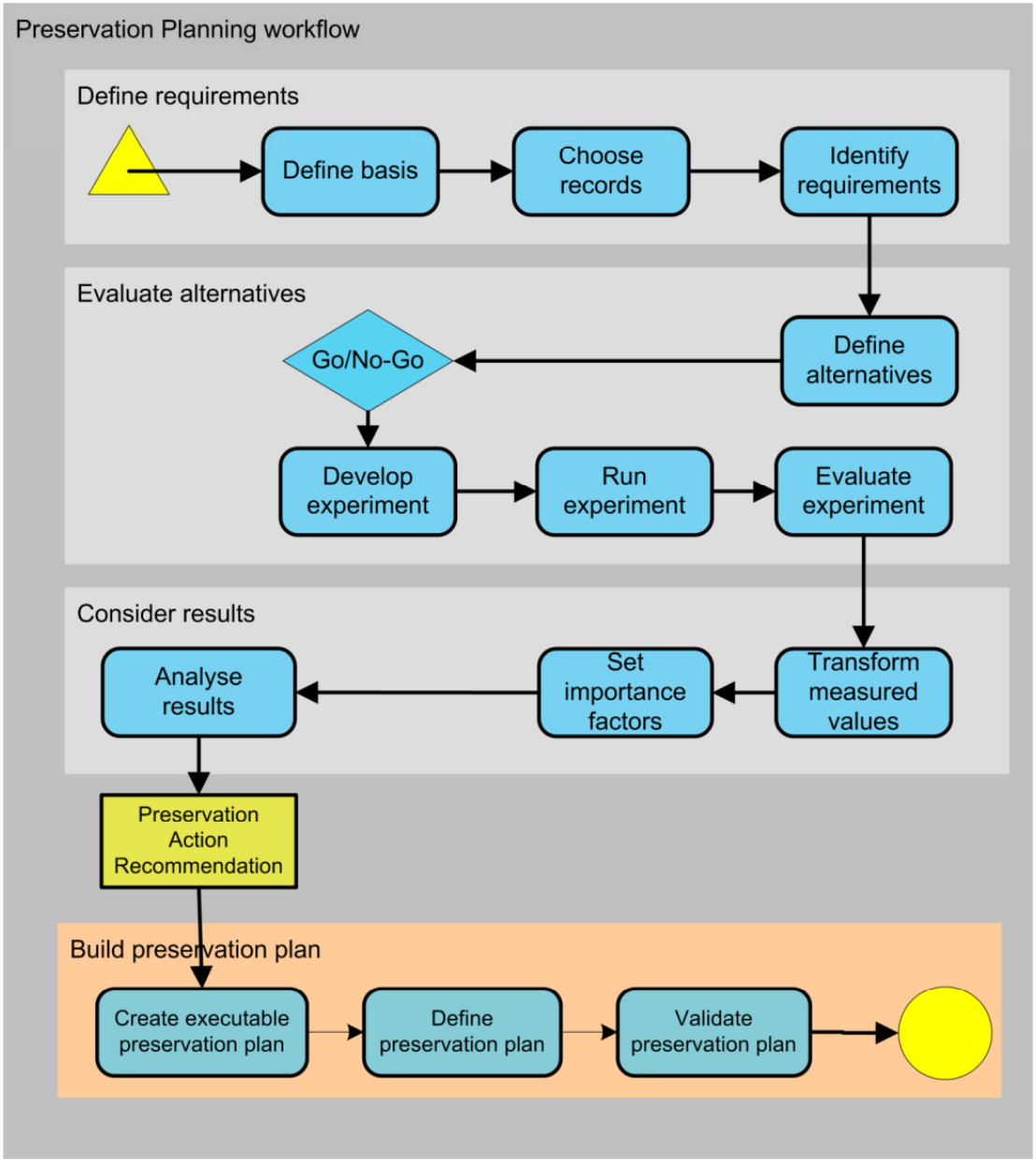


Preservation Planning



Planets Preservation Planning Workflow

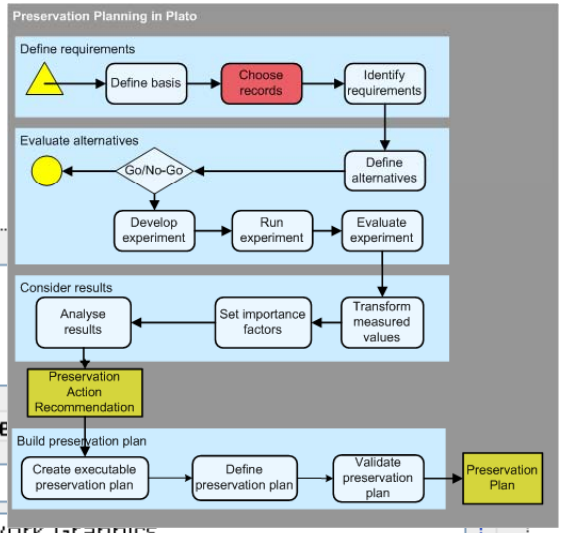
- Define requirements
  - Basis
  - Sample objects
  - Requirements
- Evaluate potential actions
- Analyse results
- Build a preservation plan



Preservation Planning in Plato

- Web based planning tool implementing the Planets preservation planning workflow
- Publicly available
- Ongoing development
- Integration of registries and services for
  - File format identification
