Logical and bit-stream preservation using Plato and EPrints

Hannes Kulovits
Andreas Rauber

Department of Software Technology and Interactive Systems

Vienna University of Technology
kulovits@ifs.tuwien.ac.at
rauber@ifs.tuwien.ac.at

David Tarrant
Adam Field

School of Electronics and Computer Science

University of Southampton, UK
dct05r@ecs.soton.ac.uk
af05v@ecs.soton.ac.uk

Vienna University of Technology

- **Vienna University of Technology**
  http://www.tuwien.ac.at
  - Faculty of Computer Science
    http://www.cs.tuwien.ac.at
    - Department of Software Technology and Interactive Systems (ISIS)
      http://www.isis.tuwien.ac.at
- People in DP
  - **Andreas Rauber**
  - Christoph Becker
  - Mark Guttenbrunner
  - Rudolf Mayer
  - Michael Kraxner
  - **Hannes Kulovits**
  - Stephan Strodl
  - Michael Greifeneder
  - Petar Petrov

TU WIEN planets UNIVERSITY OF Southampton eprints

DP Activities in Vienna

- Web Archiving (AOLA)
  in cooperation with the Austrian National Library
- DELOS DPC (EU FP6 NoE)
- DPE: Digital Preservation Europe (EU FP6 CA)
- PLANETS (EU FP6 IP)
- eGovernment & Digital Preservation
  series of projects with Federal Chancellery
- National Working Group on Digital Preservation
  of the Austrian Computer Society, in cooperation with ONB
- Digital Memory Engineering: National research studio

TU WIEN planets UNIVERSITY OF Southampton eprints

University of Southampton, UK

- **University of Southampton**
  http://www.soton.ac.uk
  - School of Electronics & Computer Science
    http://www.ecs.soton.ac.uk
- **EPrints**
  http://www.epints.org
- People in Preservation
  - Steve Hitchcock
  - **David Tarrant**
  - Chris Gutteridge
  - Tim Brody
  - Patrick McSweeny
- EPrints Services
  - **Adam Field**
  - Tim Miles-Board

TU WIEN planets UNIVERSITY OF Southampton eprints

DP Activities in Southampton

- EPrints Preservation
  - KeepIt!
  - Preserv2
  - Preserv
- P2N – Preservation Network
  - Collabotarion with Oxford Univeristy
- P2-Registry
  - Linked Data for Digital Preservation
- Web Archiving
  - ECS project to archive old project websites and Wikis

TU WIEN planets UNIVERSITY OF Southampton eprints

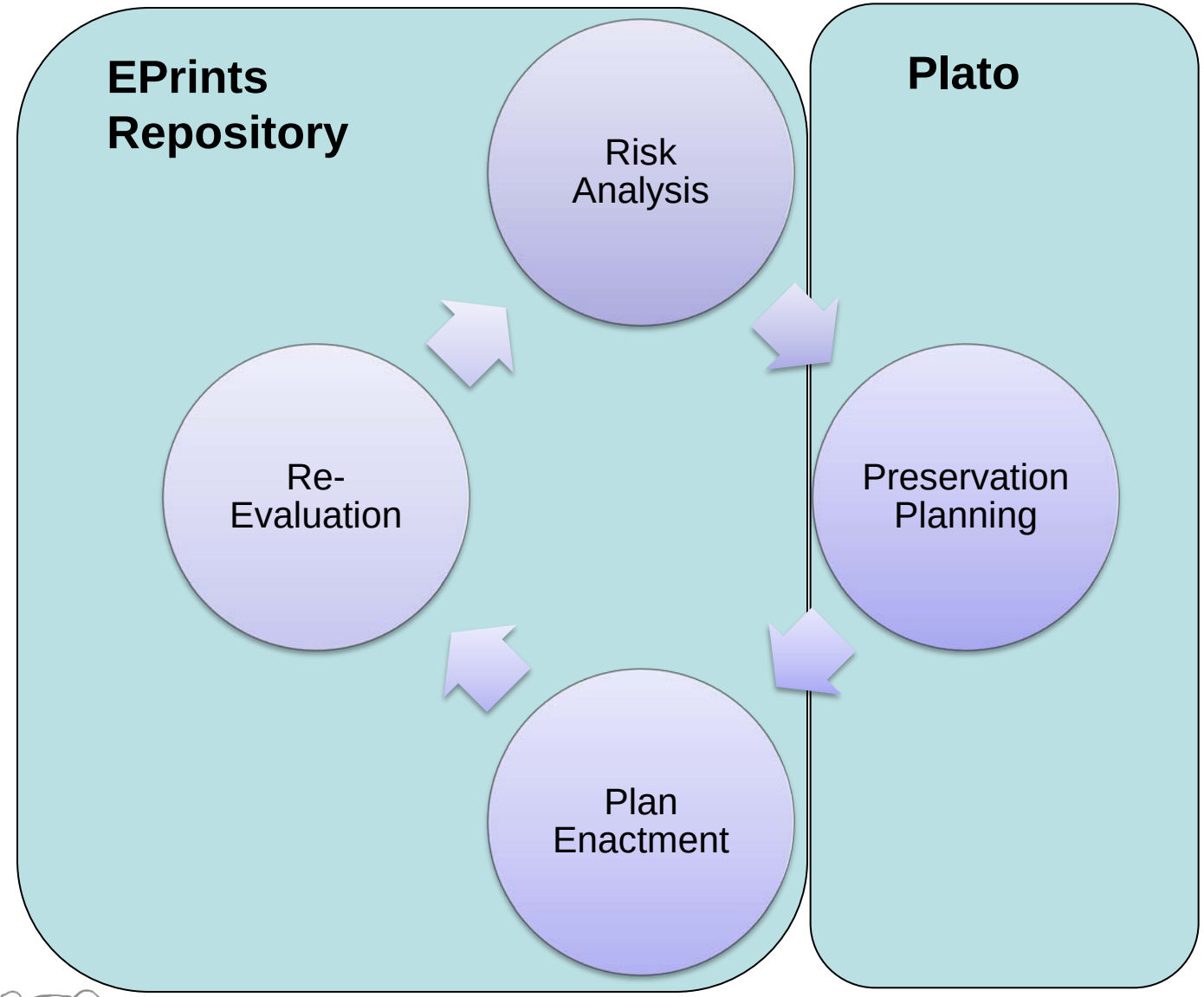
Introductions

What will you know after this tutorial?

You will:

- See the (first?) system integrating **bit stream preservation** and **logical preservation** supported by a **fully documented planning** process
- Perform **risk analysis as trigger** for preservation actions
- Understand why we need to **plan preservation activities**
- Know a **workflow to evaluate preservation strategies**
- Be familiar with Plato and EPrints
- Be able to **develop a specific preservation plan** that is optimized for
  - the objects in your institution
  - the users of your institution
  - the institutional requirements
- Be able to **execute it** in a repository (EPrints)

Integrated Preservation Cycle



Schedule

09:00 – 09:45	Introduction
09:45 – 11:00	Exercise 1 (EPrints)
11:00 – 11:15	Coffee/Tea
11:15 – 13:00	Requirements
13:00 – 14:00	Lunch
14:00 – 15:30	Evaluation/ Transformation
15:30 – 16:00	Coffee/Tea
16:00 – 17:15	EPrints
17:15 – 18:00	Discussion
(18:15 - ???	Ice breaking & Wine tasting)

Schedule

(1) Introduction

(2) Preservation in EPrints

(3) Preservation Planning with Plato

(4) Bringing it all together and Closing

Overview

Part 1: Introduction

- Quick introduction to physical preservation with EPrints
- Quick introduction to logical preservation with Plato
- Bringing it together: bit-stream and logical preservation

What is EPrints For?

- EPrints offers a safe, open and useful place to store, share and manage material in the pursuit of research and educational agendas.

administrative reporting, collaboration, data sharing, digital profile enhancement , e-learning, e-publishing, e-research, marketing, open access, preservation, publicity, research assessment, research management, scholarly collections

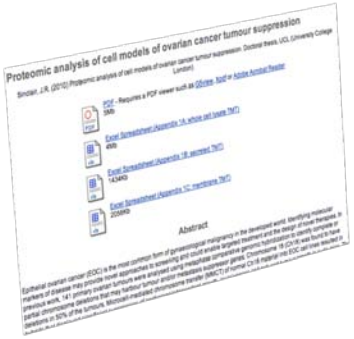
TU WIEN planets UNIVERSITY OF Southampton eprints

An EPrints repository is

- A valuable part of the researcher's information environment
  - directly integrating with the research desktop
  - offering sustainable storage and open access

- A competent and mature component of the institution's information environment
  - providing management and curation support for core business research data
  - leveraging information about research outputs to inform management strategy

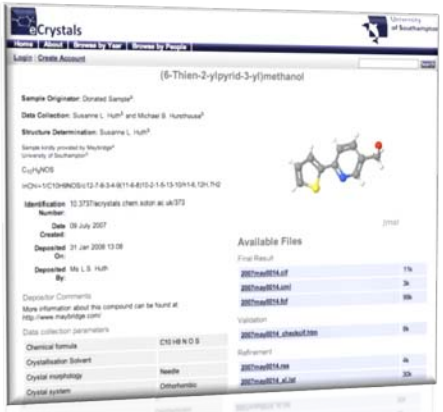
KeepIT Exemplars



Open Access to Research Outputs

Open Arts

Open Educational Resources



Open Scientific Data

TU WIEN

planets

UNIVERSITY OF Southampton

eprints

EPrints Repositories

- eprints.lse.ac.uk (institutional)
- eprints.ecs.soton.ac.uk (departmental)
- pubs.or08.ecs.soton.ac.uk (conference)
- archive.serpentproject.com (project)
- nora.nerc.ac.uk (funders)
- ecrystals.chem.soton.ac.uk (data)
- www.linnean-online.org (collection)
- ualresearchonline.arts.ac.uk (art)
- **demoprints.eprints.org** (demo)

Overview

Part 1: Introduction

- Quick introduction to physical preservation with EPrints
- Quick introduction to logical preservation with Plato
- Bringing it together: bit-stream and logical preservation

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

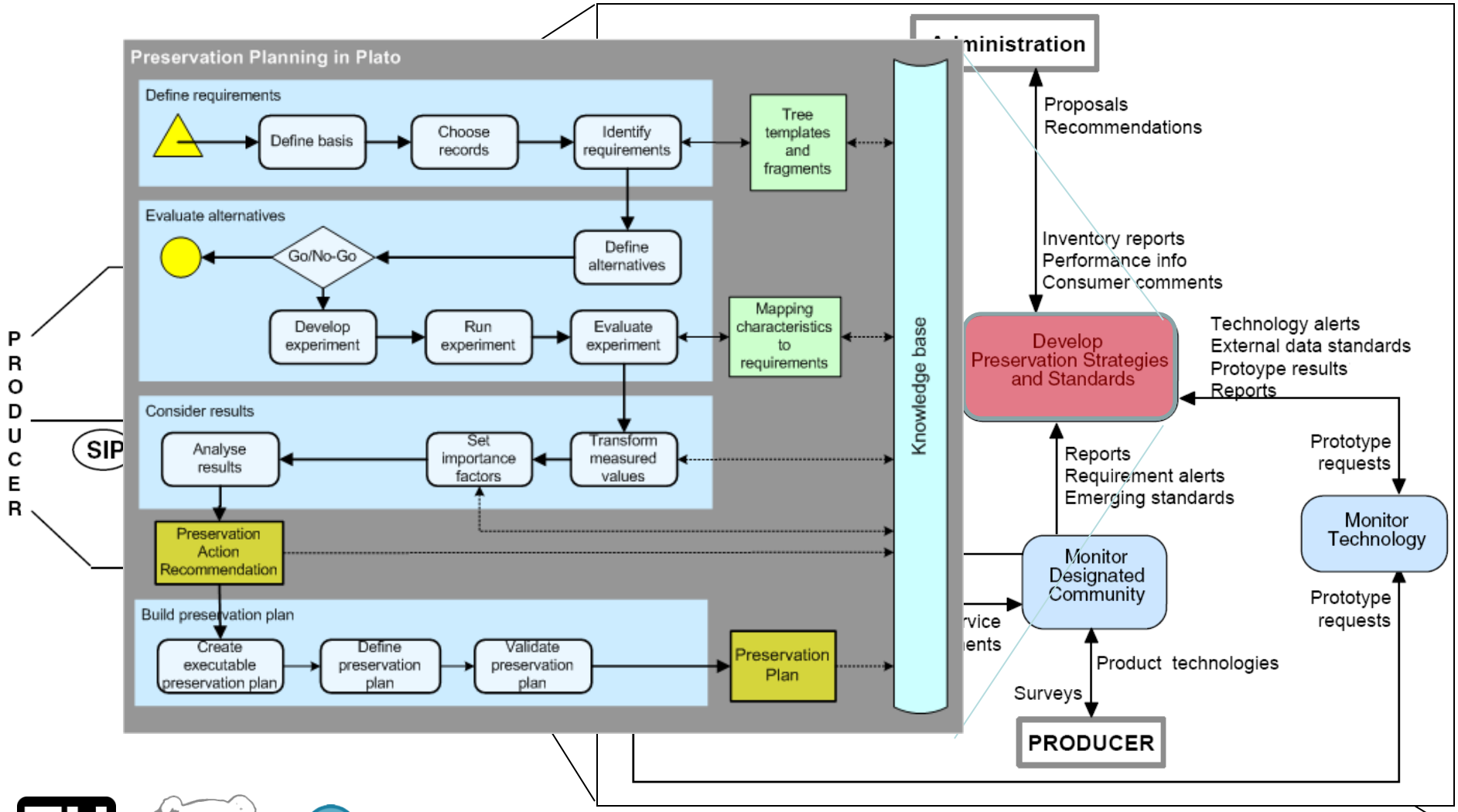
Digital Preservation

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

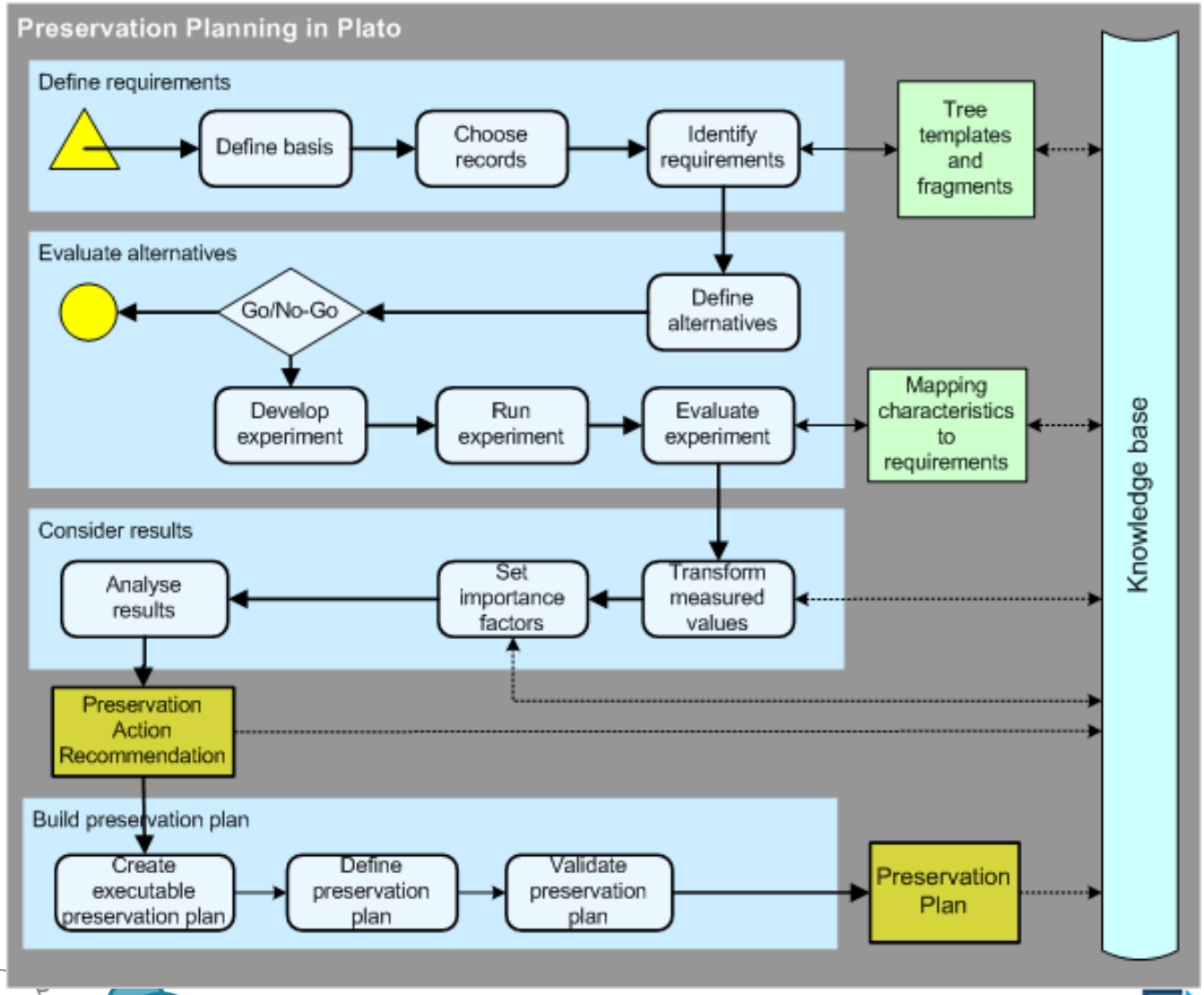
Preservation Plan Template

Preservation Planning

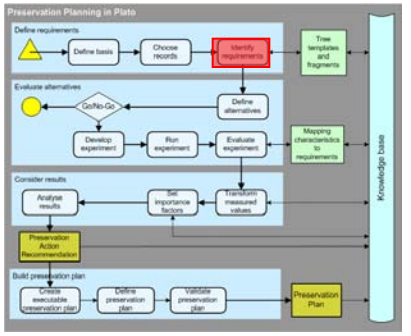


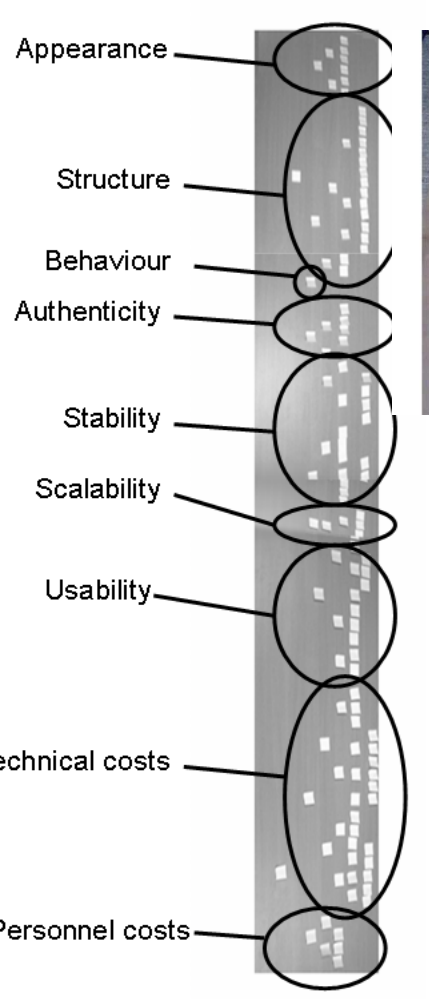
TU WIEN planets UNIVERSITY OF Southampton eprints

