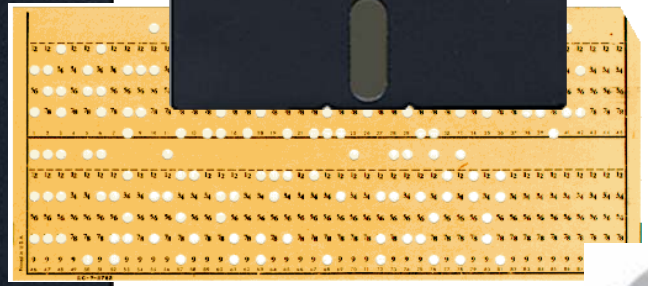
The Preservation Planning Workflow: From institutional requirements via experimental evidence to accountable preservation plans

Andreas Rauber
Department of Software Technology and Interactive Systems
Vienna University of Technology
http://www.ifs.tuwien.ac.at/~andi

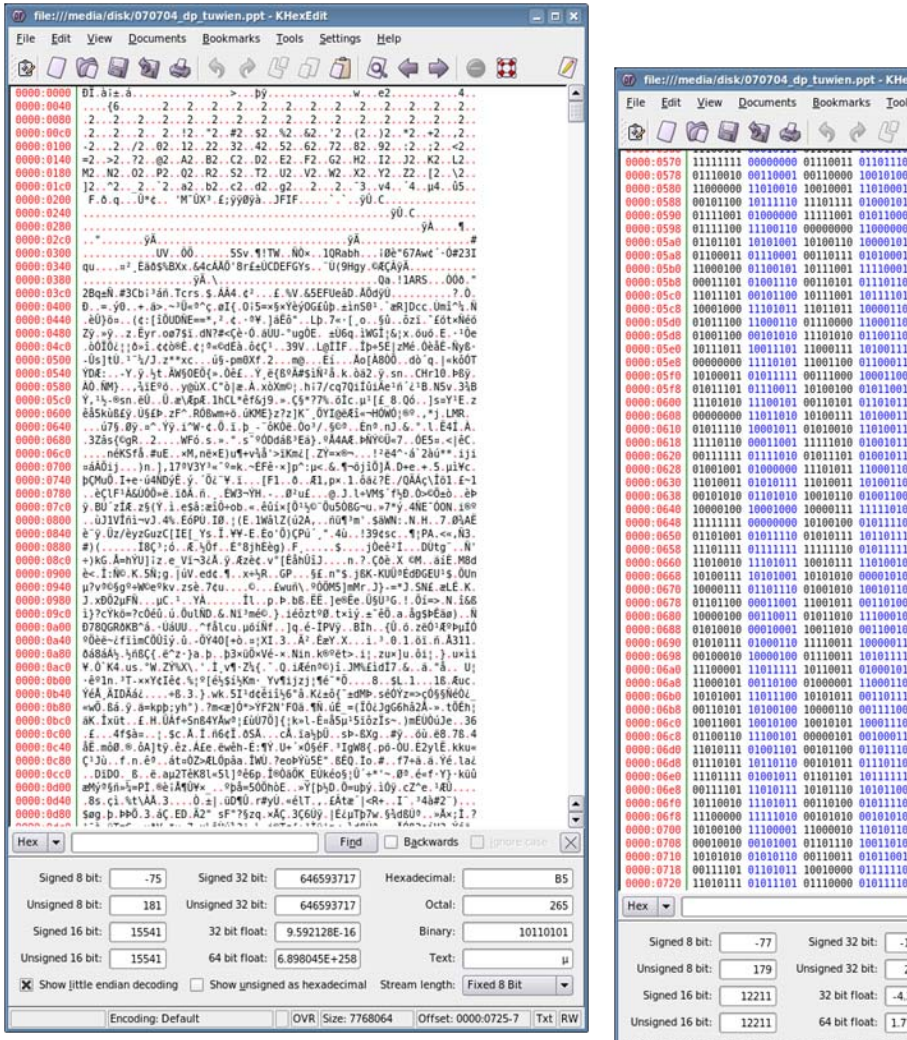
Overview

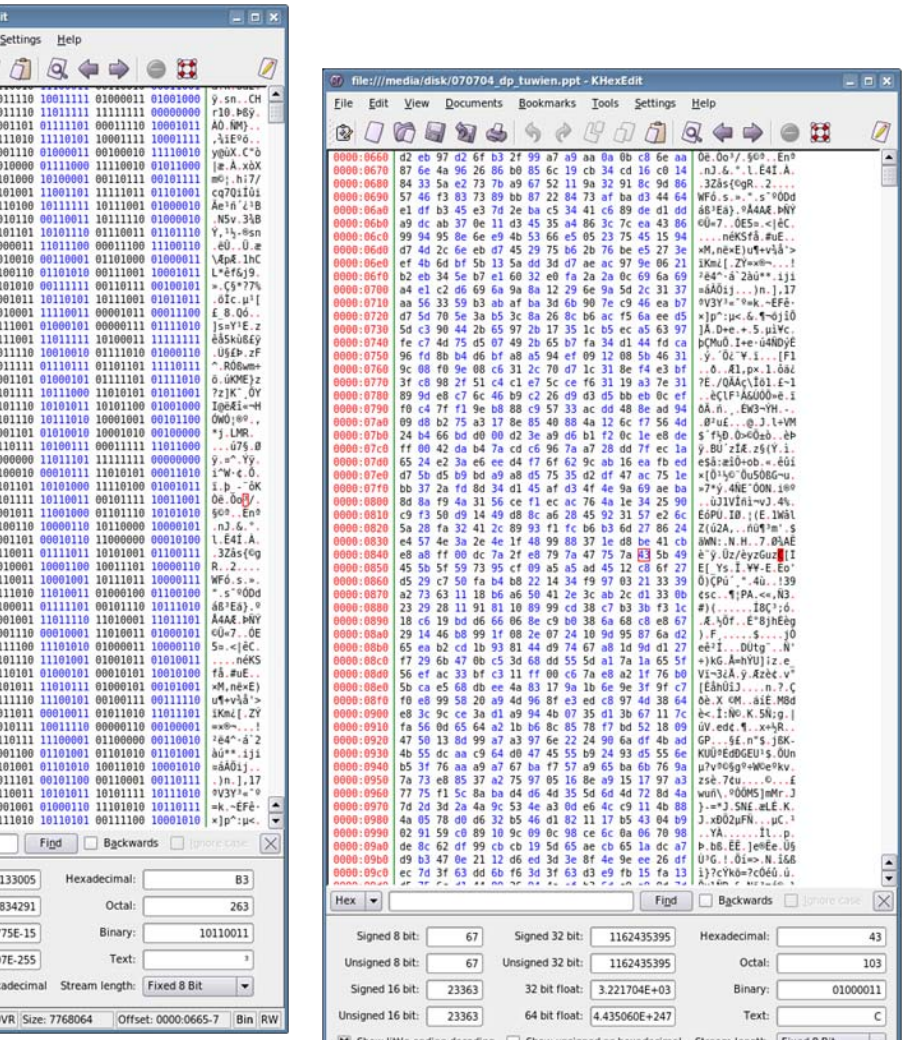
- What is Digital Preservation?
- Preservation Planning
- Plato: Preservation Planning Tool

Why do we need Digital Preservation?



Why do we need Digital Preservation?





Why do we need Digital Preservation?