  - Preservation action
  - Characterisation and comparison
- Frontend to a distributed architecture of preservation services

Format identification



[-] Sample Records

Description of sample records: some test samples

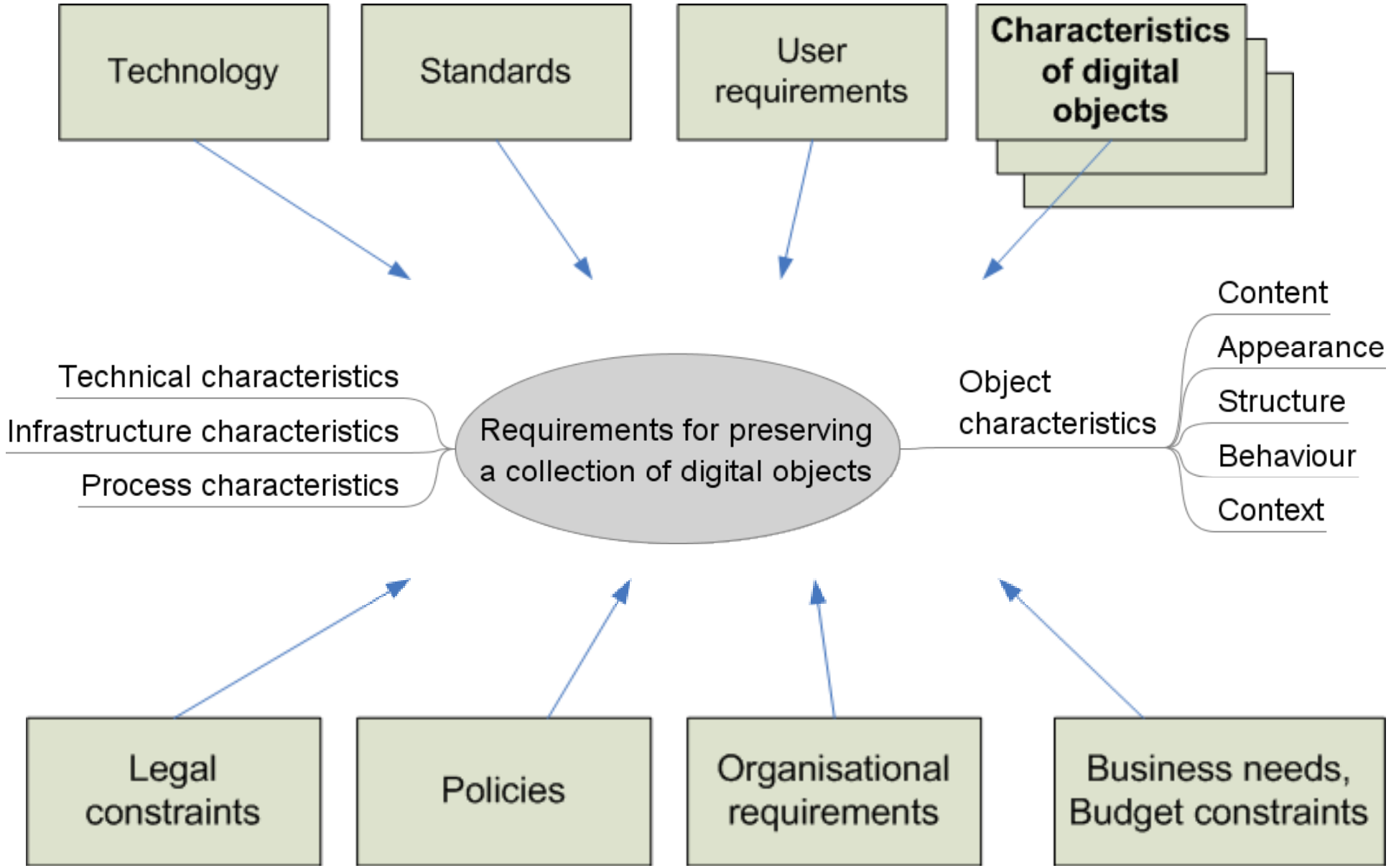
Sample Record	Object
Full name: sample 1 * ?	PUID: fmt/11
Short name: eins.png * ?	Name: Portable Network Graphics
Has data: download	Version: 1.0 ?
Original technical environment: ?	Mime-type: image/png ?
Description: ?	Identify format DROID