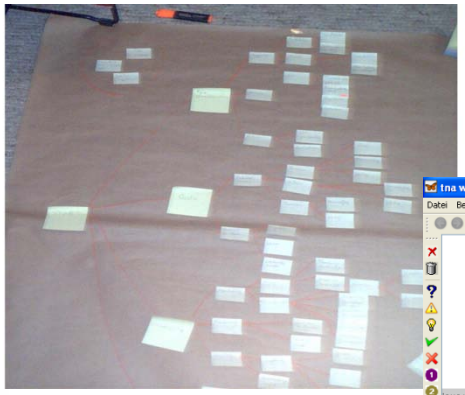
Preservation Planning Workflow



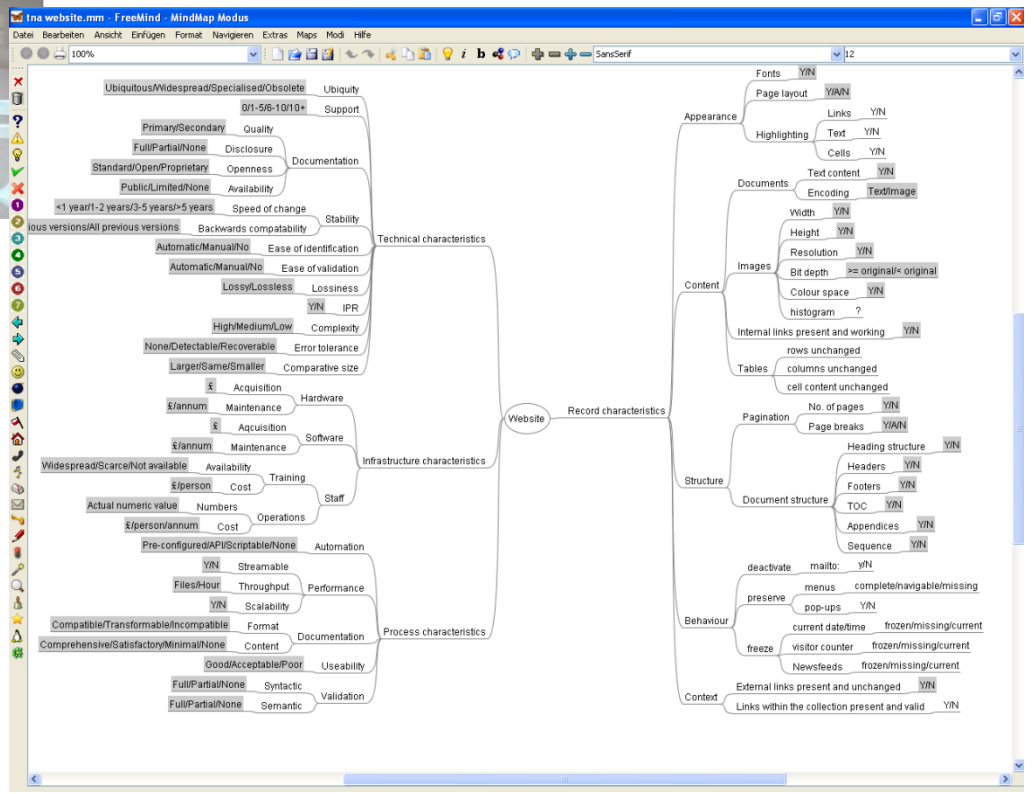
Identify requirements







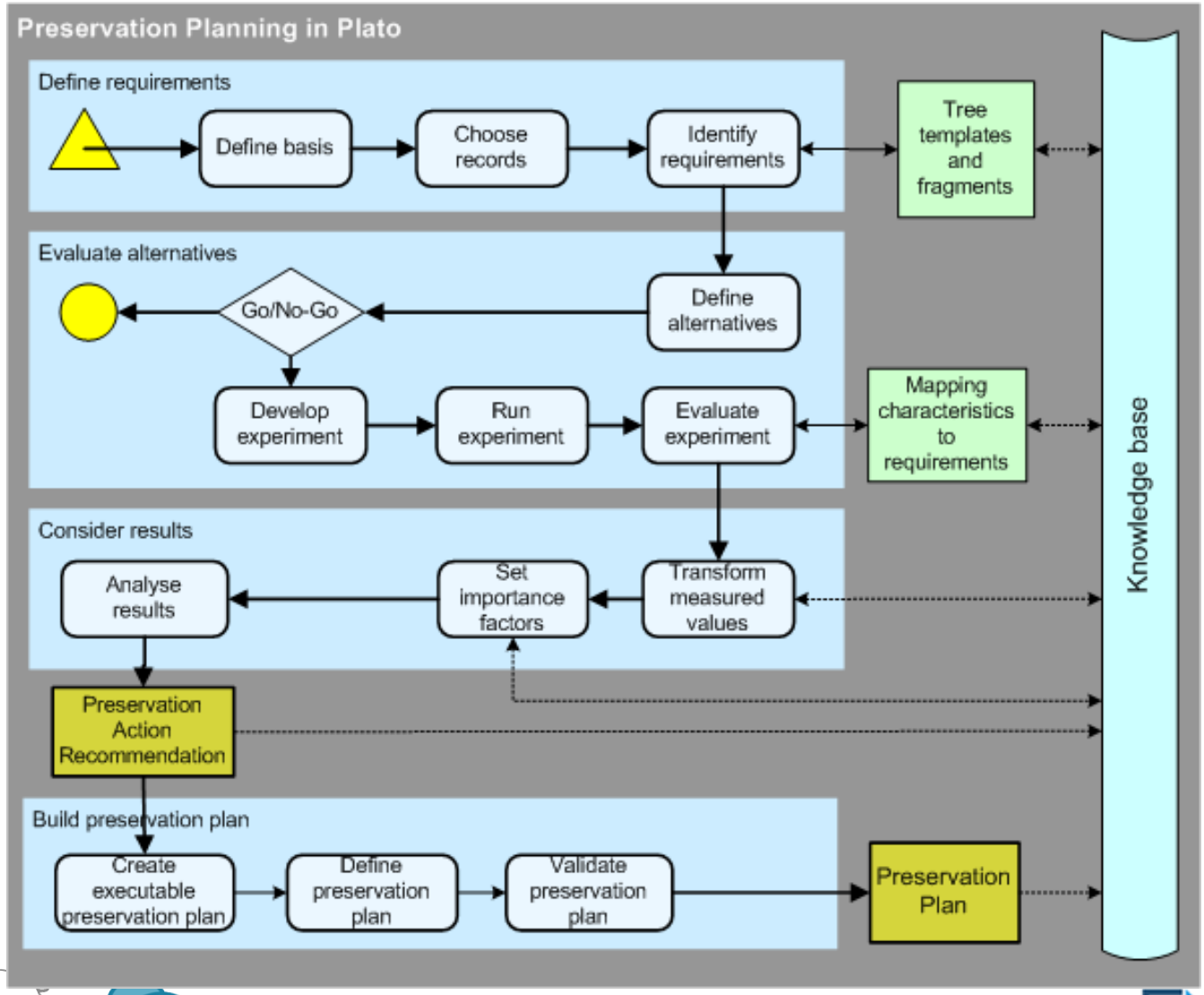
Analog…



… or born digital

TU WIEN planets UNIVERSITY OF Southampton eprints

Preservation Planning Workflow

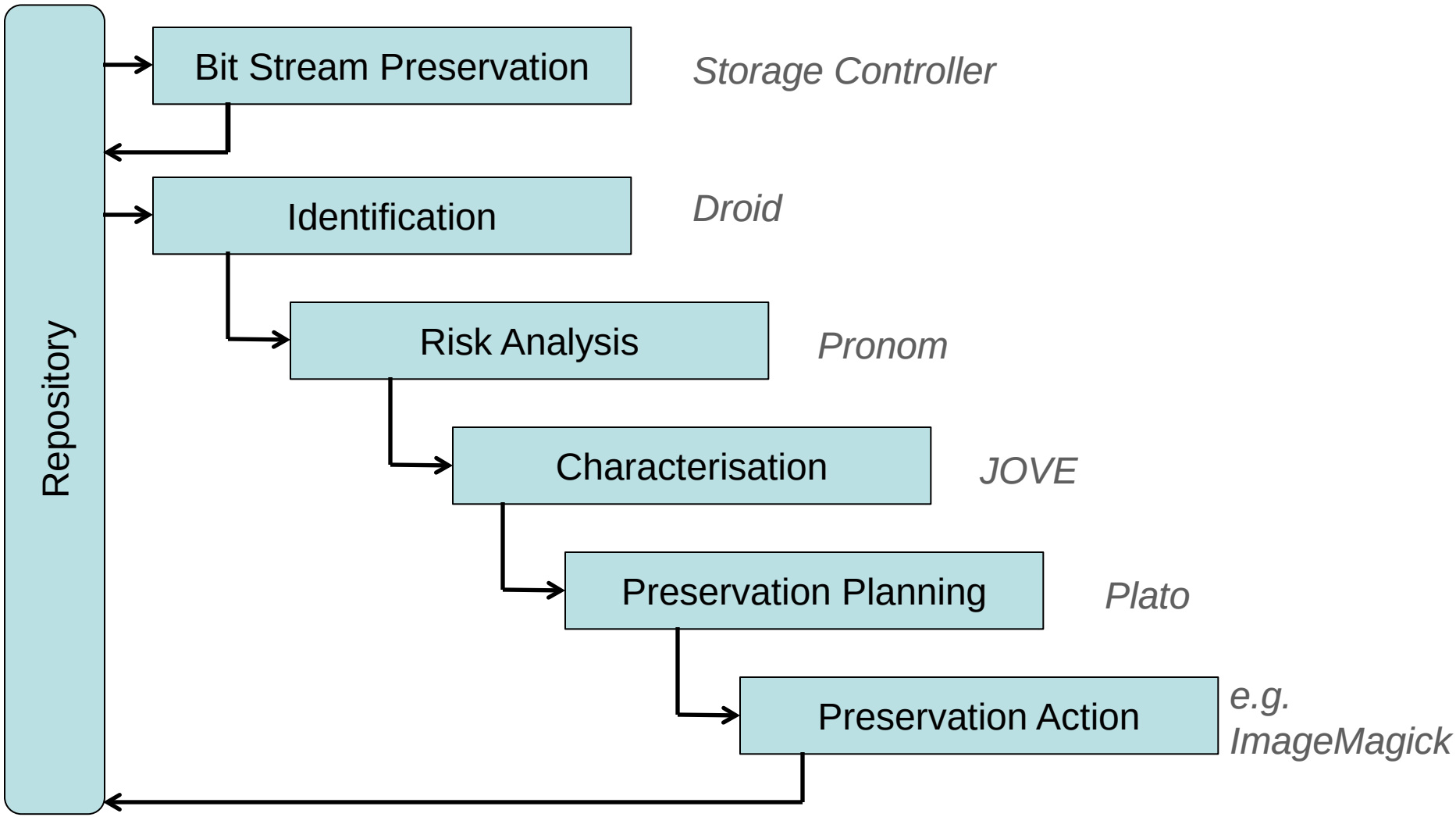


Overview

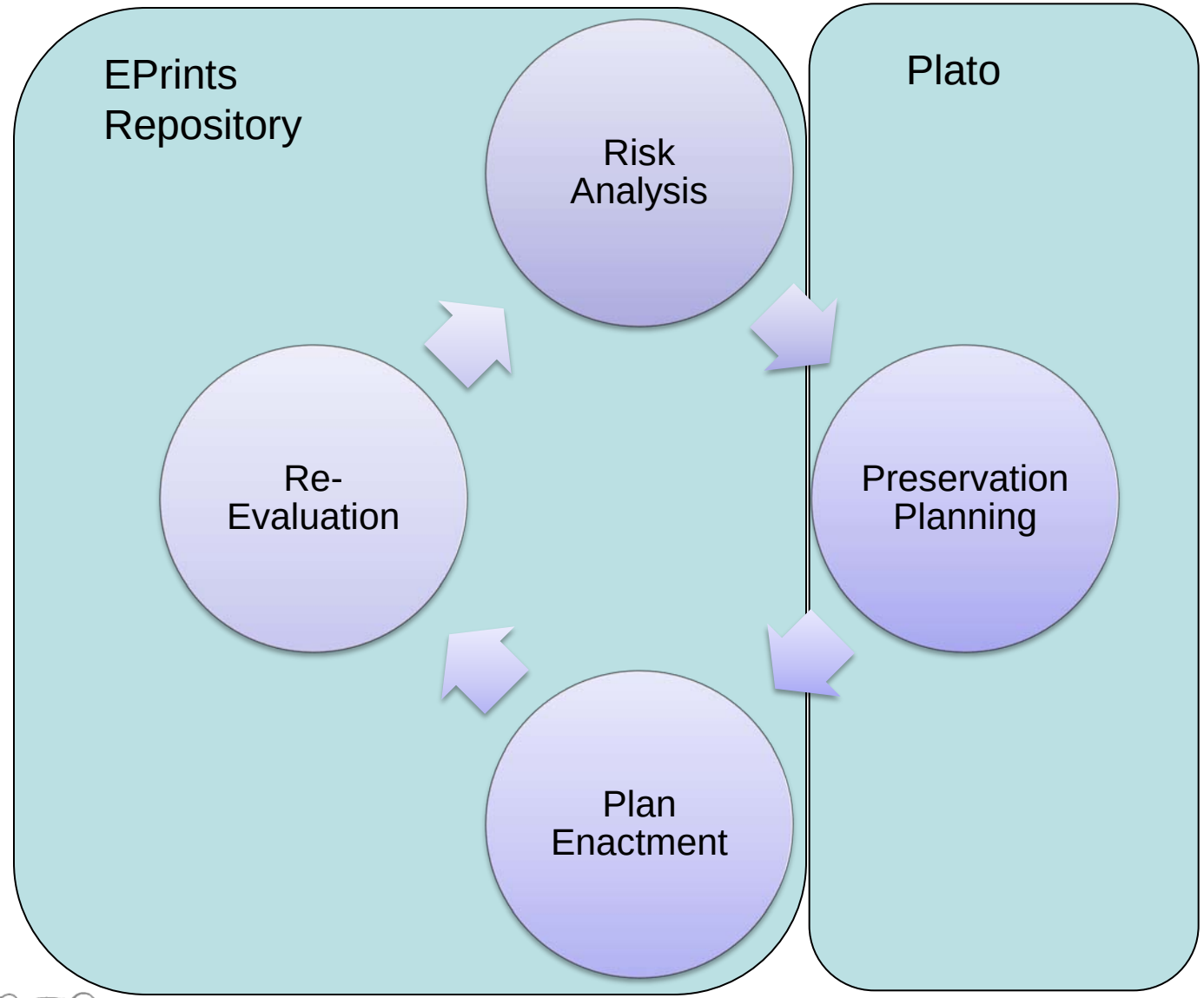
Part 1: Introduction

- Quick introduction to physical preservation with EPrints
- Quick introduction to logical preservation with Plato
- Bringing it together: bit-stream and logical preservation

Bringing it all together



Bringing it all together (3/2)



Conclusions

- Integrating bit-stream and logical preservation
- Thorough planning process
- Actionable preservation plan
- Consistent with OAIS model
- Follows recommendations of TRAC and nestor
- Generic workflow that can easily be integrated in different institutional settings
- **EPrints**:
  - Open-source repository system

    http://www.eprints.org
- **Plato**:
  - Tool support for preservation planning

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

Schedule

(1) Introduction

(2) Preservation in EPrints

(3) Preservation Planning with Plato

(4) Bringing it all together and Closing

The Preservation Process

Preservation - Check

- Resilient Storage
- Bit checking & checksum calculation

Preservation - Analyse

- What is the type of file, is the file valid?
- Is the file at risk of not having an editor/reader?
- Is there a better format available? Lossless or Lossy?

Preservation - Planning

- What is the best preservation action given requirements and constraints
- Preservation Planning (Plato)

Preservation - Action

- File migration to avert risks found by analysis.
- Movement of file to new storage.

The Storage Ecosystem

Local

- No local bandwidth costs
- Hard to expand
- Locally Managed
- High overheads cost
- Requires space and cooling
- Tied closely to the software

Archival

- Specialist
- Expensive to purchase
- Locally Managed
- Space and running costs
- Expandable

Cloud

- Scalable
- Externally controlled
- Known Costings
- Unclear retention policy
- Re-Useable (APIs)
- Global Scale

TU WIEN planets UNIVERSITY OF Southampton eprints

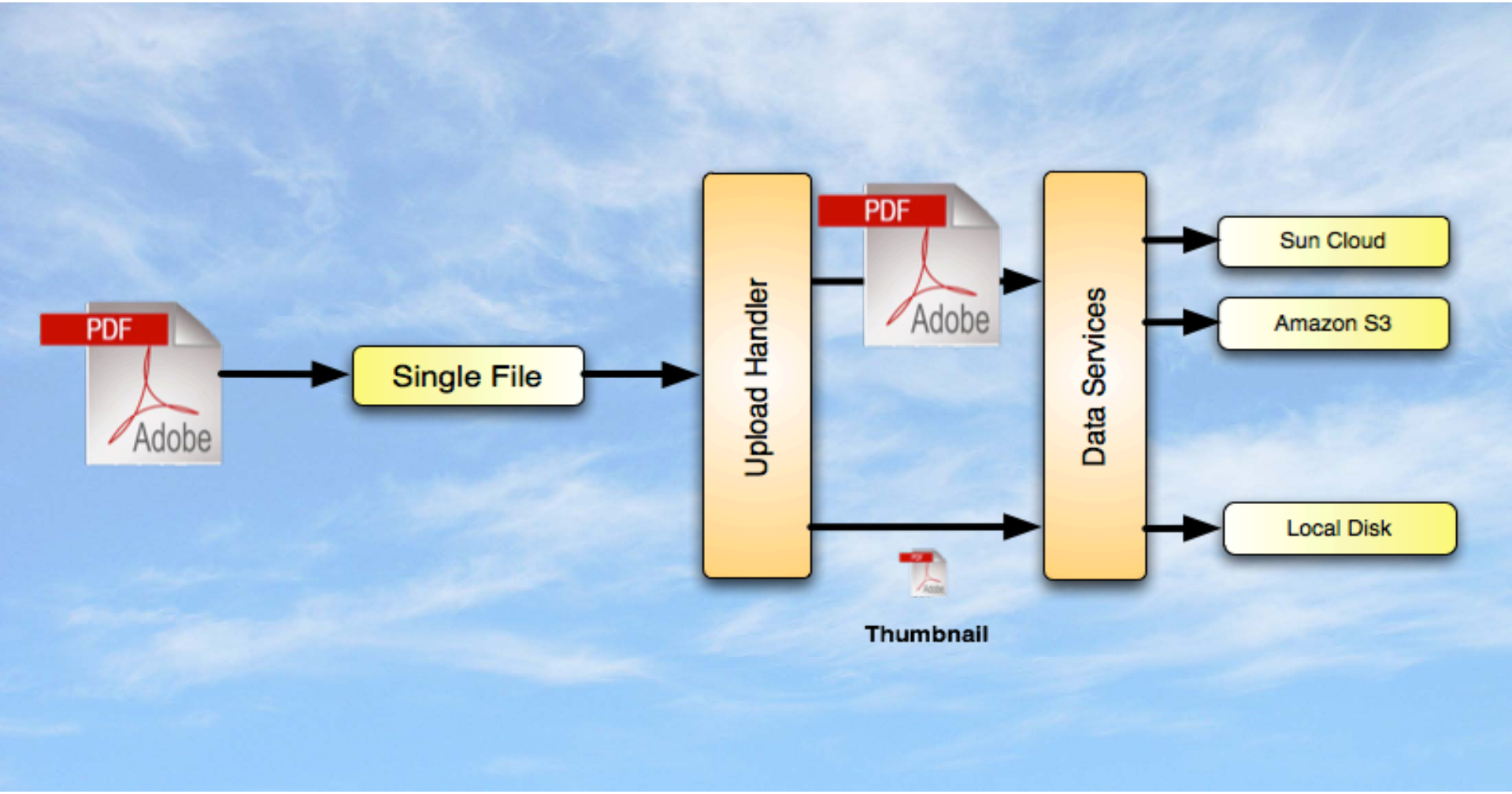
Hybrid Storage

- Use the best features of each storage type
- Performance
  - Scaling-up bandwidth
- Optimisation
  - Large-file handling
  - Multimedia streaming
- Localised Delivery
  - Local delivery from the cloud