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

Strategies for Digital Preservation

Strategies

(grouped according to Companion Document to UNESCO Charter http://unesdoc.unesco.org/images/0013/001300/130071e.pdf)

- Investment strategies:
  - Standardization, Data extraction, Encapsulation, Format limitations
- Short-term approaches:
  - Museum, Backwards-compatibility, Version-migration, Reengineering
- Medium- / long-term approaches:
  - Migration, Viewer, Emulation
- Alternative approaches:
  - Non-digital Approaches, Data-Archeology
- No single optimal solution for all objects

Migration

- Transformation into different format, continuous or on-demand (Viewer)
+ Wide-spread adoption
+ Possibility to compare to un-migrated object
+ Immediately accessible
- Unintended changes, specifically over sequence of migrations
- Cannot be used for all objects
- Requires continuous action to migrate

Emulation

- Emulation of hardware or software (operating system, applications)
+ Concept of emulation widely used
+ Numerous emulators are available
+ Potentially complete preservation of functionality
+ *Object is rendered identically*
- *Object is rendered identically*
- Requires detailed documentation of system
- Requires knowledge on how to operate current systems in the future
- Complex technology
- Emulators must be emulated or migrated themselves
- Emulators potentially erroneous/incomplete

DP & Publishing

What are some core DP challenges in scientific publishing wrt. the objects to be preserved?

- The publication
- The context of the publication
- Adjunct material (slides, notes, videos, …)
- Demos, exercises, interactive elements
- Data sets and simulations
- Community aspects, web 2.0, discussion forums, …
- Links to external material
- Other new functionalities

Digital Preservation

- Is a complex task
- Requires a concise understanding of the objects, their intellectual characteristics, the way they were created and used and how they will most likely be used in the future
- Requires a continuous commitment to preserve objects to avoid the „digital dark ages“
- Requires a solid, trusted infrastructure and workflows to ensure digital objects are not lost
- Is essential to maintain electronic publications & data accessible
- Will become more complex as digital objects become more complex

Overview

- What is Digital Preservation?
- Preservation Planning
- Plato: Preservation Planning Tool

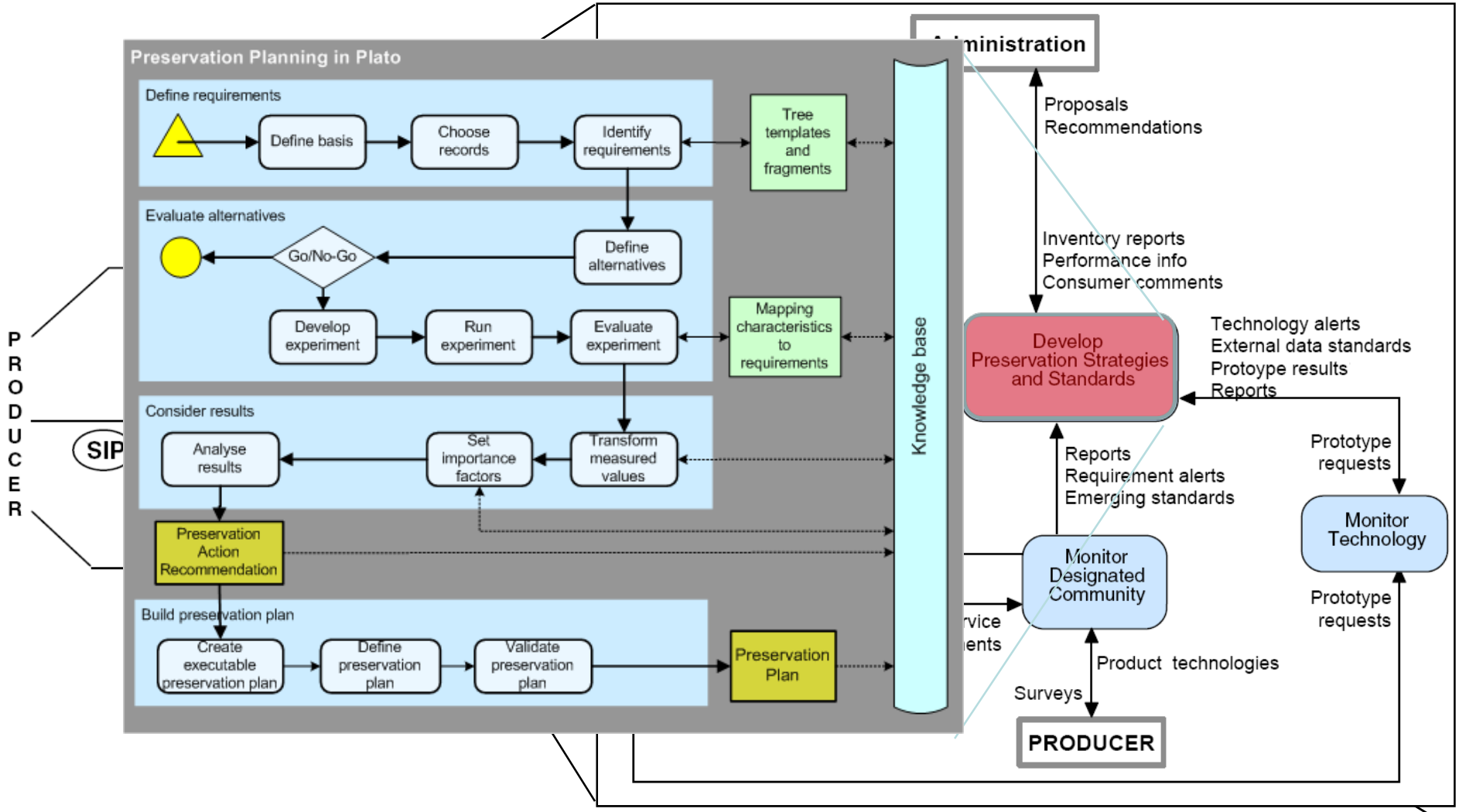
Why Preservation Planning?