Remove record

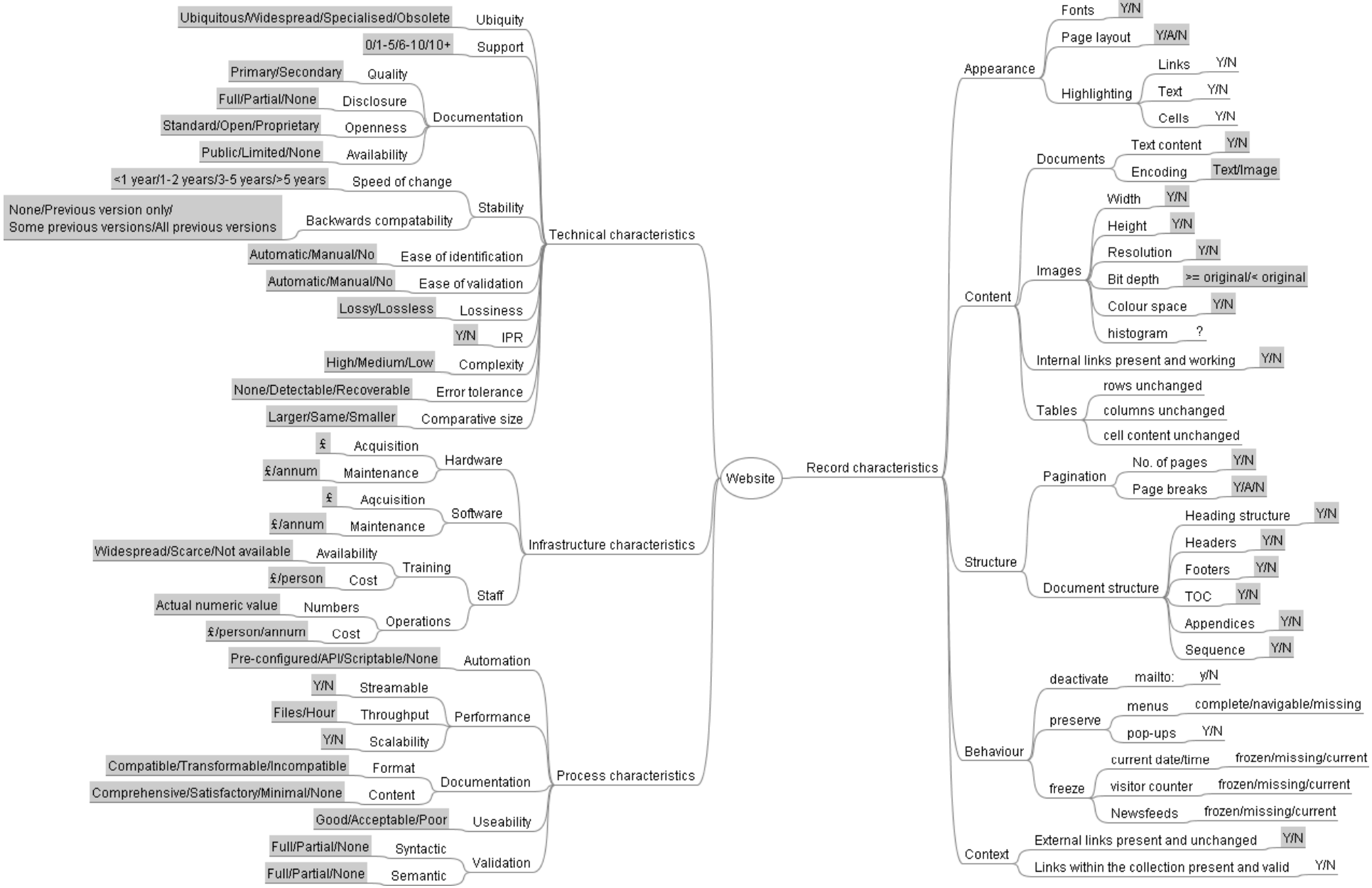
Sample Record	Object
Full name: sample number two * ?	PUID: fmt/11 ?
Short name: zwo.png * ?	Name: Portable Network Graphics ?
Has data: download	Version: 1.0 ?
Original technical environment: ?	Mime-type: image/png ?
Description: ?	Identify format DROID