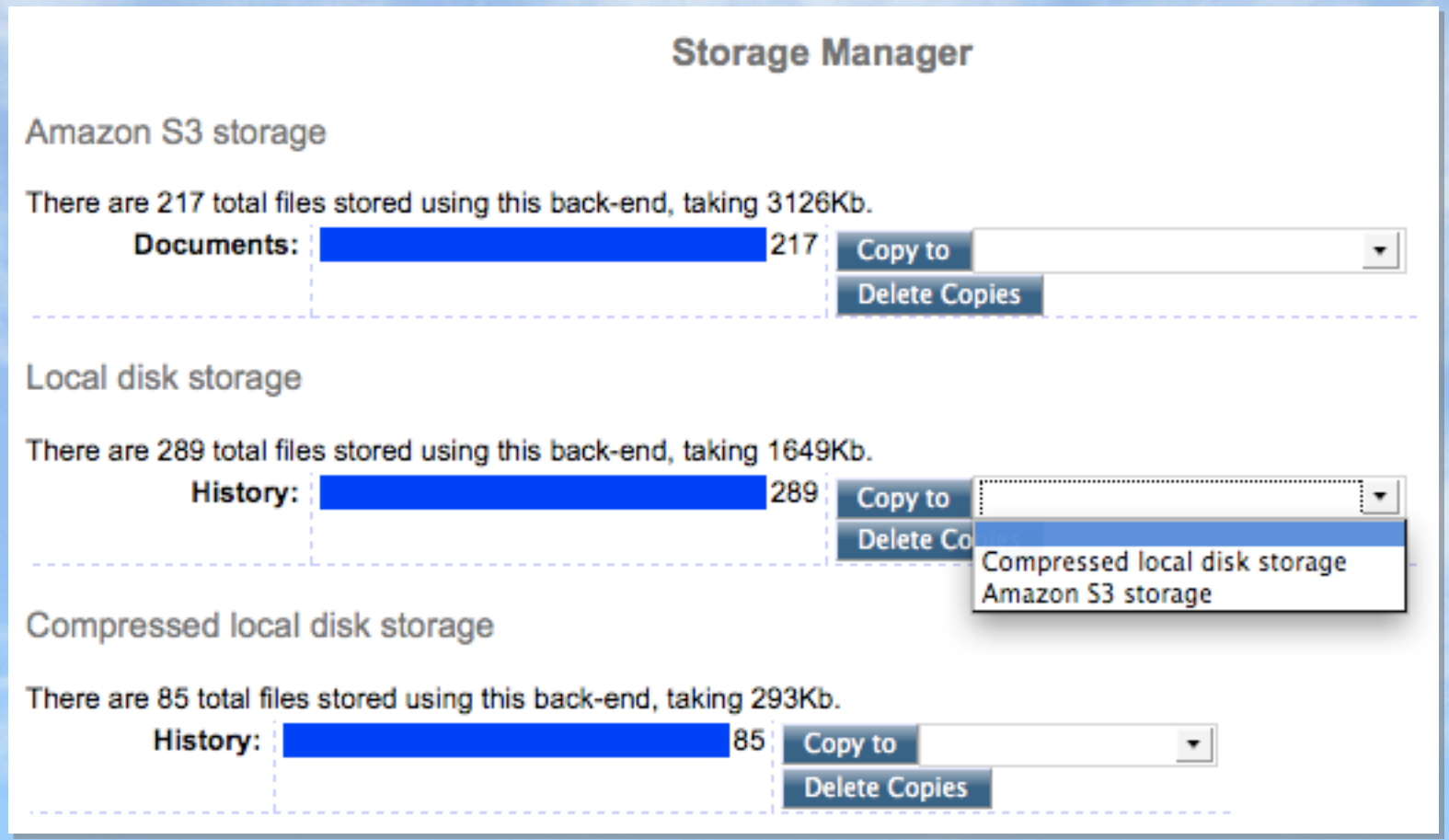
EPrints Storage Controller

- The storage controller manages the location of files.
- Uses rule based policy defined by a simple configuration file (XML)
- Examples:
  - Large binary files of scientific data (raw machine result data) can be stored in a large disk (slower access) system and sent to a tape company for long term storage.
  - Processed results can be stored locally and in the cloud ready for rapid delivery to end points.

Hybrid Storage Policies



EPrints Storage Manager



TU WIEN planets UNIVERSITY OF Southampton eprints

Recap

1. Storage Ecosystem
   - There are a great number of products and services available designed to protect your resources. Each is aimed at a market with different needs based on the type of content.
2. Storage Controller
   - Allows you to utilise a diverse range of storage services simultaneously. Take advantage of the current ecosystem.
3. Managing Stored Assets
   - If the ecosystem changes, moving of resources to a new service is a seamless operation.

The Preservation Process

Preservation - Check

- Resilient Storage
- Bit checking & checksum calculation

Preservation - Analyse

- What is the type of file, is the file valid?
- Is the file at risk of not having an editor/reader?
- Is there a better format available? Lossless or Lossy?

Preservation - Action

- File migration to avert risks found by analysis.
- Movement of file to new storage.

Analysis

Preservation - Analyse

- What is the type of file, is the file valid?
  - Droid is a good classification tool for this.

The National Archives

- Is the file at risk of not having an editor/reader?
  - Functionality is being developed in PRONOM technical registry.

The technical registry PRONOM

- Is there a better format available? Lossless or Lossy?

.

planets

TU WIEN

planets

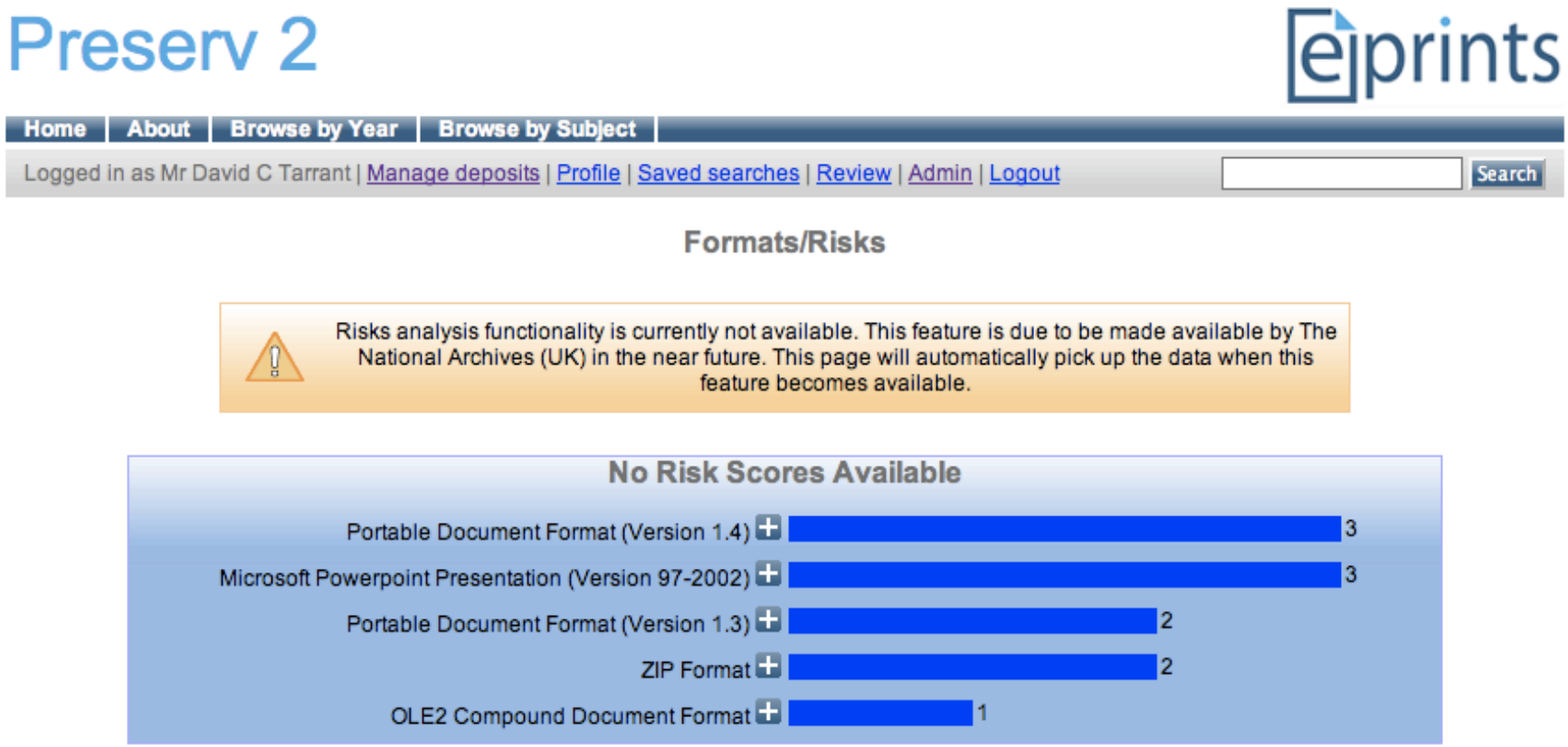
UNIVERSITY OF Southampton

eprints

File Format Analysis

Preservation - Analyse

EPrints File Classification



TU WIEN planets

UNIVERSITY OF Southampton eprints

Risk Analysis

Preservation - Analyse

The technical registry PRONOM

- Is the file at risk of not having an editor/reader?
  - Functionality is being developed in PRONOM technical registry.

- Simple SOAP web service

- Takes file format identification id's, hands back risk score.
- Breakdown of risk score may also be available in future releases.

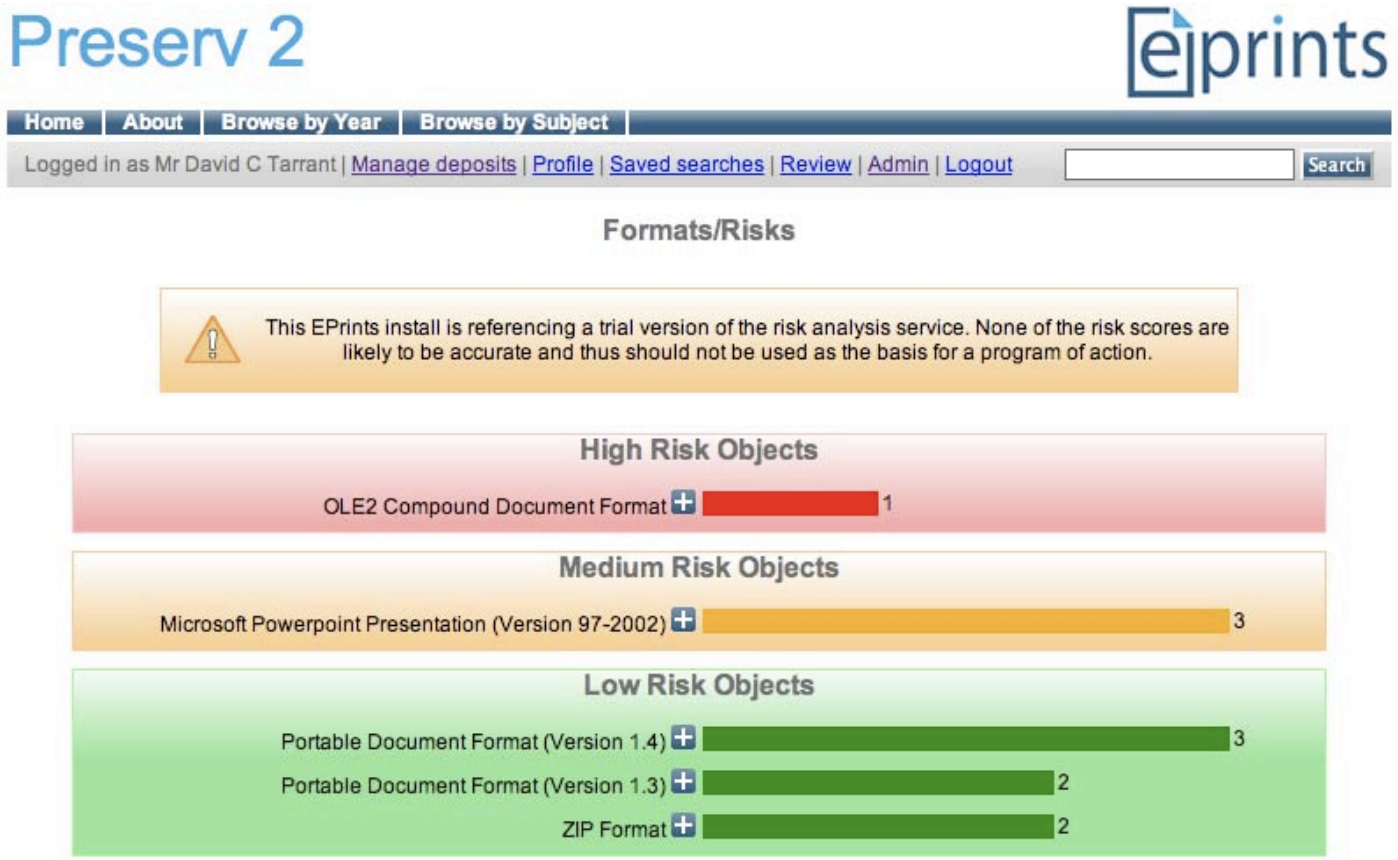
- A stub you can download and run providing this functionality before the official release with mock up risk scores is available at http://preserv2.googlecode.com

TU WIEN planets University of Southampton eprints

Risk Analysis In EPrints

Preservation - Analyse

EPrints File Classification + Risk Analysis



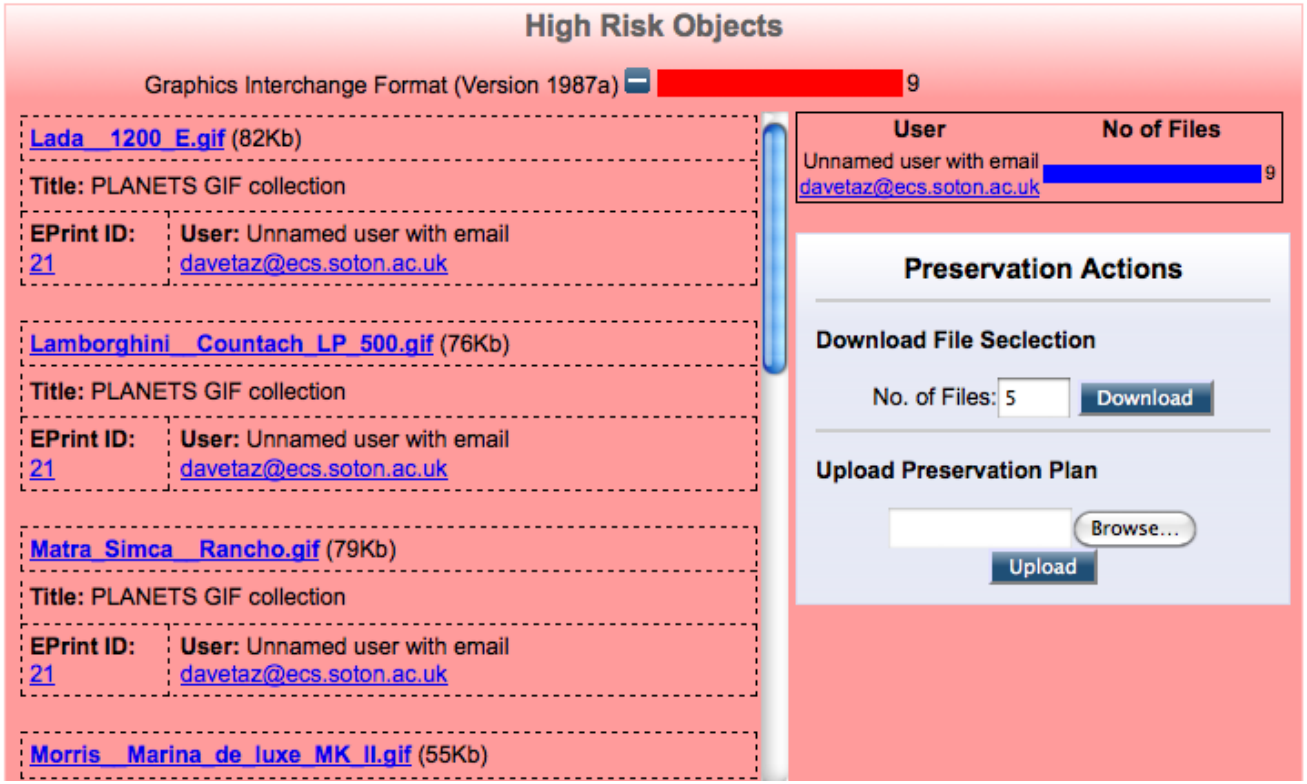
TU WIEN planets

UNIVERSITY OF Southampton eprints

Risk Analysis In EPrints - Detailed View

Preservation - Analyse

EPrints File Classification + Risk Analysis



Exercise Time

Recap

Preservation - Check

- Handled by our storage manager and reported back via the preservation interface.

Preservation - Analyse

- Parallels can be drawn with storage, in that we are integrating with and utilising currently available services to perform our analysis.
- Processing of the results leads to a powerful interface which tells us many things about the repository ecosystem and it's future.

Preservation - Action

- Future plan is to utilise further web based services to ensure information remains comprehensive and up to date set, 0day digital preservation.