- 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?

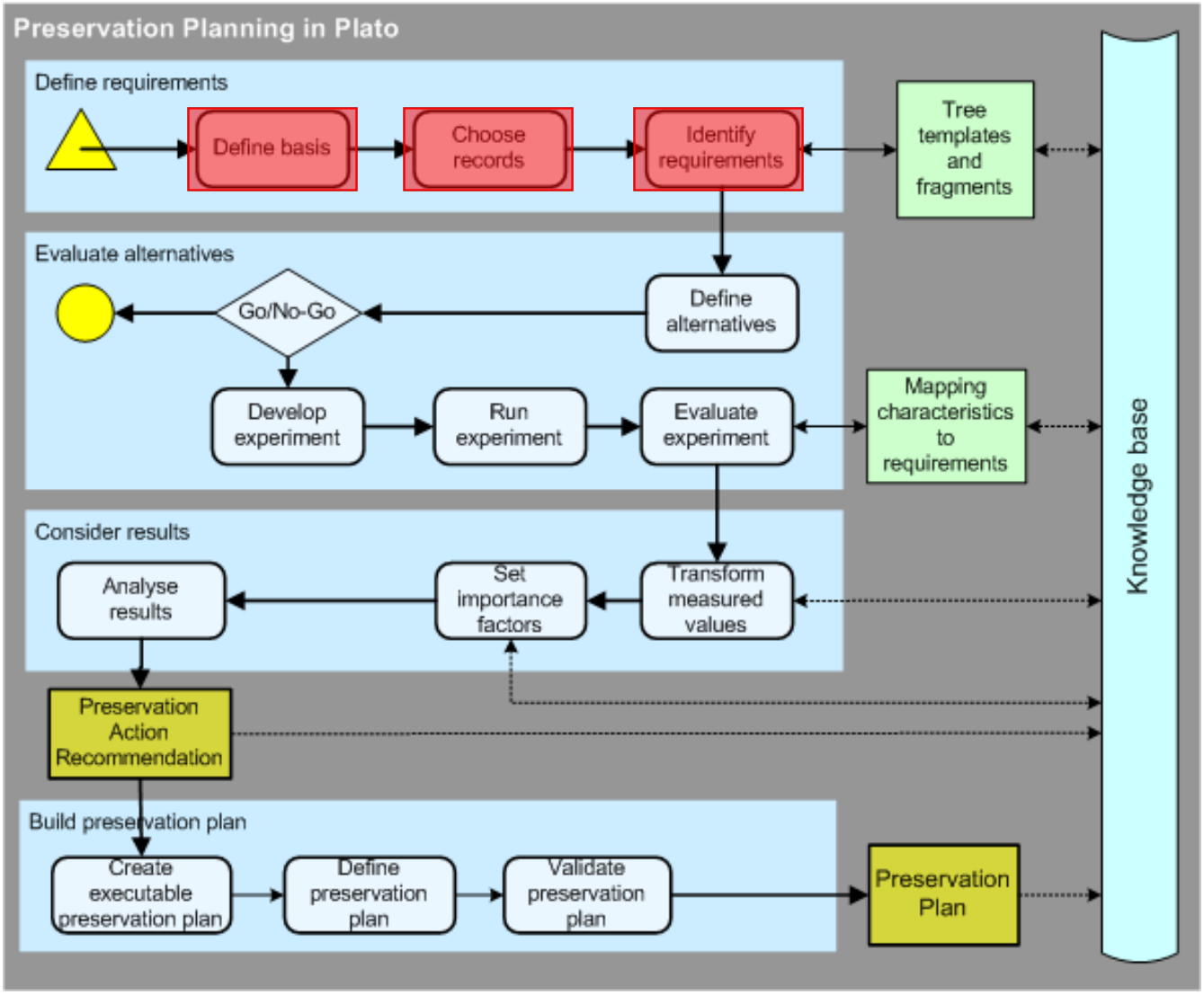
Preservation Planning Workflow

- Originally developed within the DELOS DP Cluster now refined and integrated within PLANETS
- Based on
  - Preservation Planning approach using utility analysis, developed at TU Vienna
  - Testbed for evaluation developed at Nationalarchief, The Netherlands
- Follows the OAIS model
- Consistent with requirements specified by ORLC/TRAC and Nestor criteria catalogue