Remove record

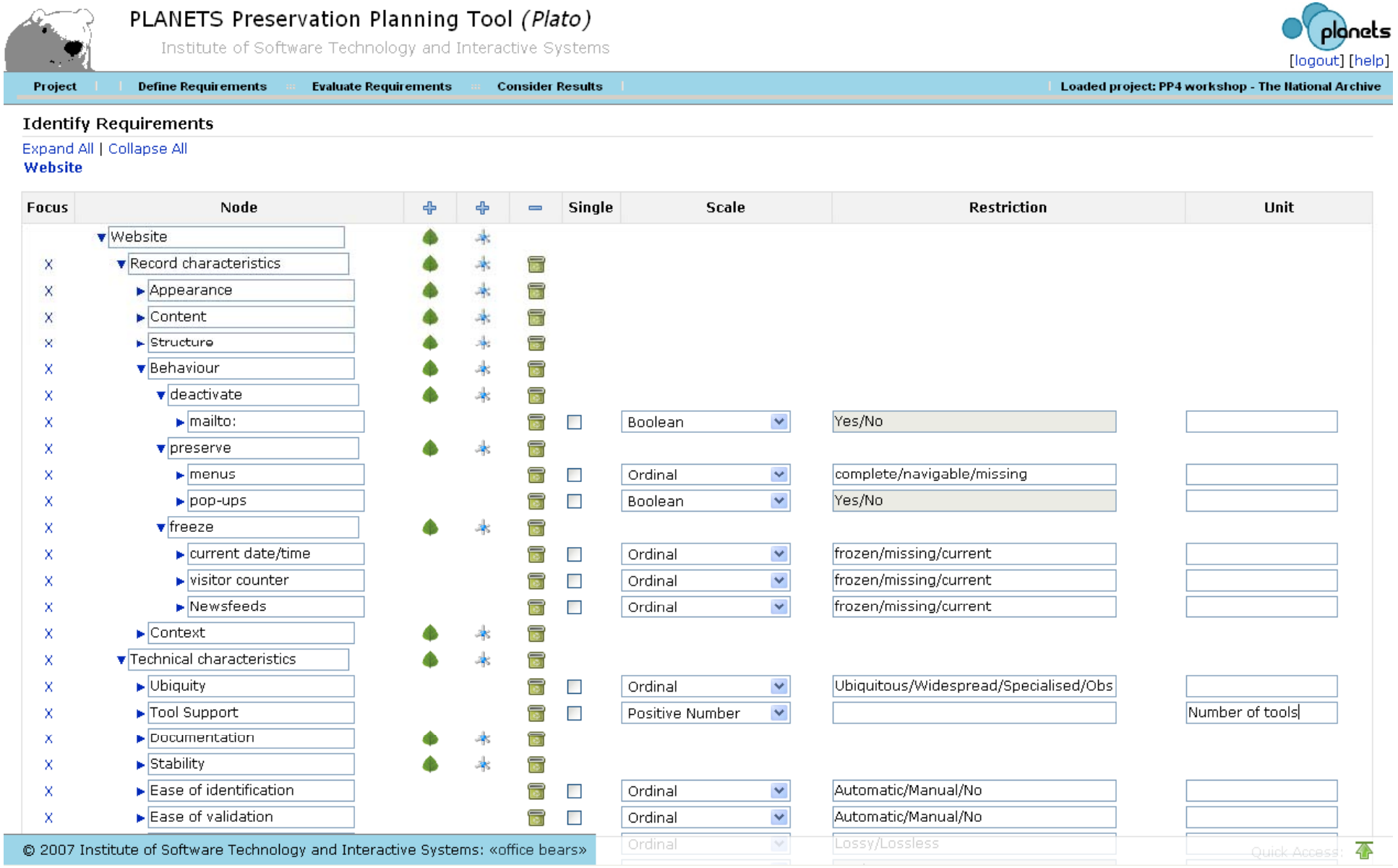
Requirements and Influence Factors

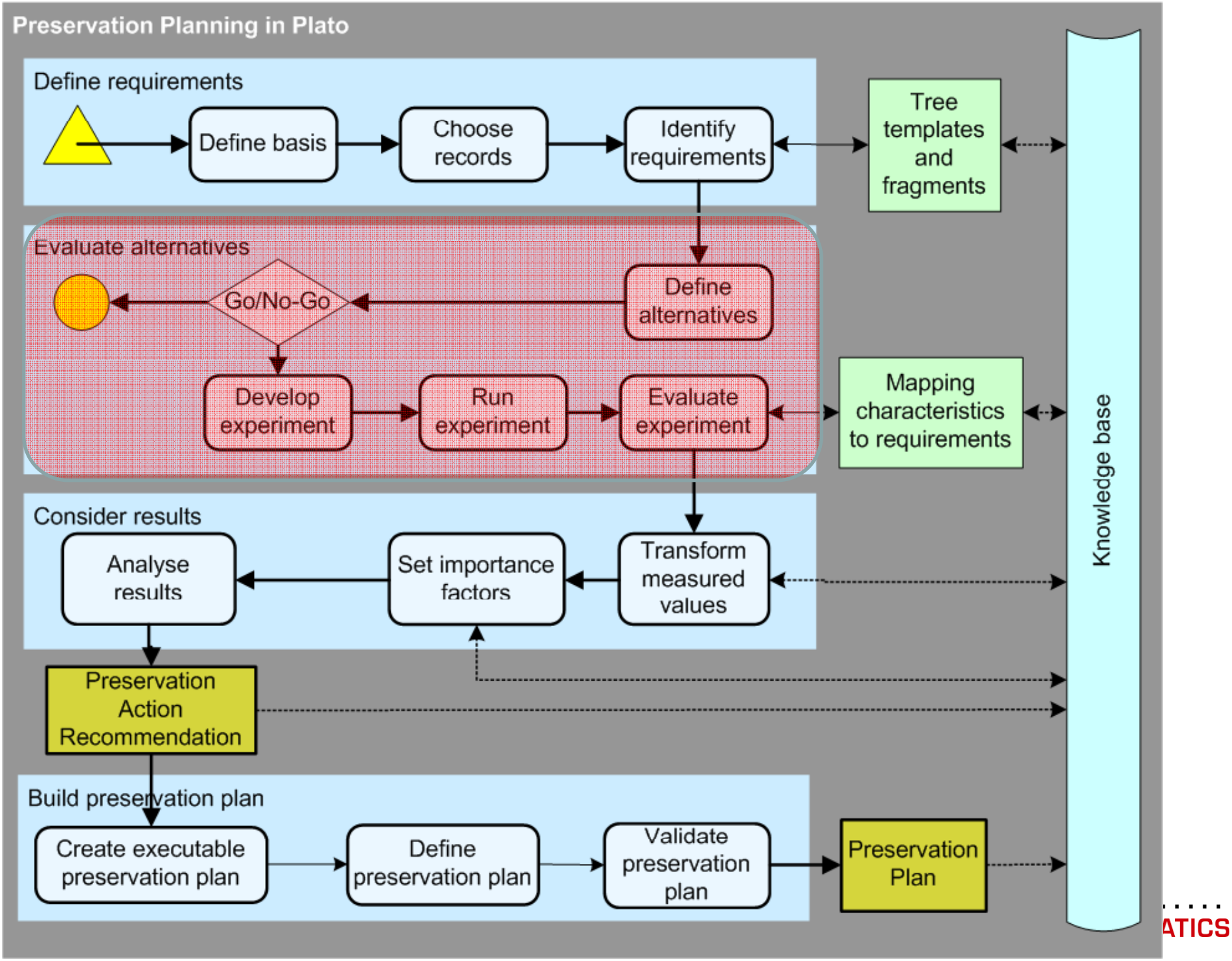


An Objective Tree

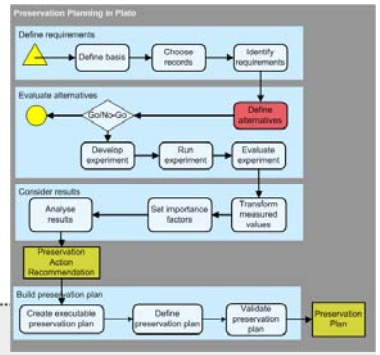


The Objective Tree in Plato




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
Knowledge base
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