Schedule

(1) Introduction
- EPrints
- Preservation Planning and Plato

(2) Preservation in EPrints

(3) Preservation Planning with Plato

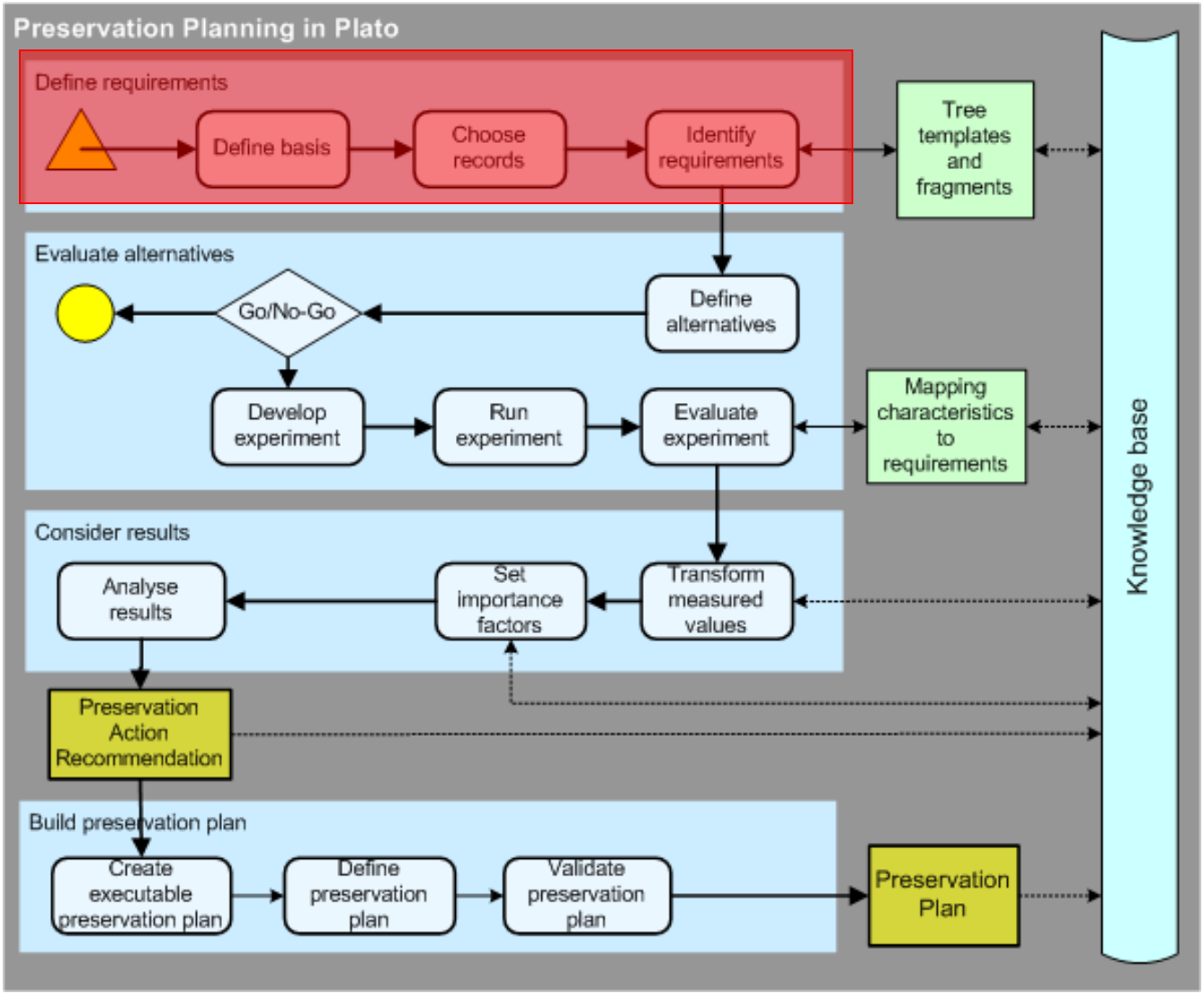
(4) Bringing it all together and Closing

Overview

Part 3: Preservation Planning with Plato

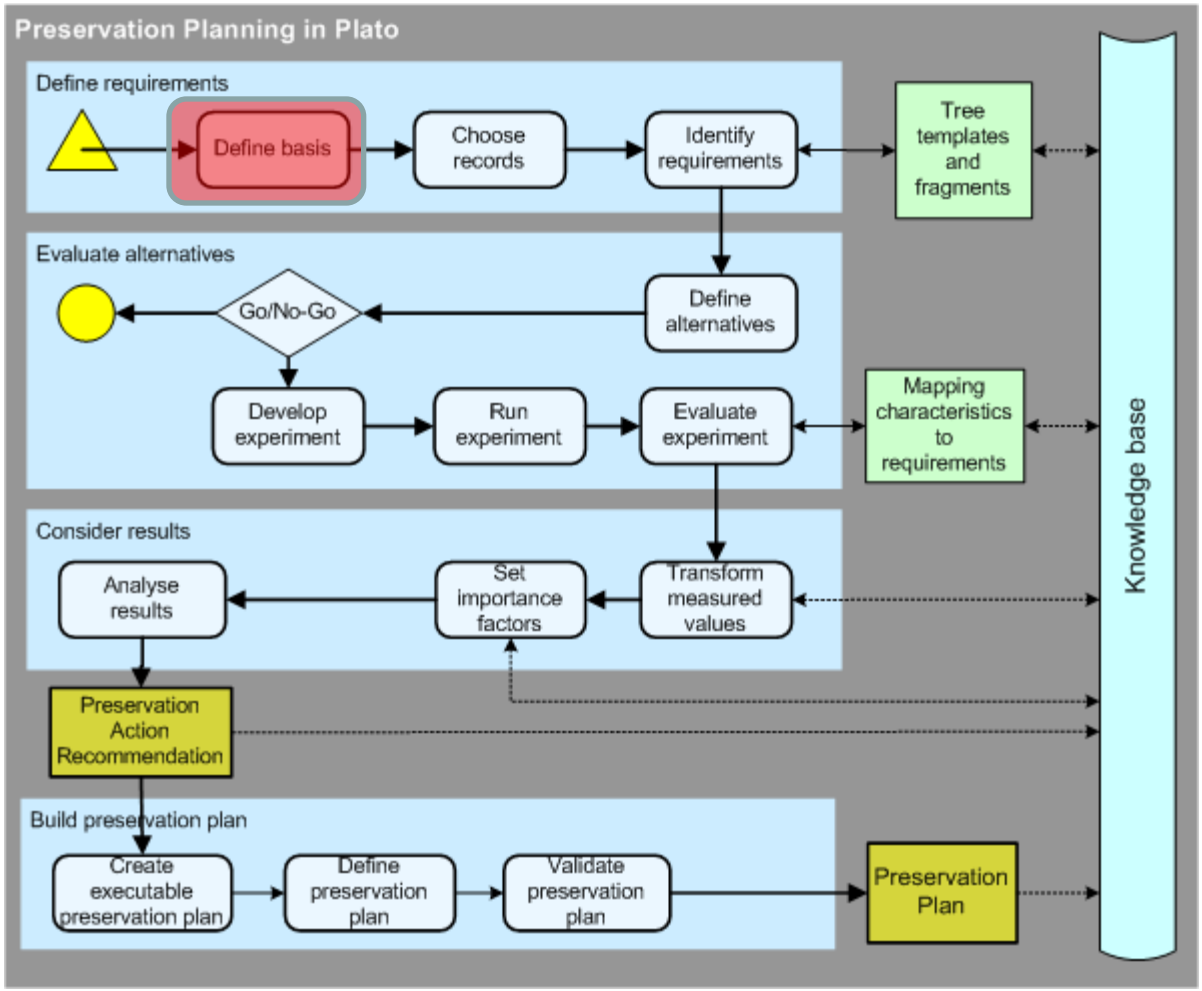
- Preservation planning workflow
- Exercises

PP Workflow

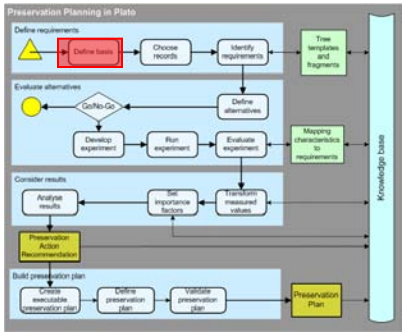


TU WIEN planets UNIVERSITY OF Southampton eprints

Orientation

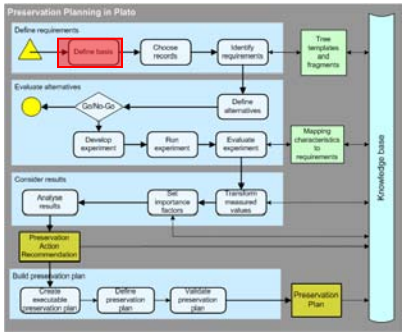


Define Basis



- 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)

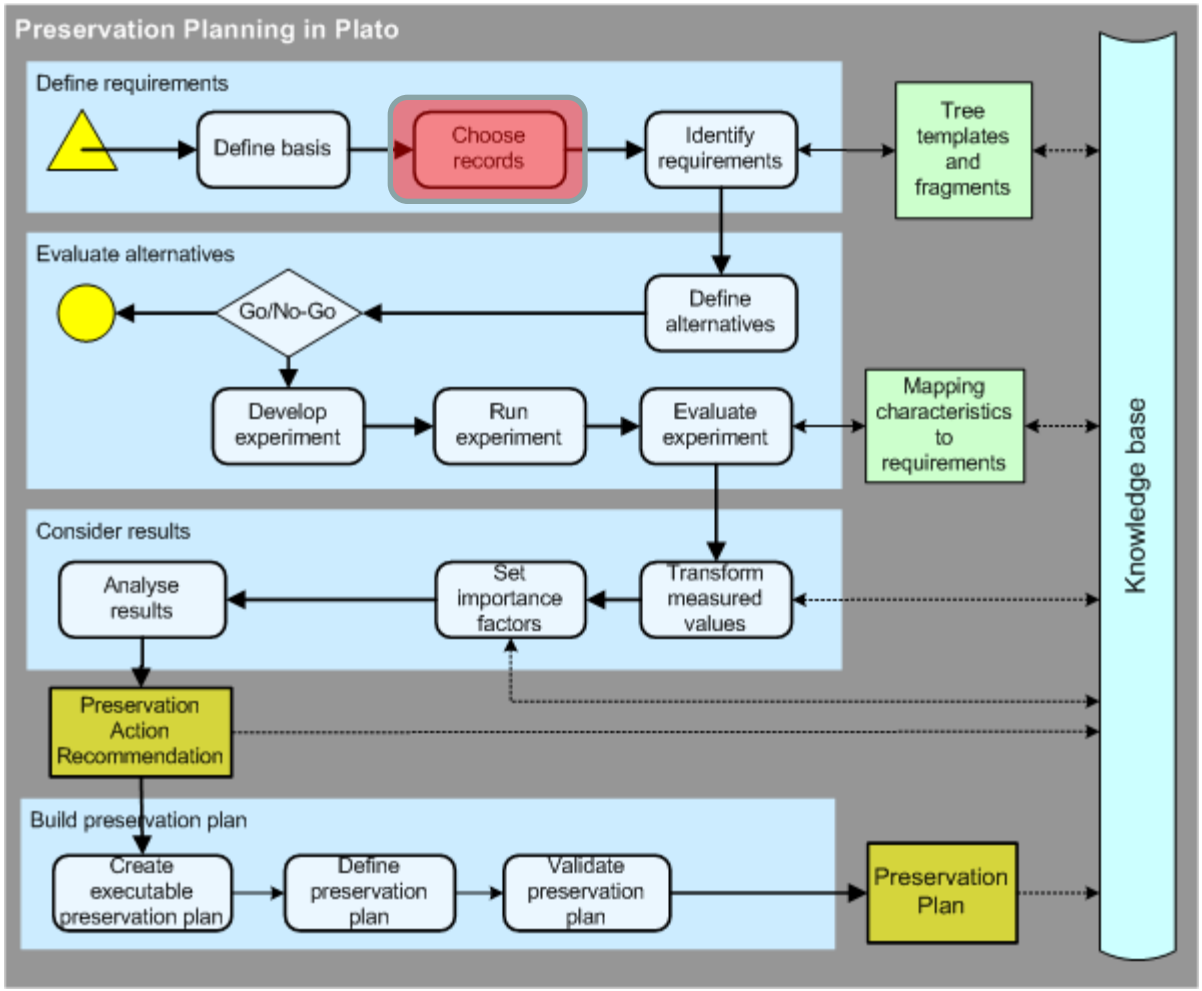
Define Basis



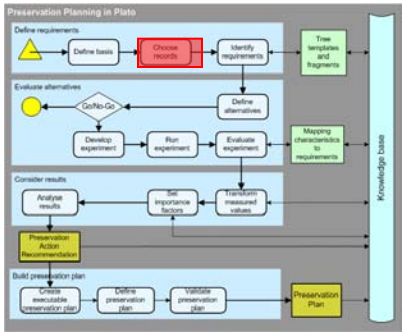
Organizational structure

- Mandate, Mission Statement
  - Provide reliable, long-term access to digital objects
  - Internet Archive: *"The Internet Archive is working to **prevent the Internet** […] and other 'born digital' materials **from disappearing** into the past. Collaborating with institutions including the Library of Congress and the Smithsonian, we are working to preserve a record for generations to come."* http://www.archive.org/about/about.php
  - Oxford Digital Library: *"Like traditional collection development **long-term sustainability** and **permanent availability** are major goals for the Oxford Digital Library."* http://www.odl.ox.ac.uk/principles.htm

Orientation

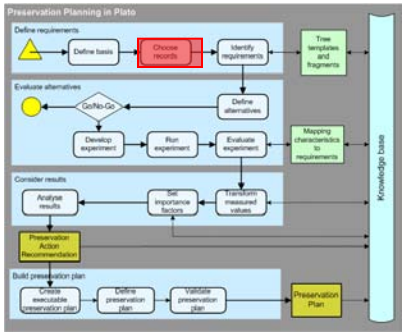


Choose Sample Objects



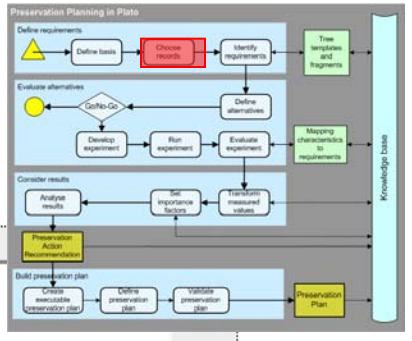
- Identify consistent (sub-)collections
  - Homogeneous type of objects (format, use)
  - To be handled with a specific (set of) tools
- Describe the collection
  - What types of objects?
  - How many?
  - Which format(s)?
- Selection
  - Representative for the objects in the collection
  - Right choice 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

Define Sample Objects



[↑] Sample Records

Description of sample records: several samples of electronic theses ?

Sample Record		Object Format	
Full name:	sample thesis 1 ?	PUID:	?
Short name:	DA1 ?	Name:	?
Has data:	☑ download	Version:	?
Original technical environment:	?	Mime-type:	?
Description:	?		Identify format

Remove record

Save | Discard changes | Proceed

Add new record without file

Add record

Upload new record

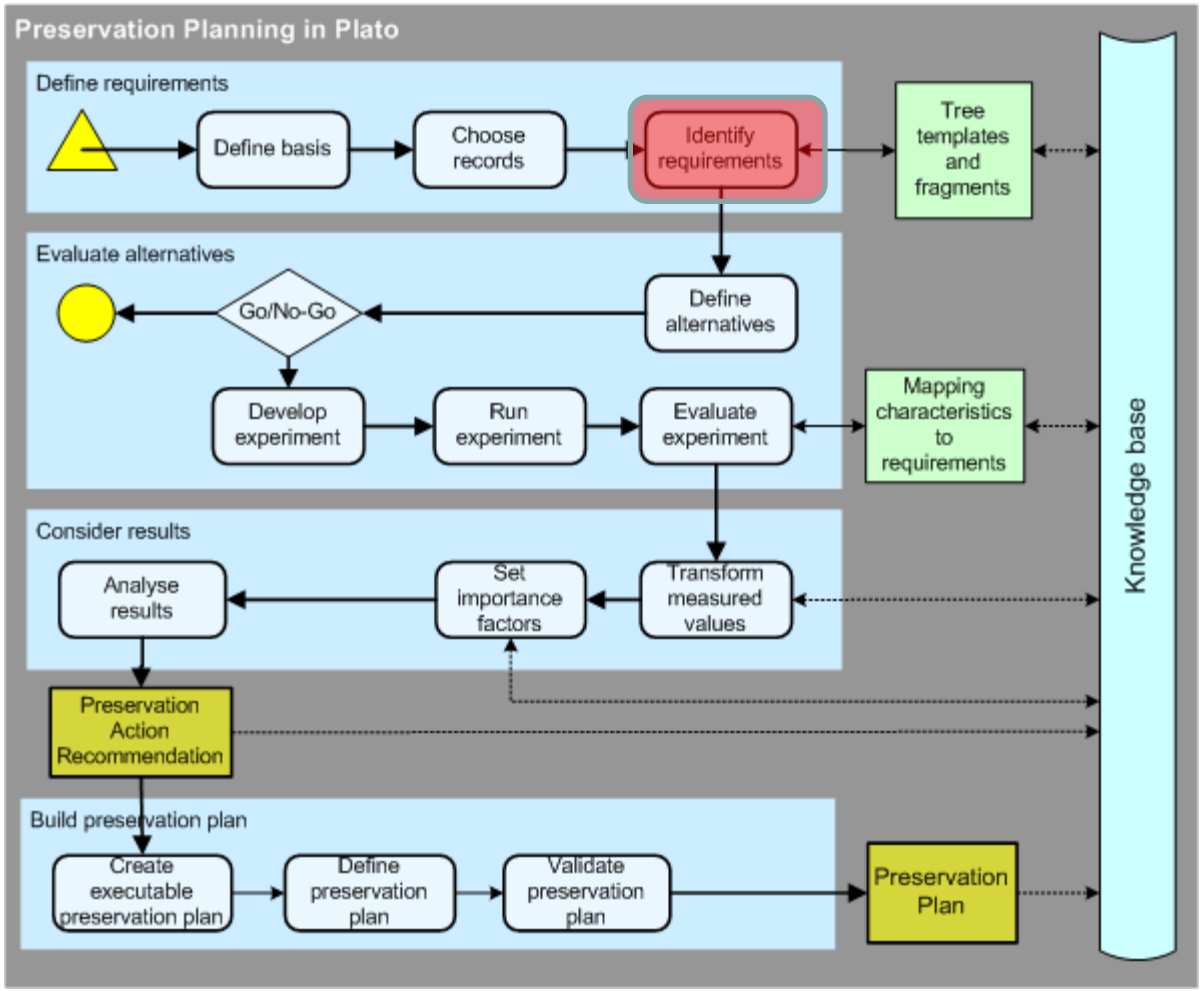
Durchsuchen... | Upload File ?

TU WIEN planets UNIVERSITY OF Southampton eprints

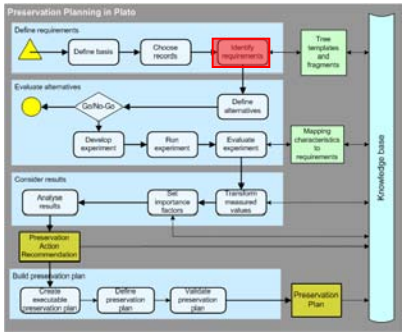
Practise time!