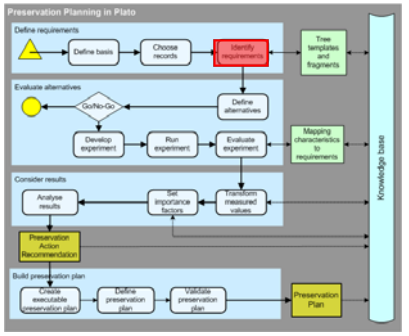
Preservation Planning

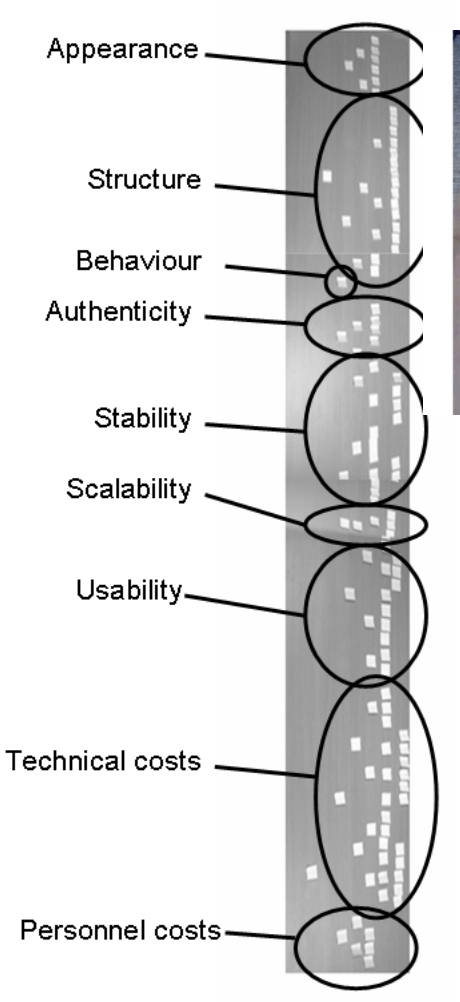


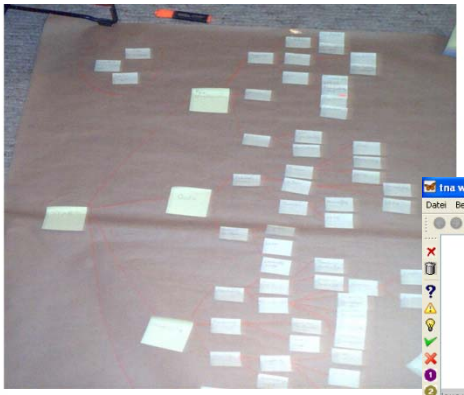
Preservation Planning Workflow