Service discovery and invocation



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

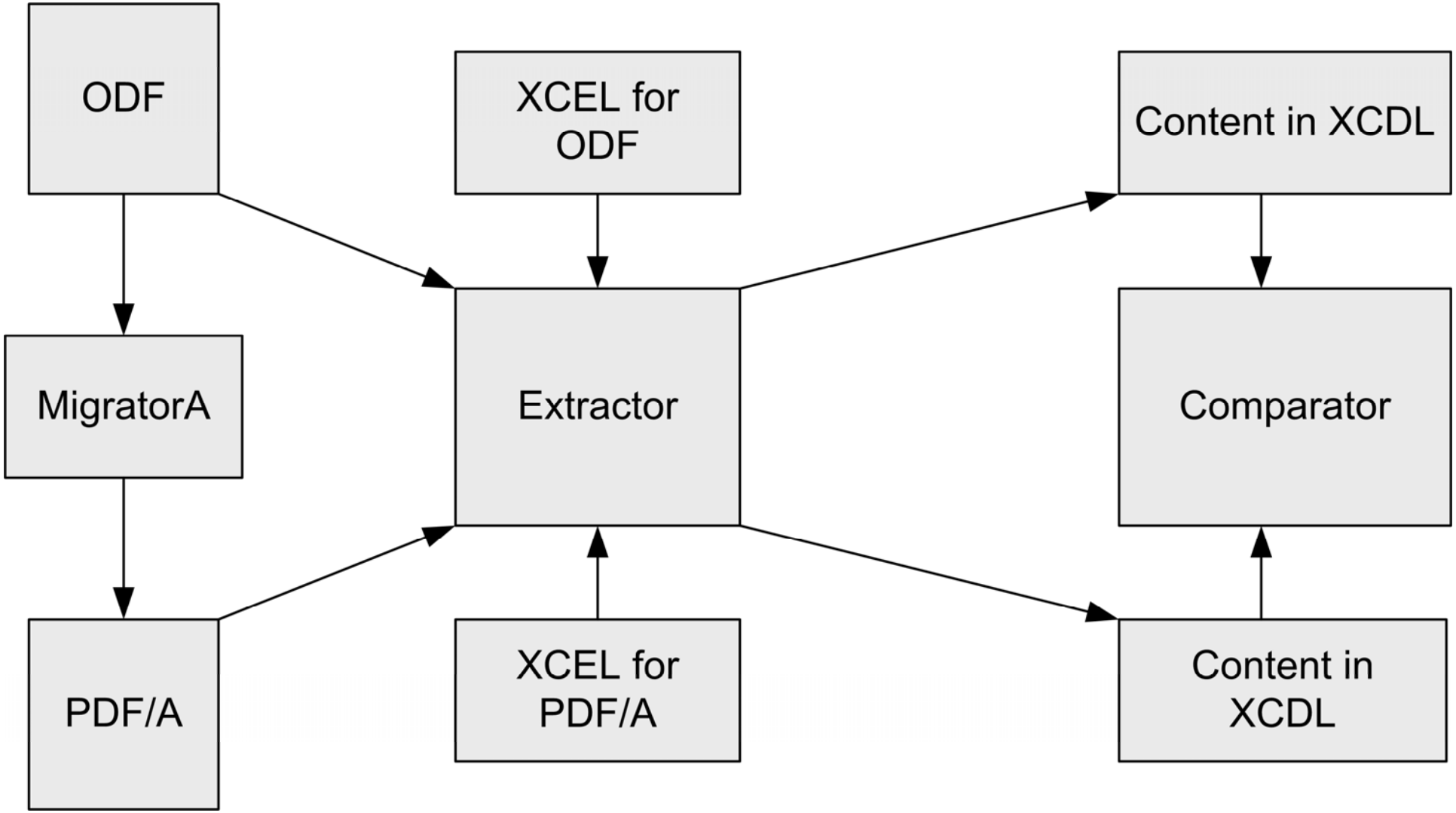
Create alternatives for selected services