- Public institution – State and University Library
- Mission to preserve the state's cultural heritage in the form of any publication
- Scanned collection of yearbooks, 9000 objects
  - One file per page
  - Scans are black and white
- Copyright held for the physical material, same for digital content
- Objects are provided

Orientation

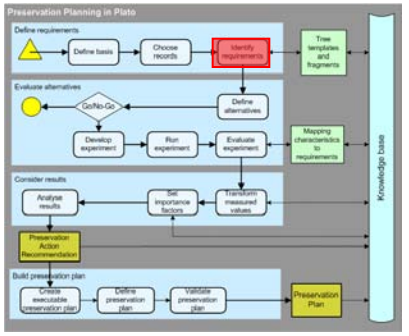


Identify Requirements

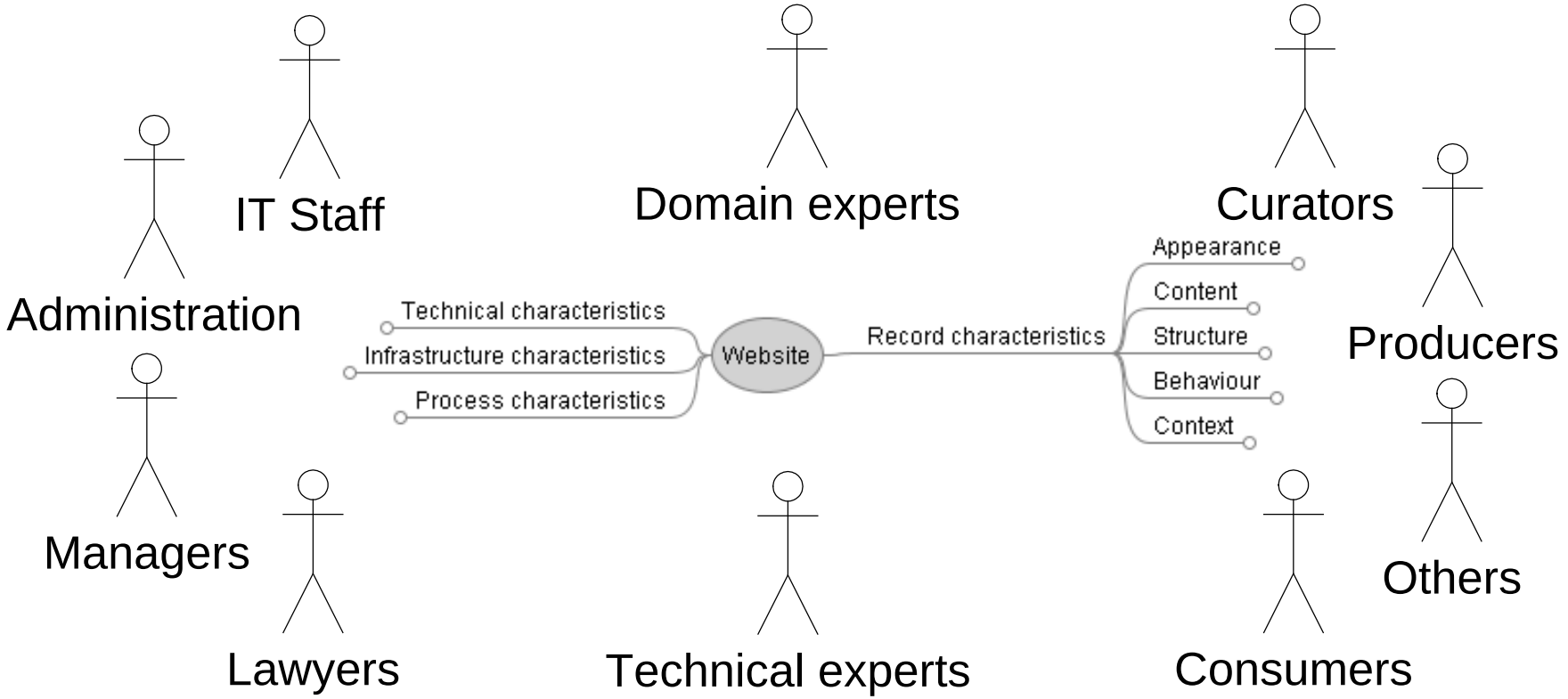


- Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain
- Put the requirements in relation to each other
  → Tree structure
- Top-down or bottom-up
  - Start from high-level goals and break down to specific criteria
  - Collect criteria and organize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

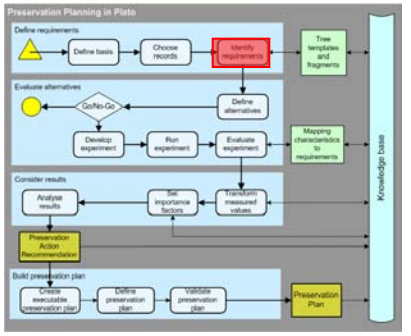
Identify Requirements



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

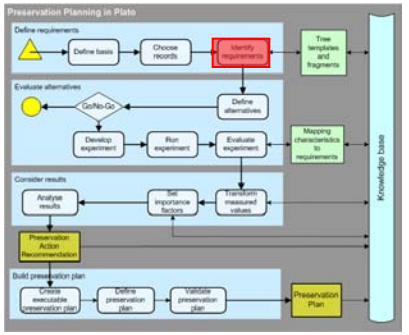


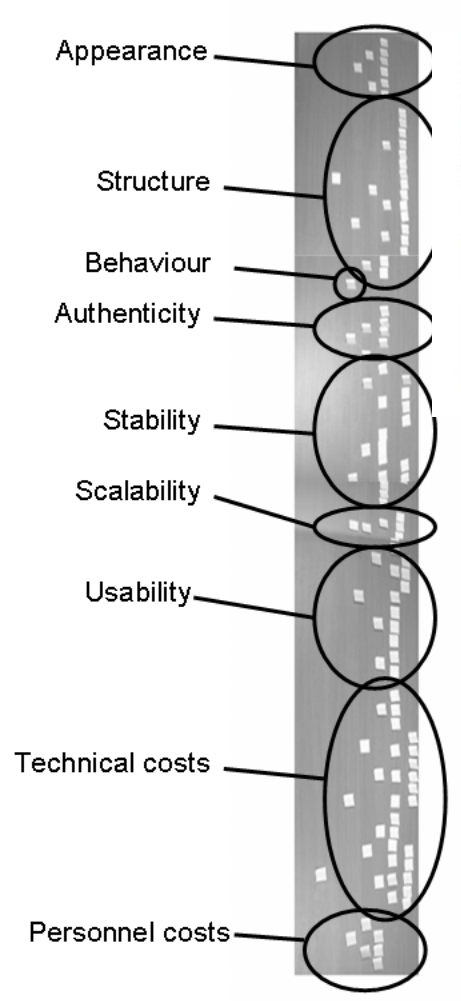
Identify requirements

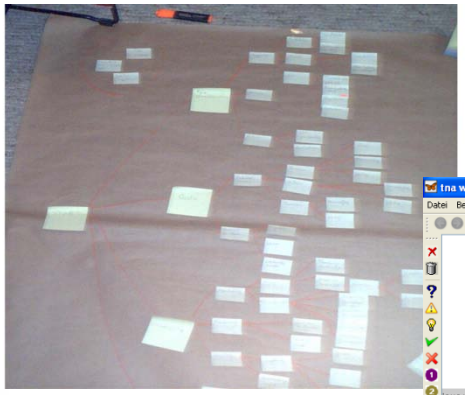


- Core step in the process
- Define all relevant goals and characteristics (high-level, detail) with respect to given application domain
- Usually four major groups
  - Object characteristics (content, metadata,…)
  - Record characteristics (context, relations,…)
  - Process characteristics (scalability,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

Identify requirements

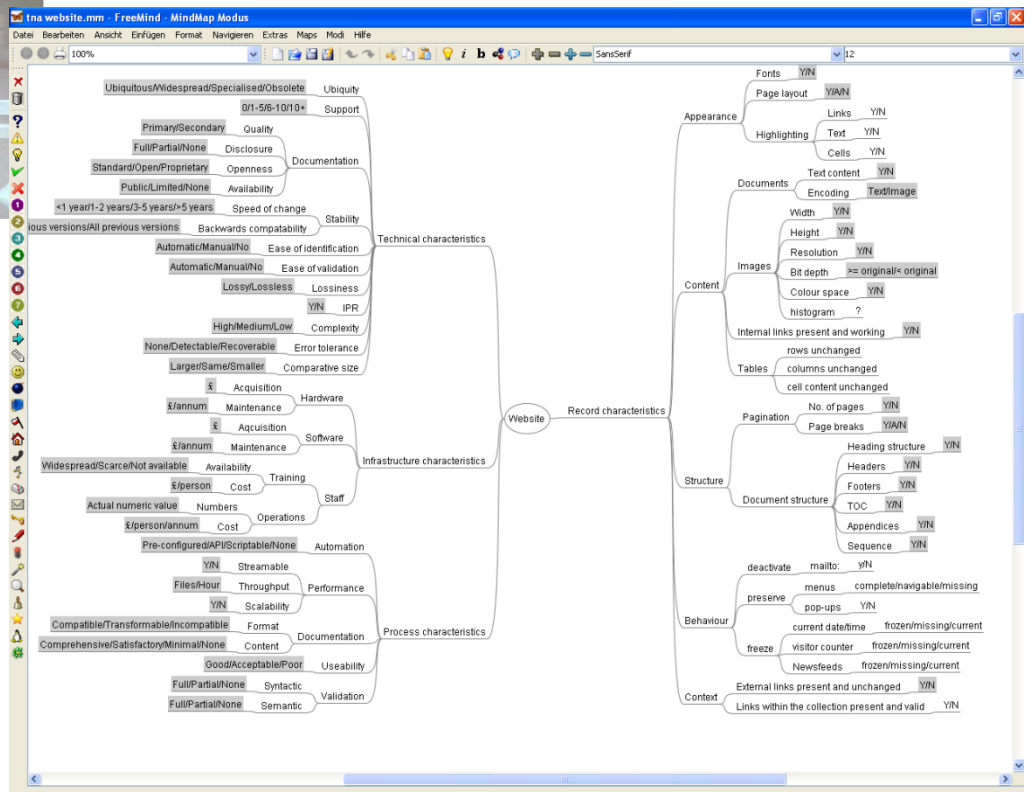






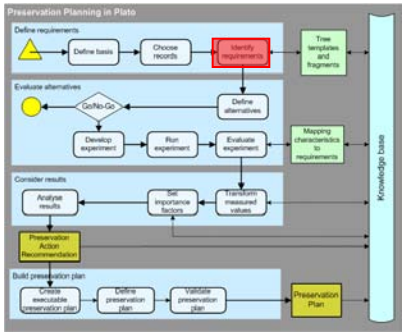
analogue…

… or digital

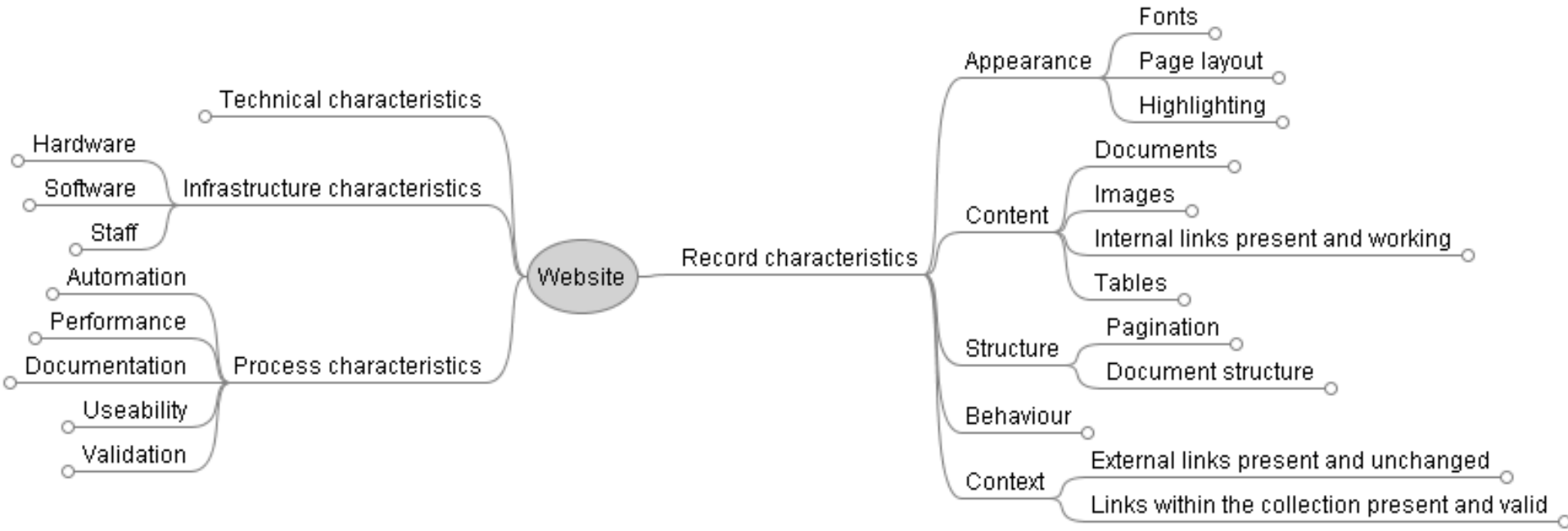


TU WIEN planets UNIVERSITY OF Southampton eprints

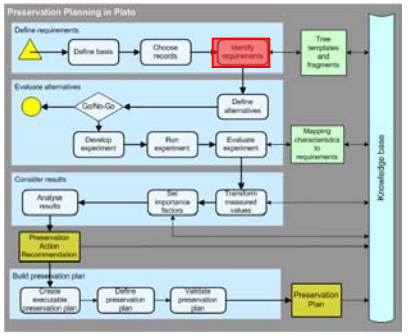
Identify requirements



Example: Webarchive



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

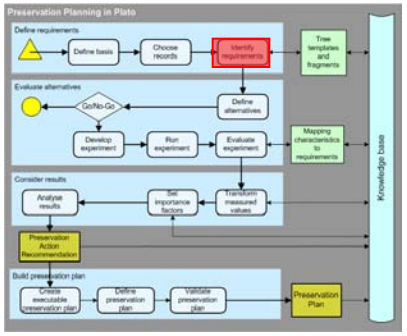
Website

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	
X	▶ Error tolerance	☐	Ordinal	None/Detectable/Recoverable	

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

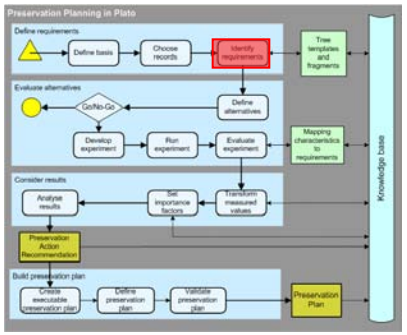
Quick Access:

Identify requirements



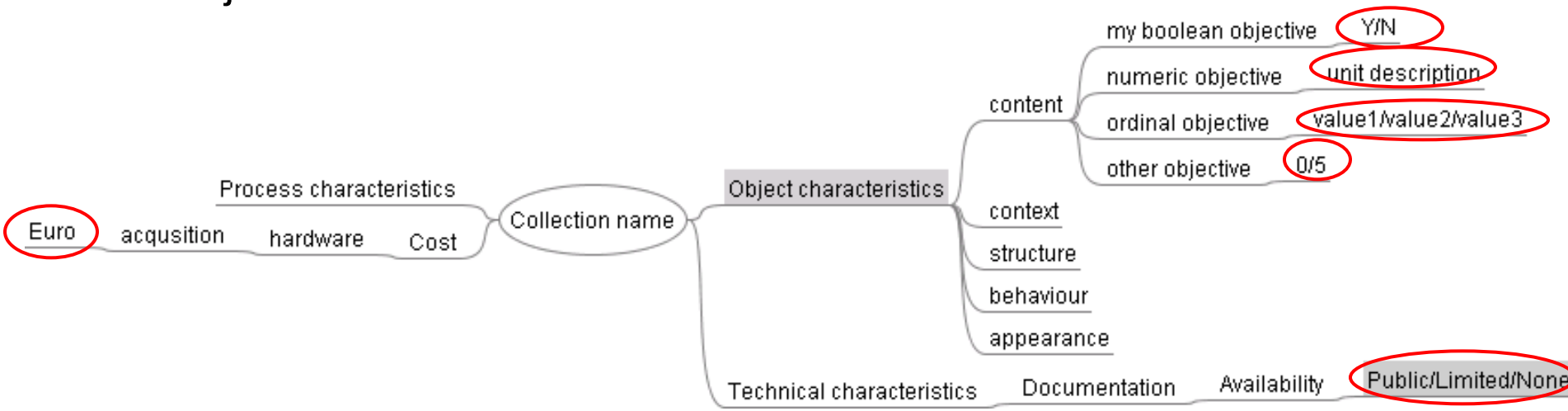
- Assign measurable unit to each leaf criterion
  - As far as possible automatically measurable
    - seconds / Euro per object
    - colour depth in bits
    - ...
  - Subjective measurement units where necessary
    - diffusion of file format
    - amount of expected support
    - ...
- No limitations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

Identify requirements

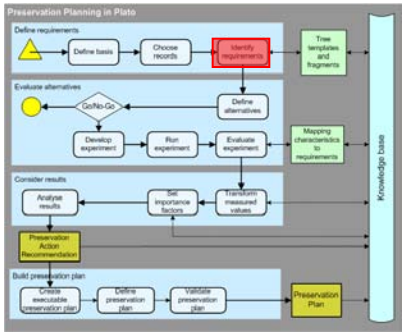


Types of scales

- Numeric
- Yes/No (Y/N)
- Yes/Acceptable/No (Y/A/N)
- Ordinal: define the possible values
- Subjective 0-to-5



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

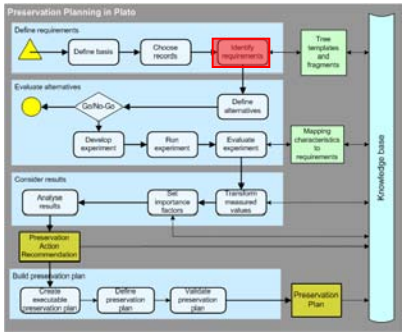
Website

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	
X	[illegible]		Ordinal	None/Detectable/Recoverable	

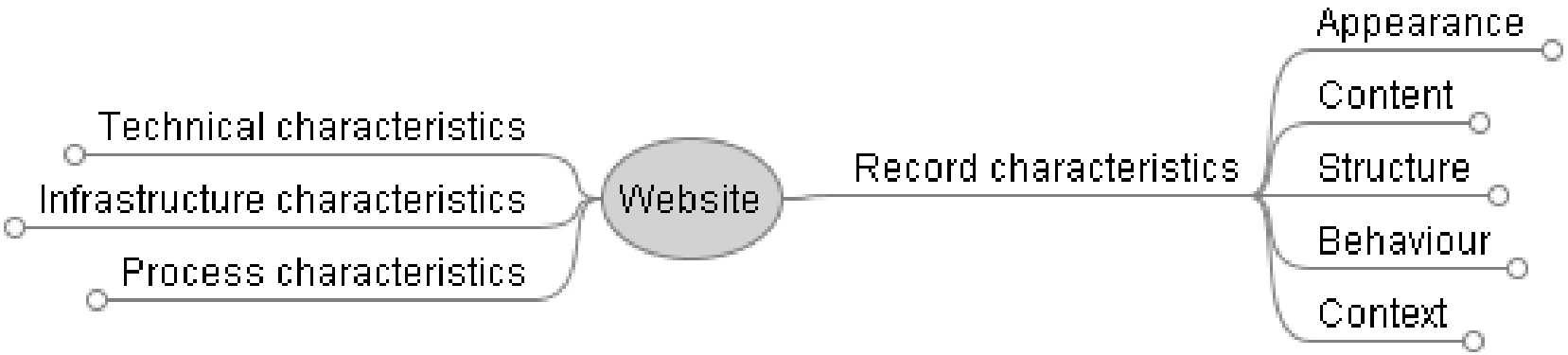
Release 1.1 - Institute of Software Technology and Interactive Systems: «off-ice bears»

Quick Access:

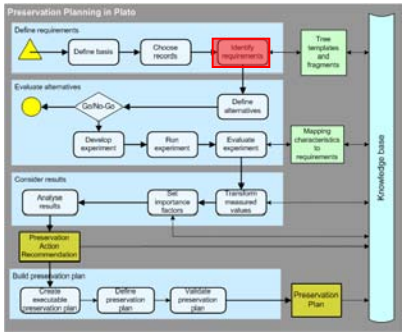
Identify Requirements: Example

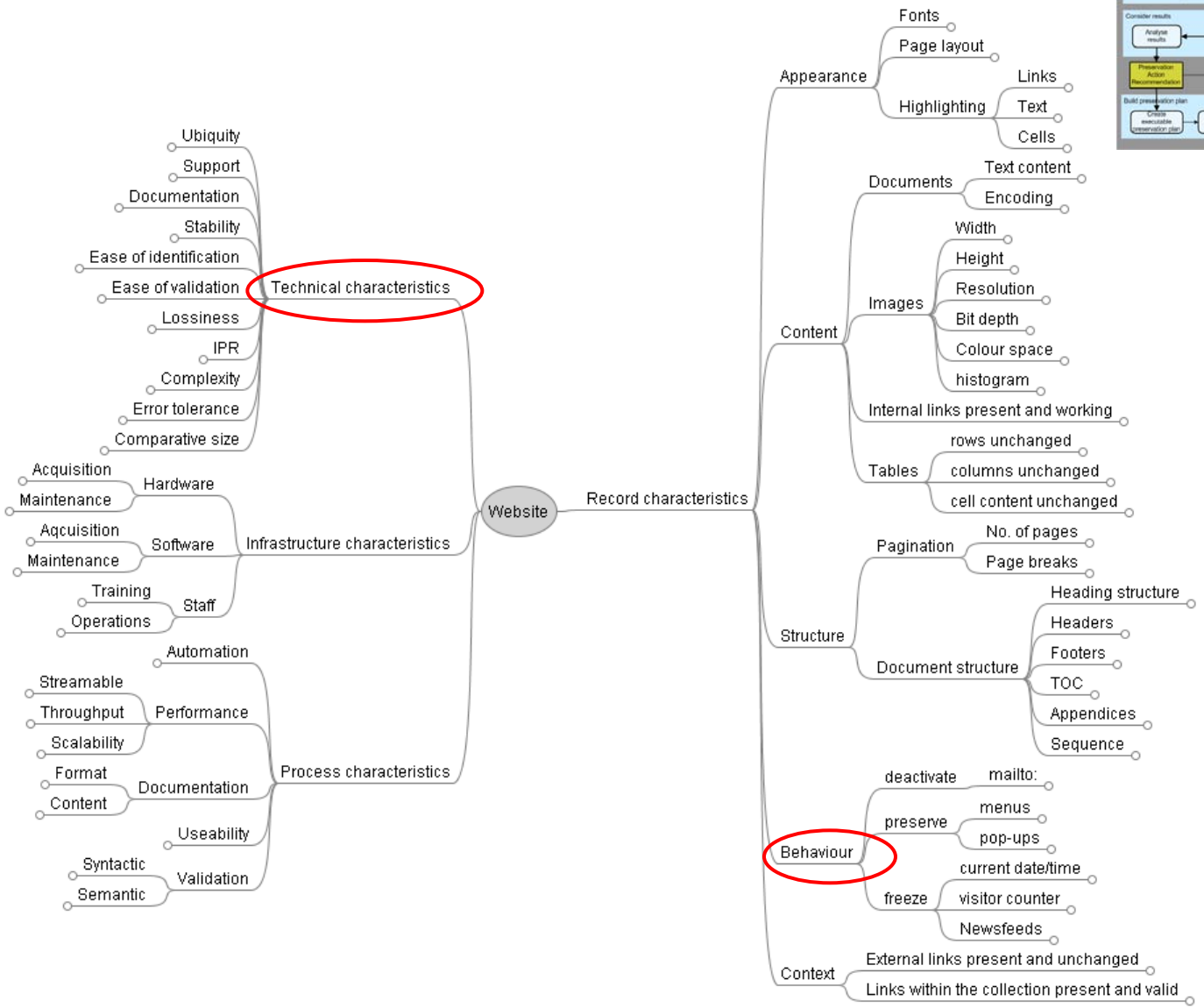


- Example Webarchiving:
  - Static Webpages
  - Including linked documents such as doc, pdf
  - Images
  - Interactive elements need not be preserved

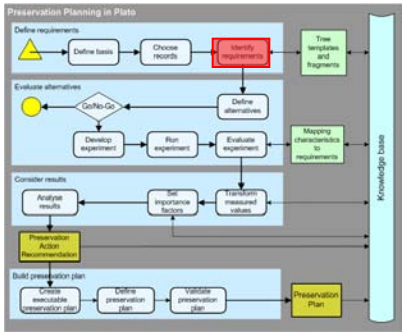


Identify Requirements: Example



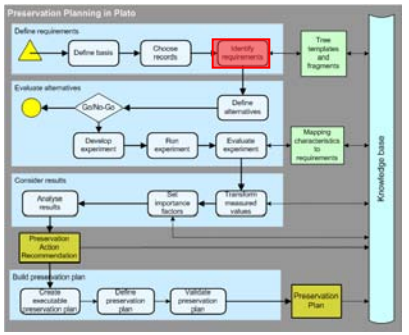


Identify Requirements: Example

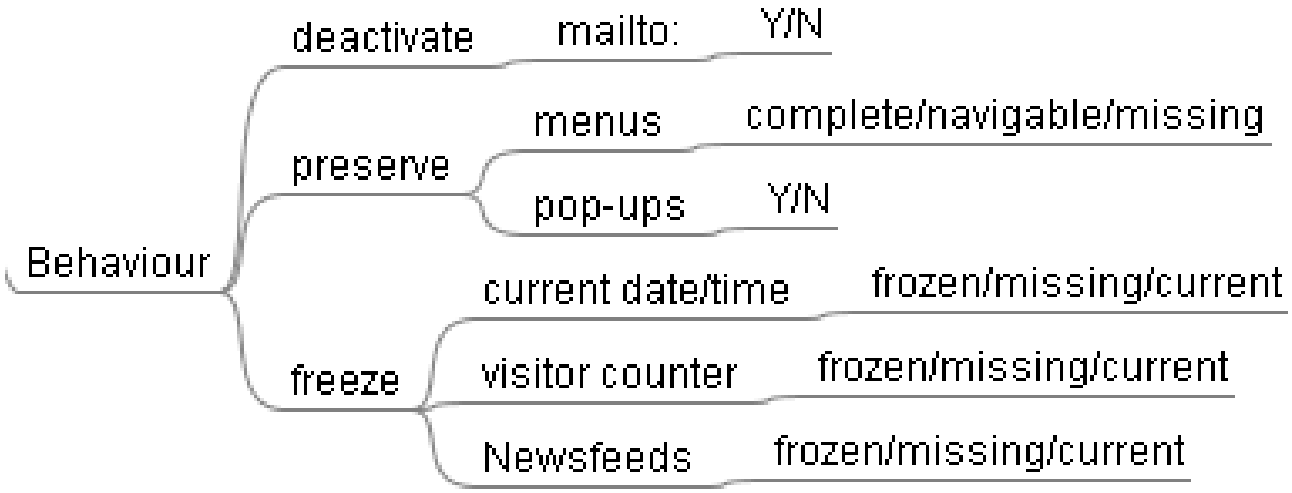


- Technical characteristics
  - Ubiquity: Ubiquitous/Widespread/Specialised/Obsolete
  - Support: number of tools
  - Documentation
    - Quality: Primary/Secondary
    - Disclosure: Full/Partial/None
    - Openness: Standard/Open/Proprietary
    - Availability: Public/Limited/None
  - Stability
    - 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?)

Practise time!

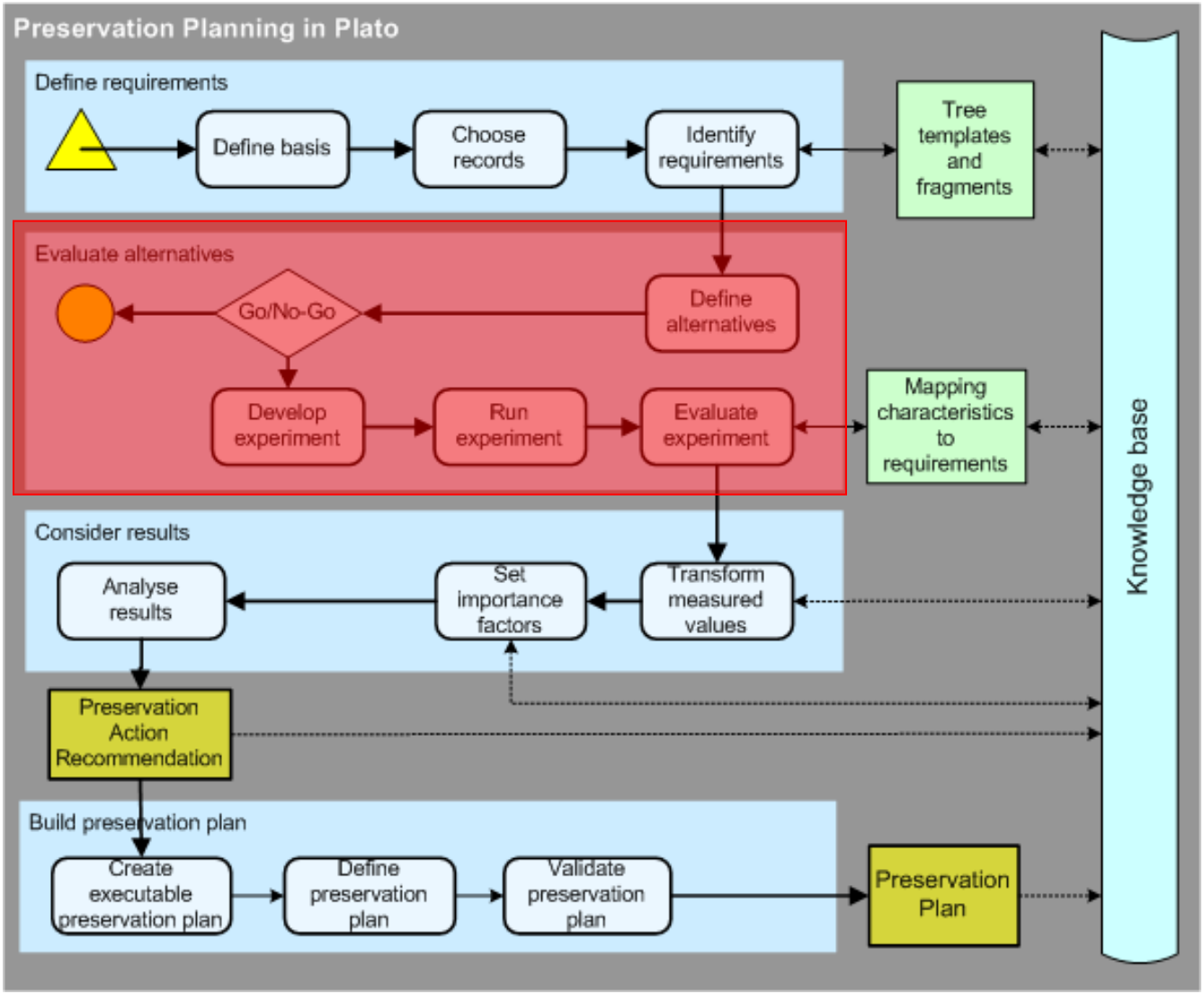
- Go to Plato: http://www.ifs.tuwien.ac.at/dp/plato
- Log into Plato with group account
- Click “List my preservation plans”

- Open preservation plan named

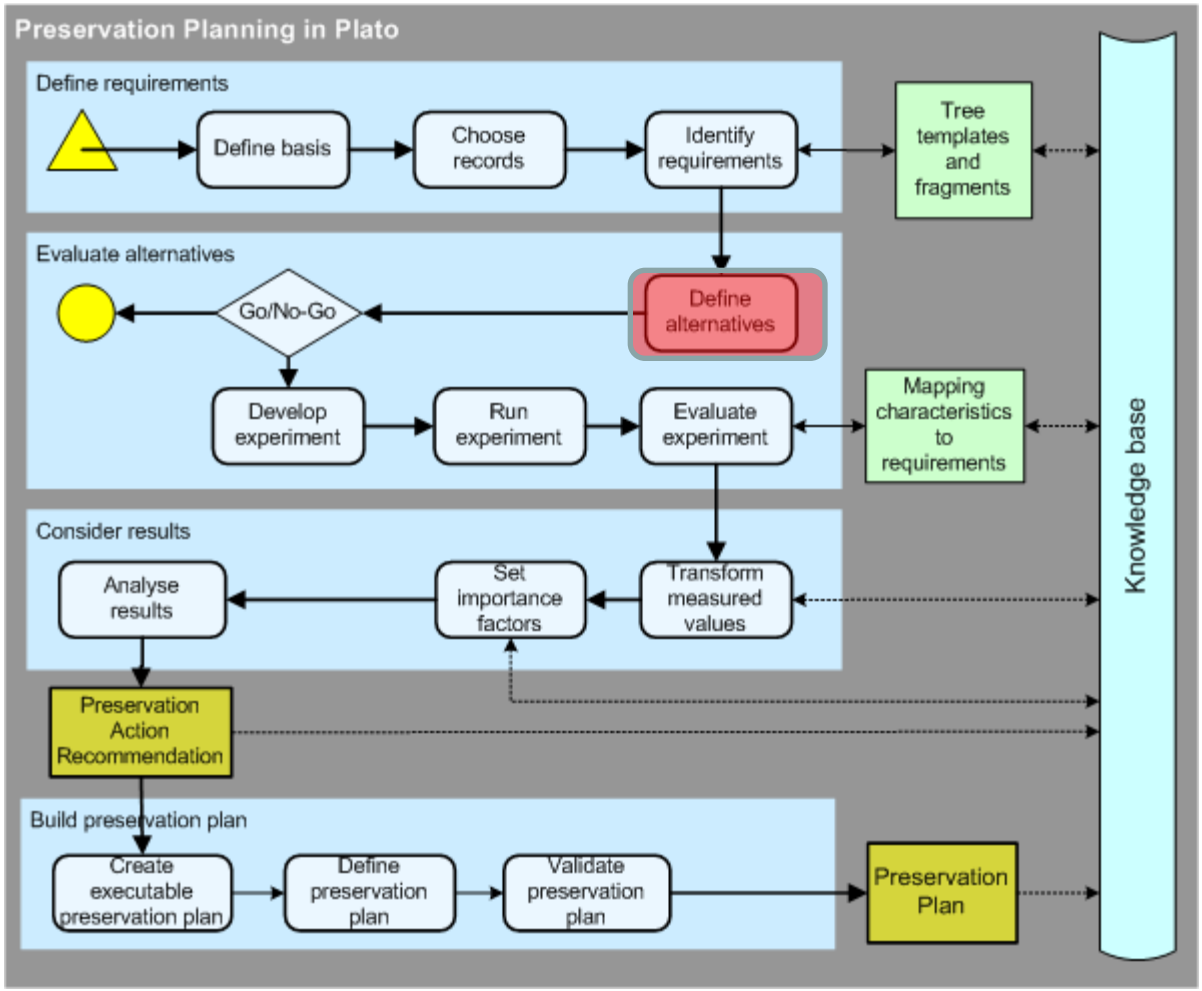
  “**Scanned yearbooks archive (IDENTIFY REQUIREMENTS)”**

- Enter further requirements

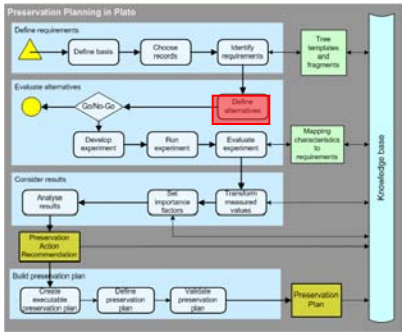
PP Workflow



Orientation

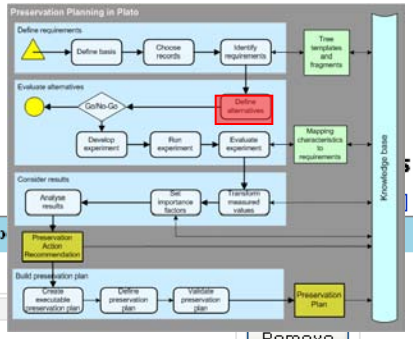


Define Alternatives



- Given the type of object and requirements, what strategies are possible and which is most suitable
  - Migration, emulation, other?
- For each alternative, precise definition of
  - Which tool (OS, version)
  - Which functions of the tool
  - Which parameters
  - Resources that are needed (human, technical, time and cost)
- Define manually or use registries via web services

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

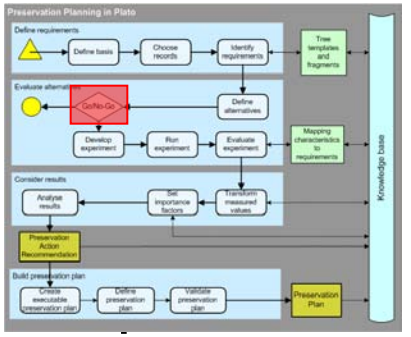
CRiB Service Registry

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

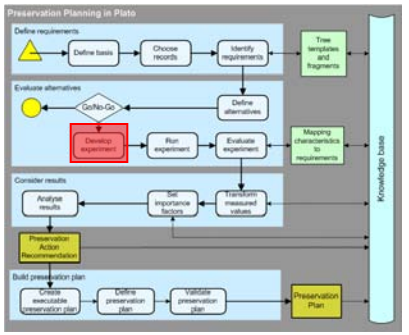
Create alternatives for selected services

Go/No-Go



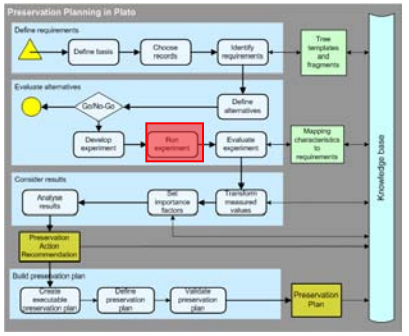
- Deliberate step for taking a decision if it will be useful and cost-effective to continue the procedure, given
  - The resources to be spent (people, money)
  - The availability of tools and solutions,
  - The expected result(s).
- Review of the experiment/ evaluation process design so far
  - Is the design complete, correct and optimal?
- Need to document the decision
- If insufficient: can it be redressed or not?
- Decision per alternative: go / no-go / deferred-go

Develop experiment



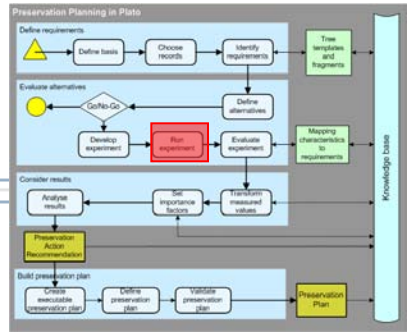
- Plan for each experiment
  - steps to build and test SW components
  - HW set-up
  - Procedures and preparation
  - Parameter settings, capturing measurements (time, logs...)
- Standardized Testbed-environment simplifies this step (PLANETS Testbed)
- Ideally directly accessible Preservation Action Services
- Ensures that results are comparable and repeatable

Run experiment



- Before running experiments: Test
- Call migration / emulation tools
- Local or service-based
- Capture process measurements
  (Start-up time, time per object, throughput, ...)
- Capture resulting objects, system logs, error messages,…

Develop and Run Experiment



JPG > JP2

Save | Discard changes | Proceed

Result Files

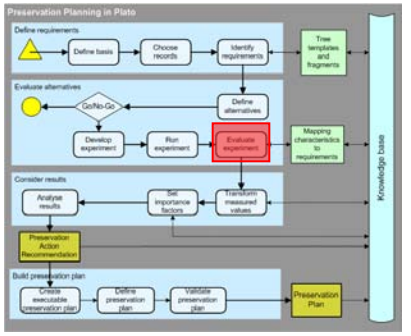
Results	Sample Records		
TIFF (tool A) ?	Polar bear 1	Upload File	
	Polar bear 2	Upload File	
TIFF (tool B) ?	Polar bear 1	Upload File	
	Polar bear 2	Upload File	
GIF (tool C) ?	Polar bear 1	Upload File	
	Polar bear 2	Upload File	
PNG (tool D) ?	Polar bear 1	Upload File	
	Polar bear 2	Upload File	
JPG > BMP ?	Polar bear 1	Upload File	Polar bear 1.bmp Download File
	Polar bear 2	Upload File	Polar bear 2.bmp Download File
JPG > TIF ?	Polar bear 1	Upload File	
	Polar bear 2	Upload File	

Release 1.2 - Institute of Software Technology and Interactive Systems: « off-ice bears »

TU WIEN planets UNIVERSITY OF Southampton eprints

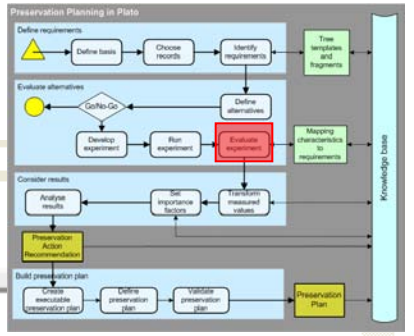
Demo!