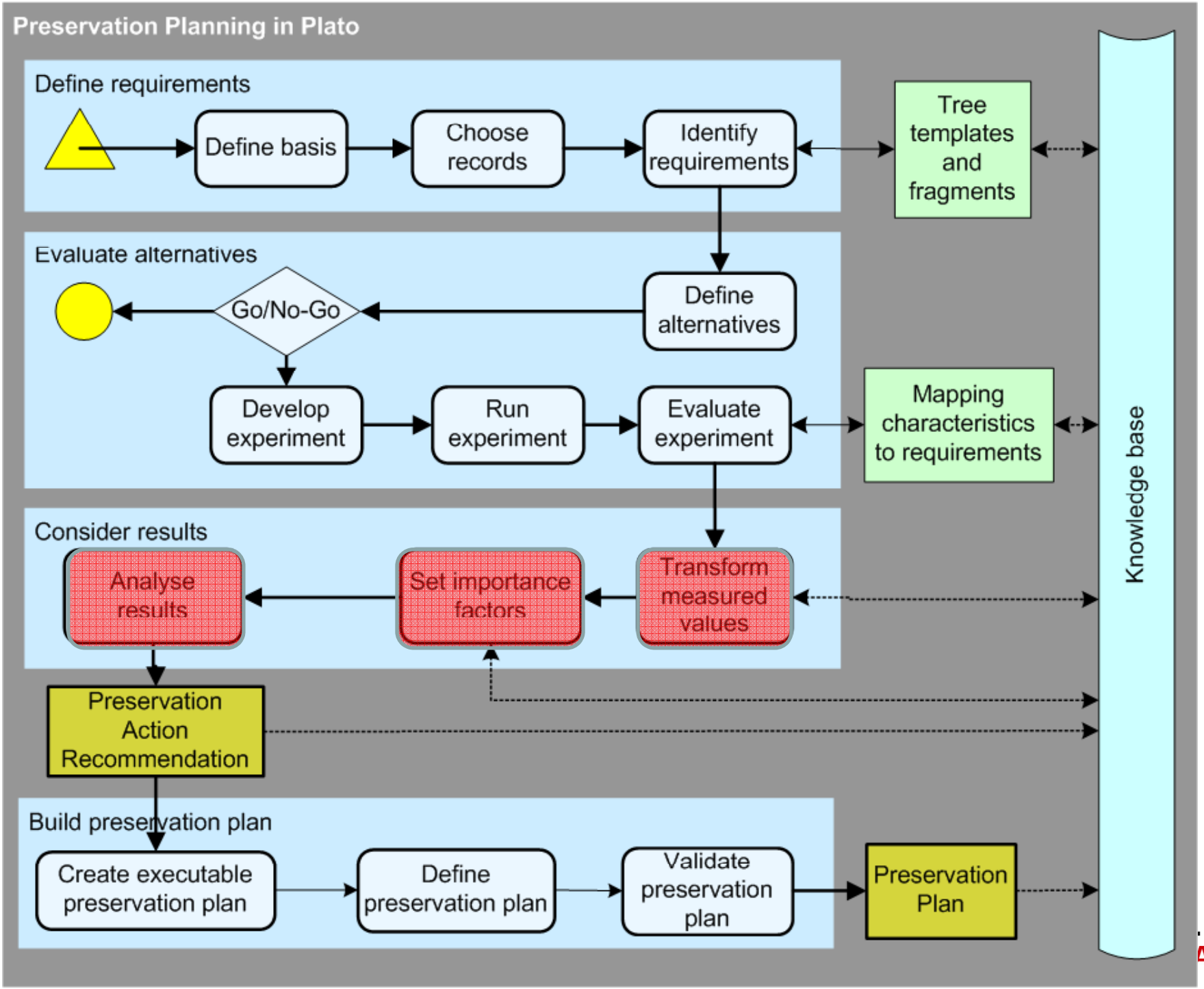
Evaluation of actions

- Apply actions on sample content
- Evaluate outcomes

- Significant properties of objects
  - Content
  - Appearance
  - Structure
  - Behaviour
  - Context
- Image width in Pixel, colour depth in bit...
- Technical characteristics extracted from objects

Comparing migrated documents




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

Aggregation method: Sum

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

Show

Alle einblenden | Alle ausblenden

Minimalist root node

Fokus	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	▼Amount of Pixel	PDF/A ToolA: 3.50 PDF/A ToolB: 4.00
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