Evaluate experiment



- Analyse the results according to the criteria specified in the Objective Tree
- Preservation Characterization: *Characterization Services*
- Evaluation analyses
  - Experiment measurements, results
  - Necessity to repeat an experiment
  - Undesired / unexpected results
- Technical and intellectual aspects

Evaluate Experiment



Designated community:

Durchsuchen...

Applying policies:

Durchsuchen...

Release 1.2 - Institute of Software Technology and Interactive Systems: « off-ice bears »

Figure 8 Define

In the first step of the workflow (Figure 8), the number environment are described in the text boxes. The proje

Evaluate Experiment

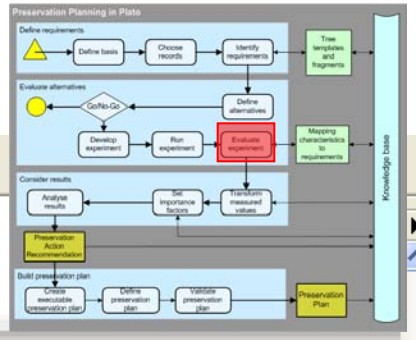
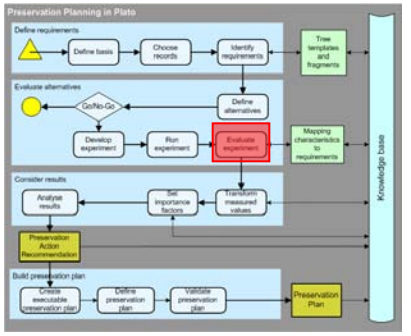


Figure 8 Define

In the first step of the workflow (Figure 8), the number
environment are described in the text boxes. The proje

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 *	

TU WIEN planets UNIVERSITY OF Southampton eprints

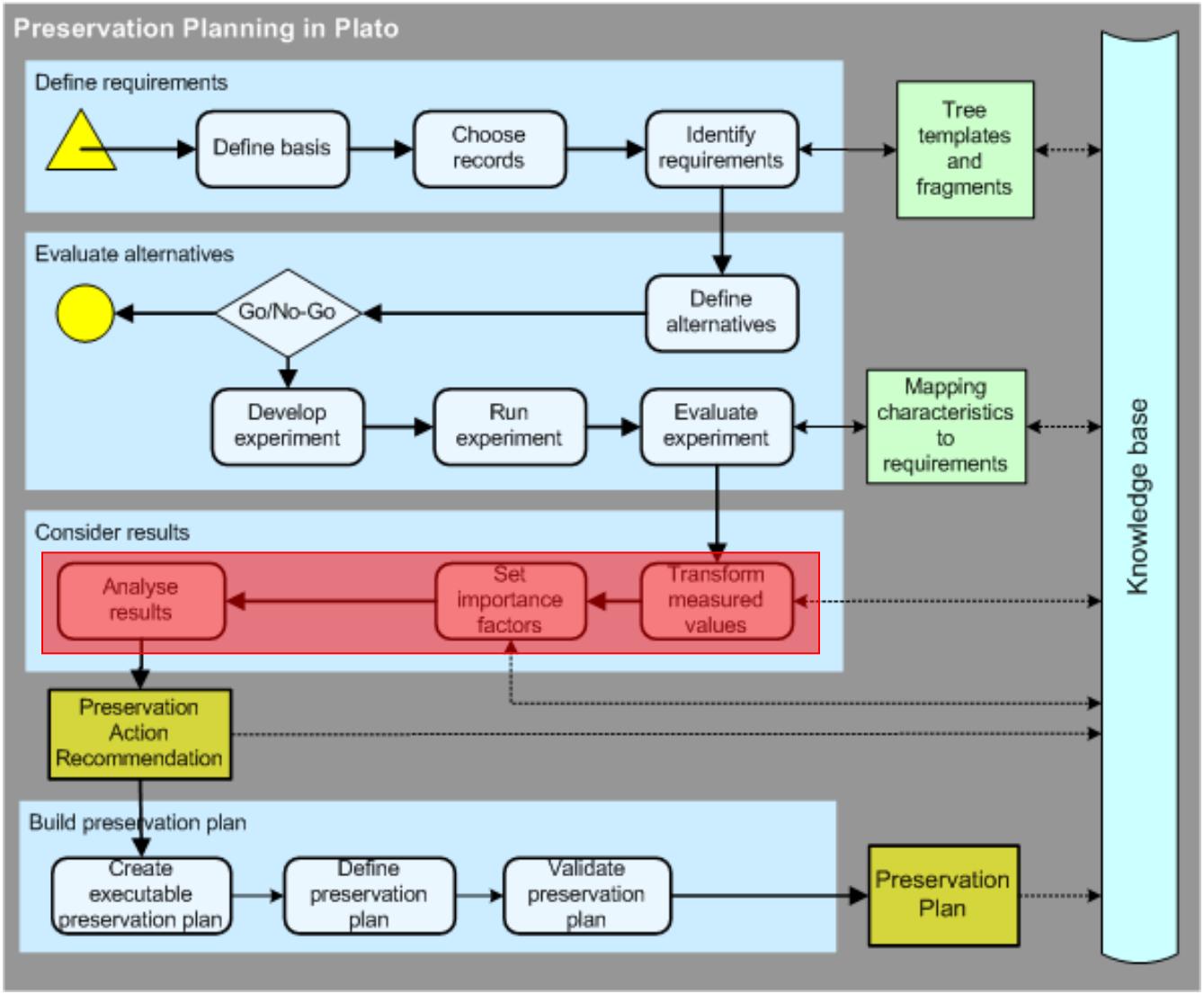
Practise time!

- Log into Plato at: http://www.ifs.tuwien.ac.at/dp/plato
- Download http://www.ifs.tuwien.ac.at/~kulovits/sample-files.zip
- Download http://www.ifs.tuwien.ac.at/~kulovits/experiment-results.zip
- Open preservation plan named

  “Scanned yearbooks archive (EVALUATE EXPERIMENTS)”

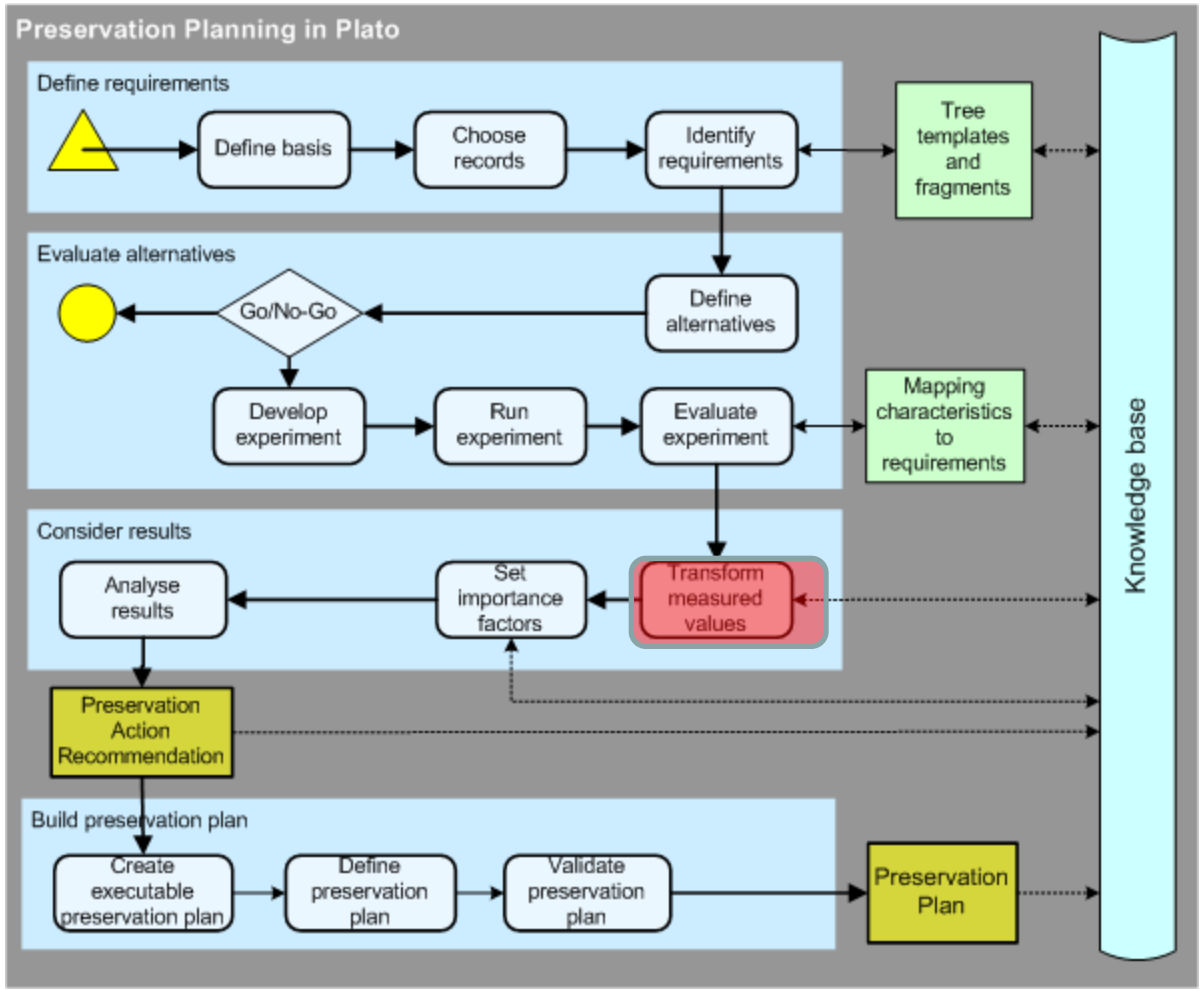
- Evaluate requirements

PP Workflow

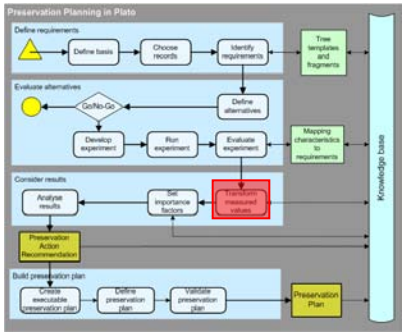


TU WIEN planets UNIVERSITY OF Southampton eprints

Orientation

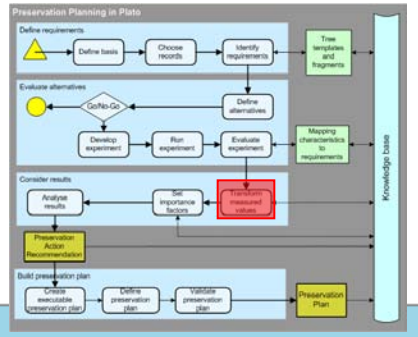


Transform measured values



- 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"

Transform Measured Values



PLANETS Preservation Planning Tool *(Plato)*

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

Transform Measured Values

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: The limit for the process cost was decided to be 150¢ per picture in the last polar-bear-enthusiasts-association-

Technical characteristics > Official standard

Ordinal Value	Target Value
Yes	-> 5.0 *
No	-> 0.0 *

Results	Single
TIFF (tool A)	Yes
TIFF (tool B)	Yes
GIF (tool C)	No
PNG (tool D)	Yes

Aggregation mode:

(•) Worst result ( ) Arithmetic mean

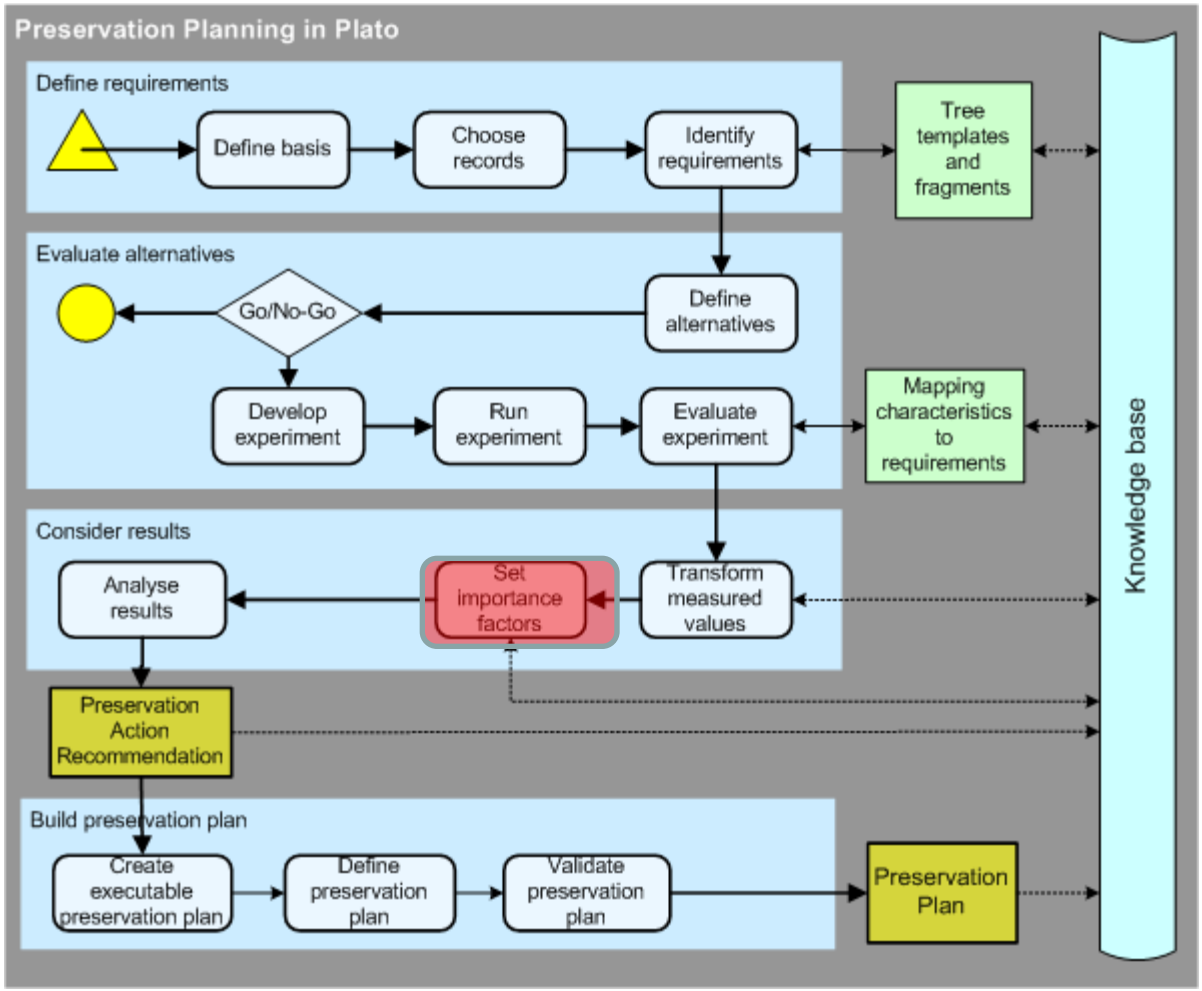
Technical characteristics > Filesize (in Relation to Original)

Threshold	Target value
10.0 *	x -> 1
5.0 *	x -> 2
2.0 *	x -> 3
1.2 *	x -> 4
0.8 *	x -> 5

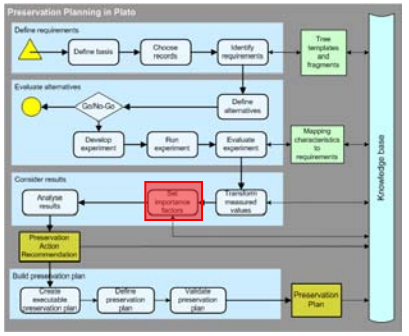
Results	1	2
TIFF (tool A)	2.73	2.6
TIFF (tool B)	2.3	3.1
GIF (tool C)	1.1	0.9
PNG (tool D)	1.3	1.5

TU WIEN planets UNIVERSITY OF Southampton eprints

Orientation

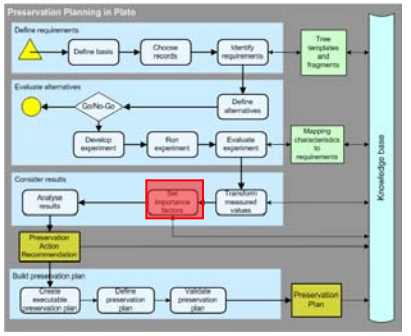


Set Importance Factors



- 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

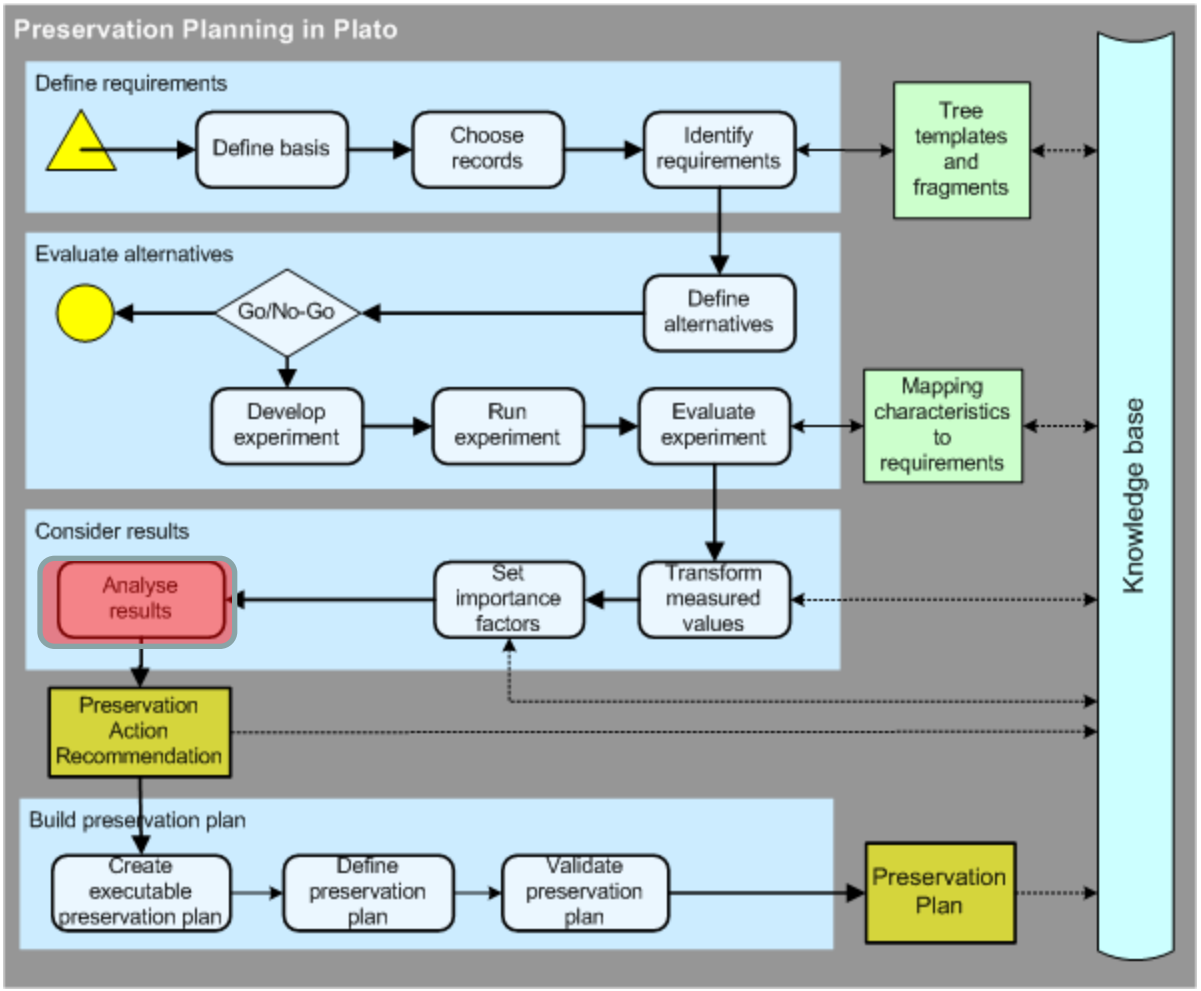
Set Importance Factors



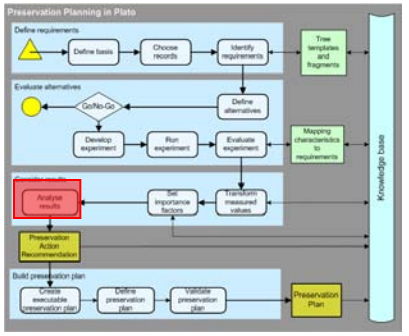
- Object characteristics
  - 25% structure
    - 35% menus and navigation path
    - 35% complete
    - 30% overall page layout
  - 10% appearance
    - 25% mouse
      - animated pointer
    - 20% transitions
      - speed
      - effects
        - graphic
        - sound
    - 15% animations
      - speed
    - 5% background
      - colour gradient
    - 25% menu speed
    - 10% fonts
      - type
      - colour
      - style
      - size
  - 15% behaviour
    - 10% navigation
    - 80% reaction to activity
      - mouse
      - keyboard
    - 10% video/sound control
  - 10% context
    - 20% documentation material
    - 80% metadata reference valid
  - 40% content
    - 22% sound
      - Loops
      - Effects
      - content identical
      - quality
    - 22% video
      - sound
      - picture
      - synchronisation
    - 22% image
      - resolution
      - colour
      - position
      - size
      - content identical
      - quality
    - 22% text
      - content identical
      - line breaks
    - 12% user manual

TU WIEN planets UNIVERSITY OF Southampton eprints

Orientation

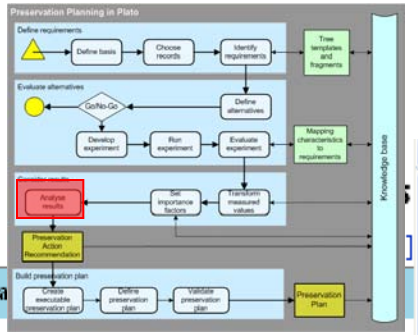


Analyse results



- Aggregate values in Objective Tree
  - Multiply transformed measurements in leaves with weights
  - Sum up across tree
- Results in accumulated performance value per alternative at root level
  → ranking of alternatives
- Also results in performance value for each alternative in each sub-branch of the tree
  → combination of alternatives
- Basis for well-informed and accountable decisions
- Different aggregation methods, e.g. sum and multiplication

Analyse Results



PLANETS Preservation Planning Tool *(Plato)*

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

Analyse Results

Aggregation method: Sum

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

Show

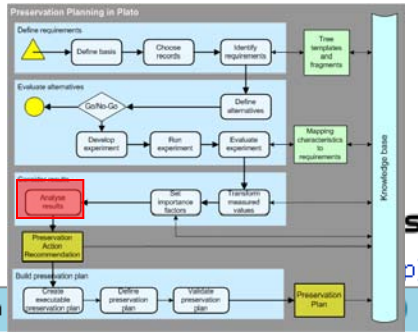
Expand All | Collapse All

Minimalist root node

Focus	Name	Result
	▼Minimalist root node	PDF/A ToolA: 2,88 PDF/A ToolB: 3,19
X	▶Image properties	PDF/A ToolA: 0,60 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

TU WIEN planets UNIVERSITY OF Southampton eprints

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

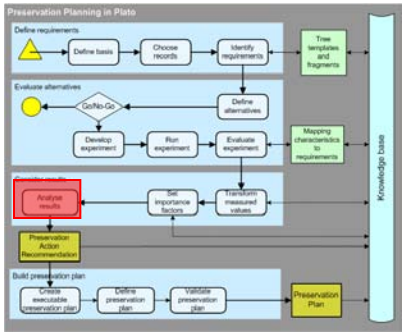
Expand All | Collapse All

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

TU WIEN planets UNIVERSITY OF Southampton eprints

Analyse results



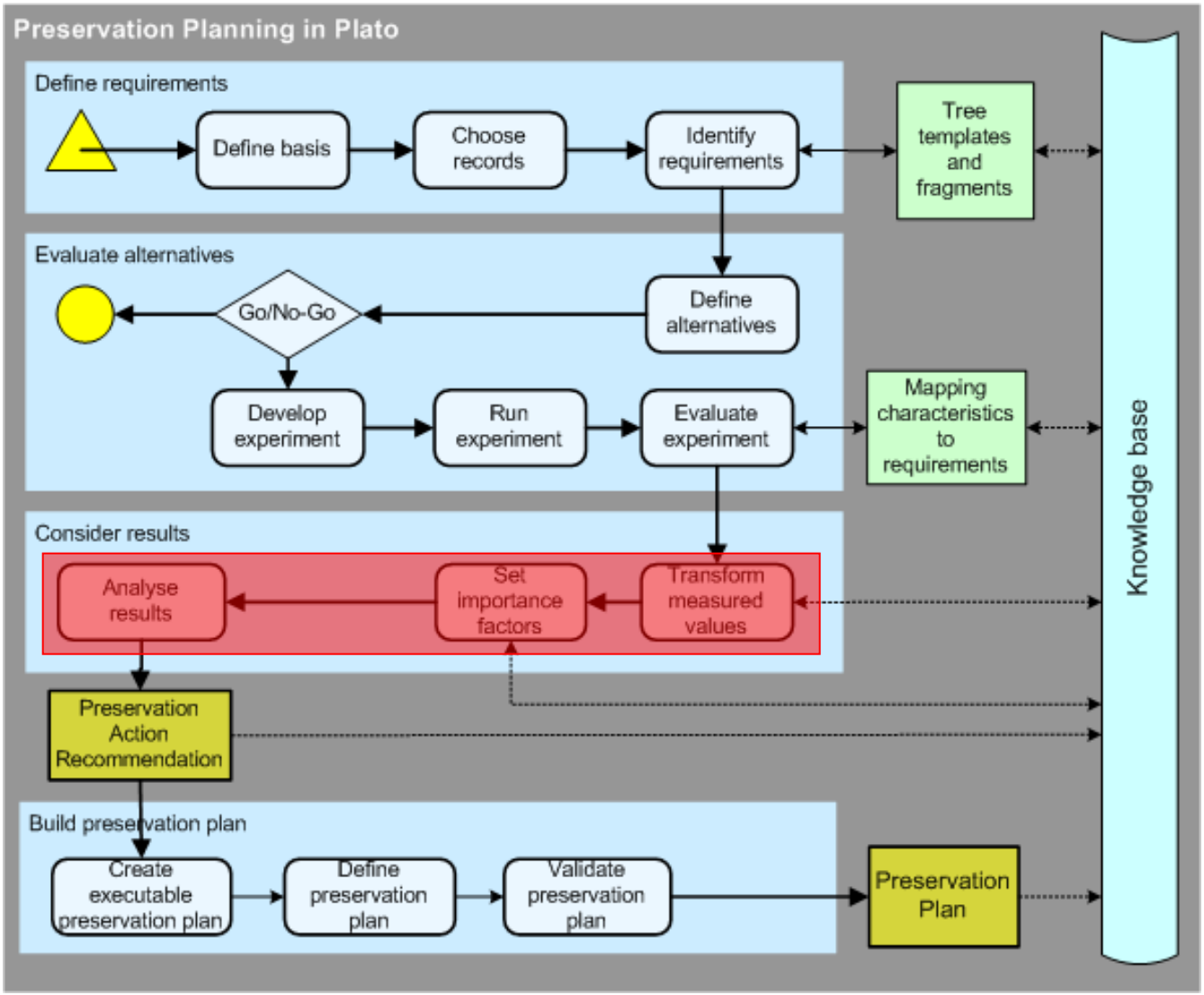
Example: Electronic documents

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

TU WIEN planets UNIVERSITY OF Southampton eprints

PP Workflow



Practise time!

- Log into Plato at: http://www.ifs.tuwien.ac.at/dp/plato
- Open preservation plan named

  “Scanned yearbooks archive (ANALYSE)”

- Proceed to “Validate Preservation Plan”
- Export the preservation plan

Schedule

(1) Introduction
- What is Digital Preservation?
- EPrints
- Preservation Planning and Plato

(2) Preservation in EPrints

(3) Preservation Planning with Plato

(4) Bringing it all together and Closing

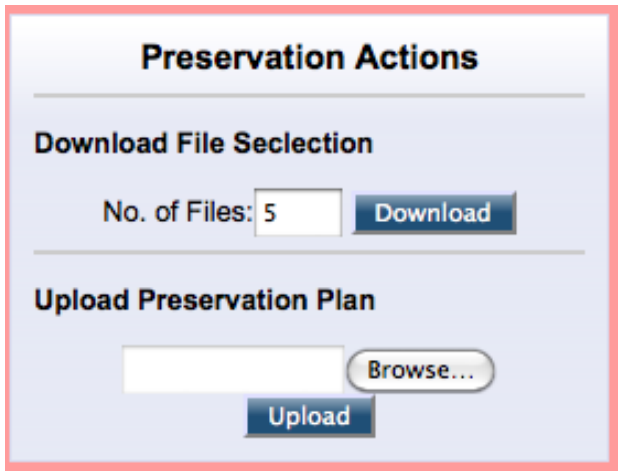
TU WIEN planets UNIVERSITY OF Southampton eprints

The Preservation Process

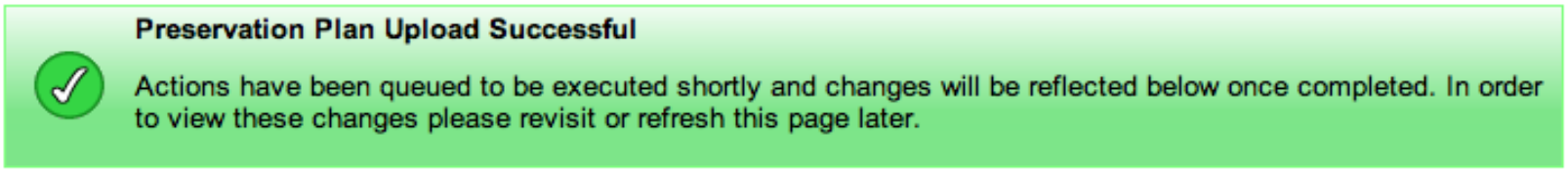
Preservation - Action

- Uploading a Preservation Plan in EPrints
- Viewing resultant actions
- Managing your plans
- Re-enacting the Plan
- Viewing Provenance Information

Uploading a Plan



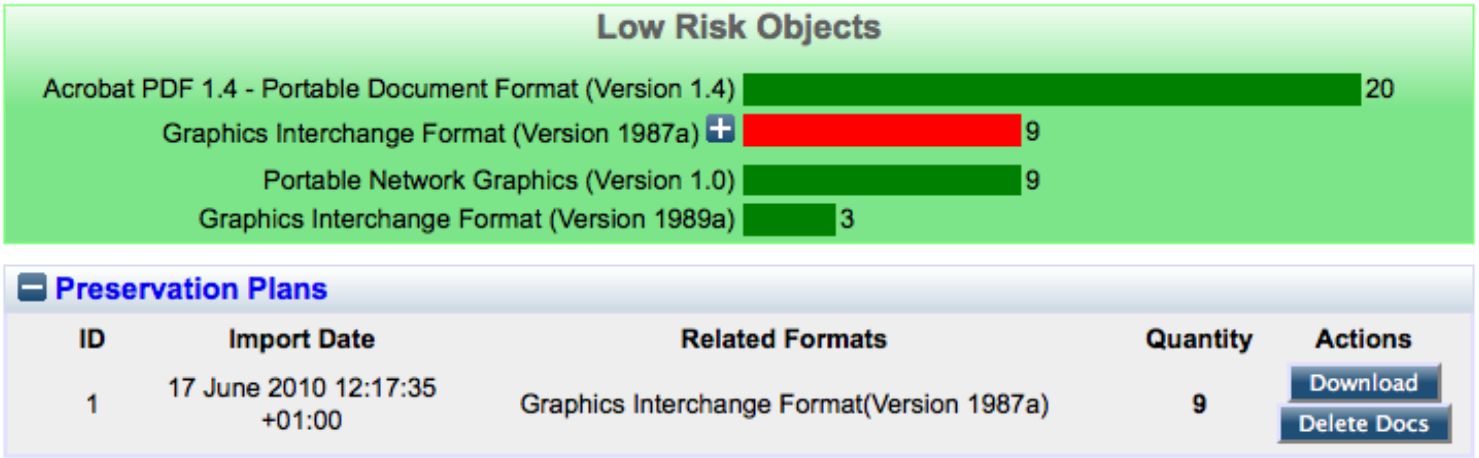
- Each set of “at risk” classified files can have a single related preservation plan.
- Once uploaded, any defined actions will be performed on **all** files of that classification.



Plan Management

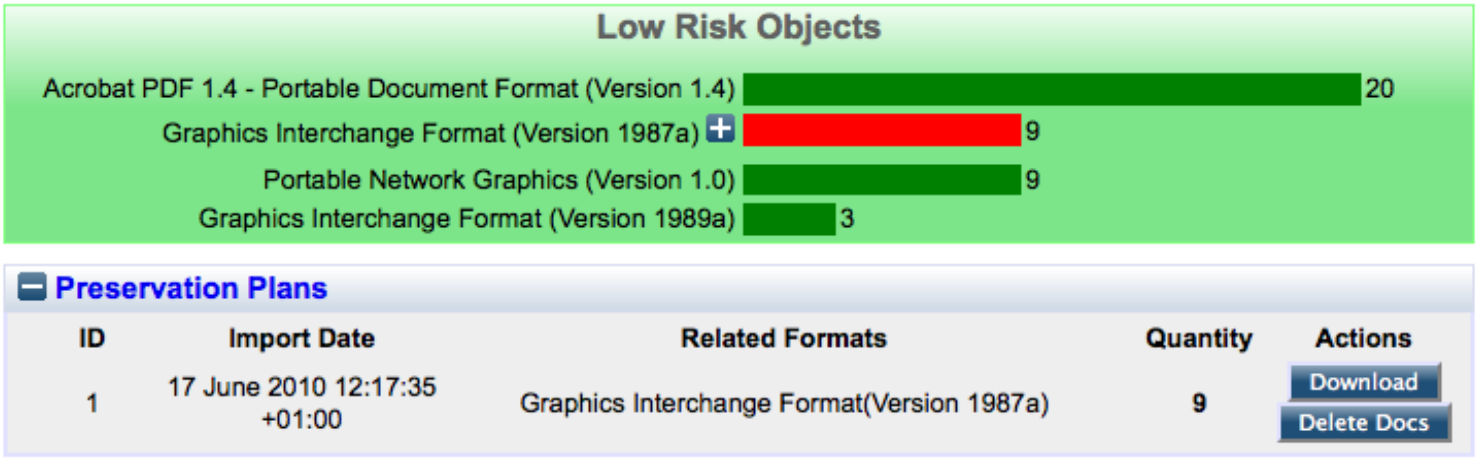
- No plan can cause files to be deleted.

- A plan controls any files it has created.
- While these files exist, the plan cannot be deleted.

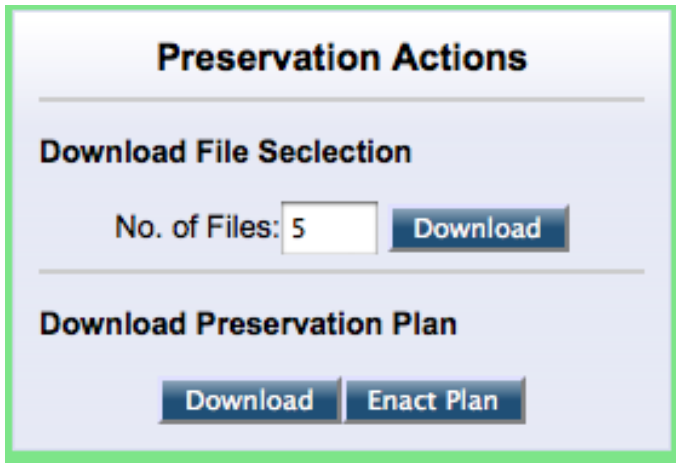


Viewing the Result

- Previously high risk objects are still represented by a red bar, but are now in the low risk category.



Preservation Actions Panel



- Download plan for reviewing in planning software.
- Re-enact plan

Viewing the Result

- Before

Image (GIF)
Download (76Kb) | Preview

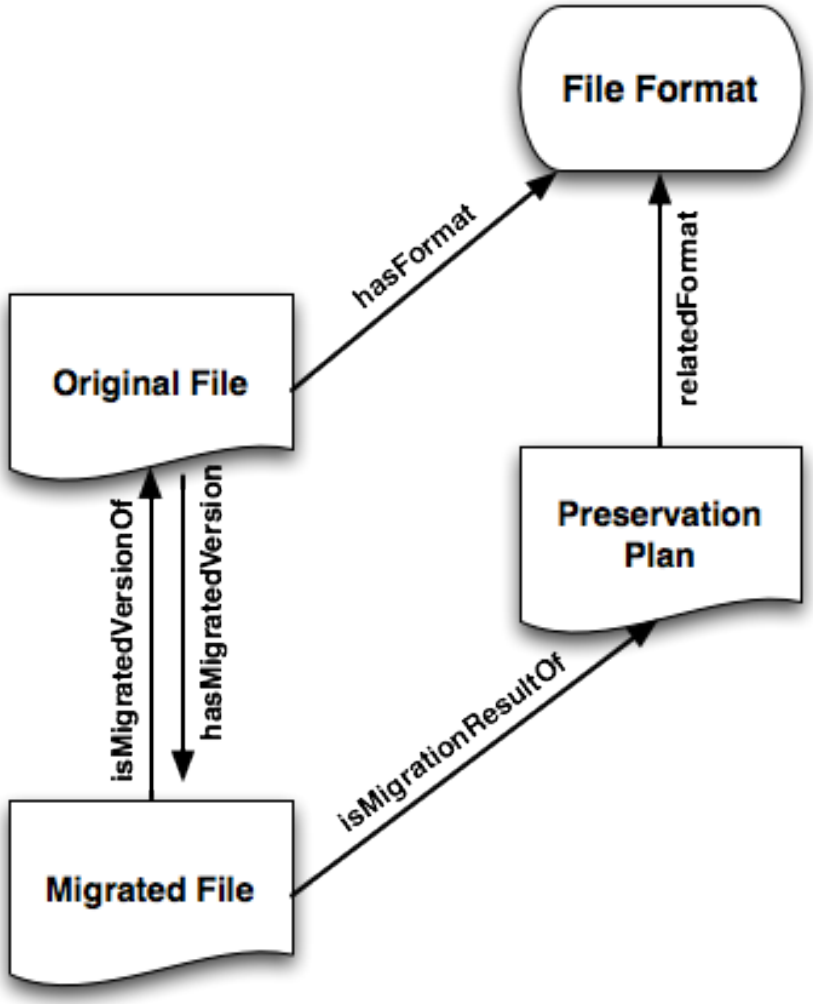
- After

Image (PNG) (Migrated (Preservation) from Document ID: 41 (image/gif))
Download (76Kb) | Preview

- Image (GIF) (Original Version)

Provenance Information

- Open Provenance Model (OPM) compliant
- Stored in RDF triple form using the EPrints relation manager added in 3.2



Exercise Time

Conclusions

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

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 lifecycle

Thank you!

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

http://www.eprints.org/