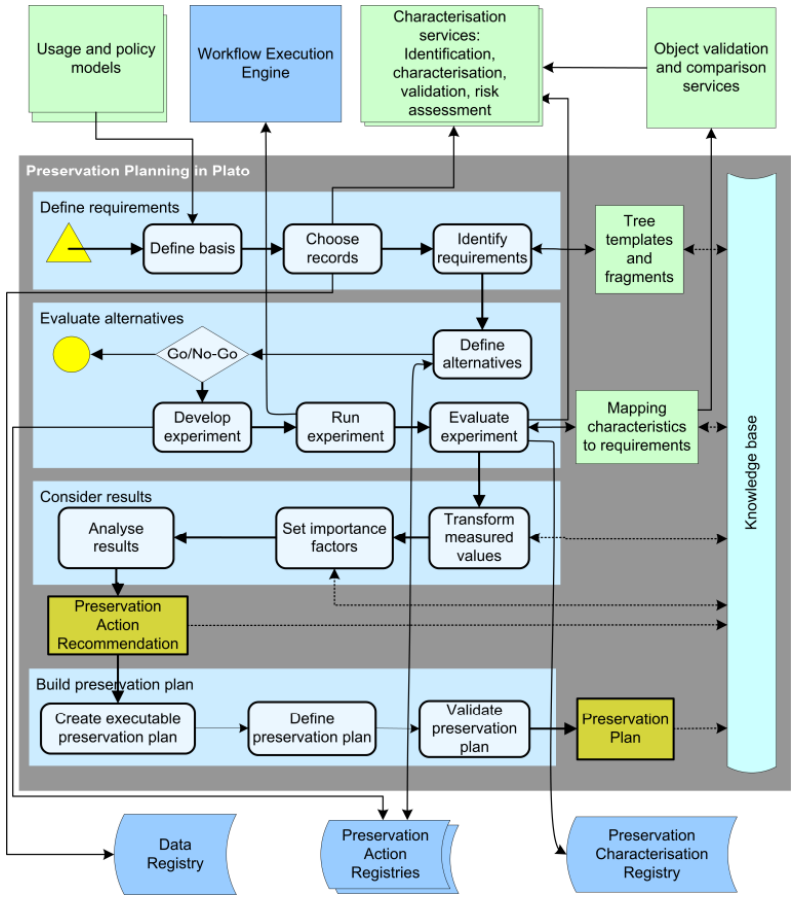
Recommendation

Recommendation:

Reasoning: ?

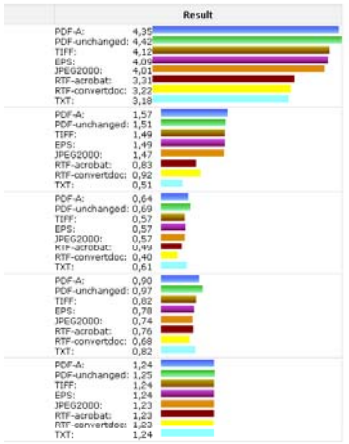
Summary

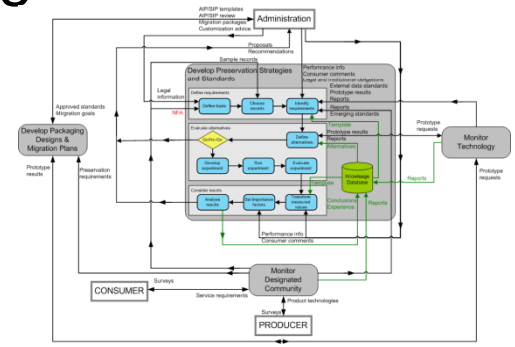
- Preservation planning in Plato
- Workflow
- Characterisation
  - Format identification
  - Collection profiling
  - Risk assessment
  - XCL and comparison
- Discovering applicable actions
- Building a preservation plan



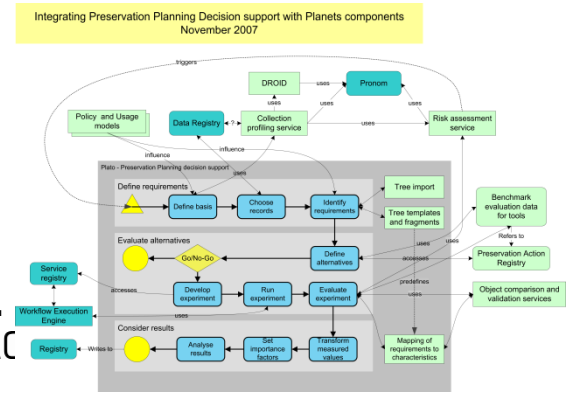
Future work

- Pluggable infrastructure for the automated evaluation of preservation actions
- Integrated knowledge base and recommender systems
- Case studies on electronic documents, image archives, electronic art, computer games...





- Plato 1.3 in July
- Plato 2.0 scheduled for October 2008
  - Service integration
  - Preservation plan
- Plato 3.0 early 2010



Questions?

www.ifs.tuwien.ac.at/~becker

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

www.planets-project.eu