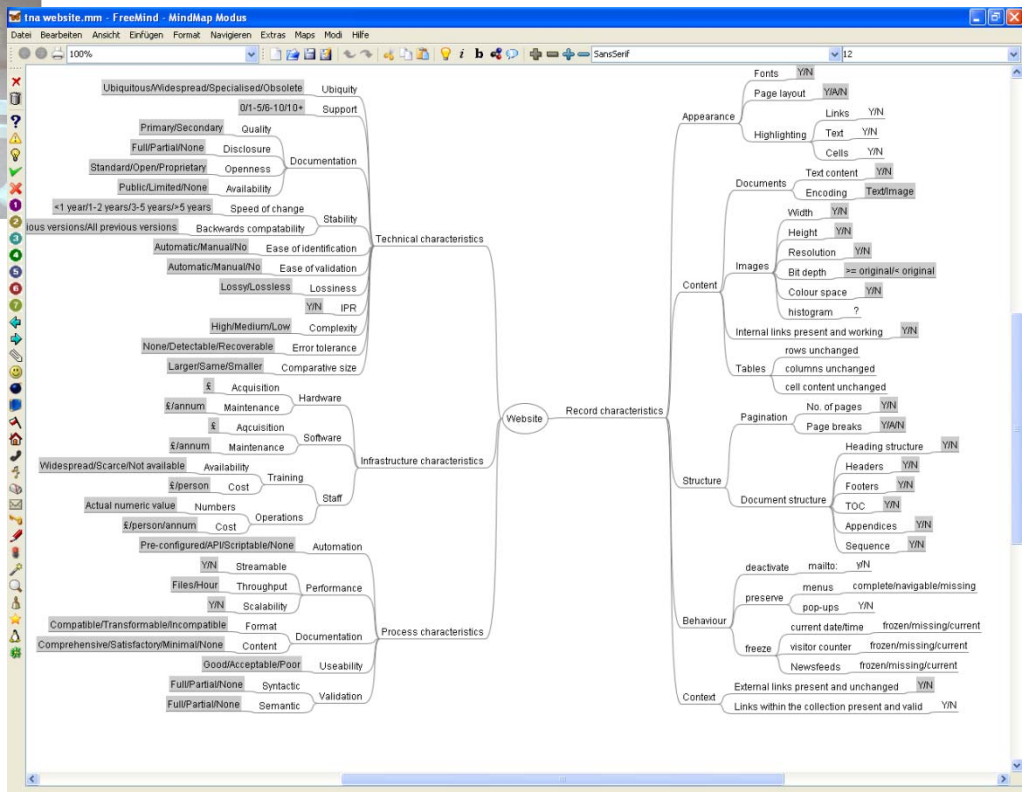
Identify requirements





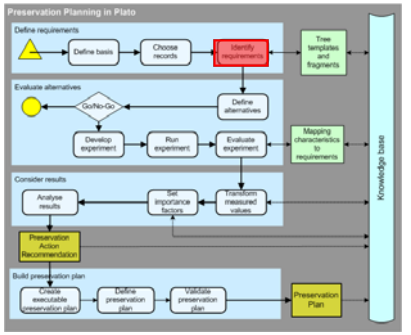


Analog…



… or born digital

Identify requirements



OENB Master thesis.mm - FreeMind - MindMap Modus

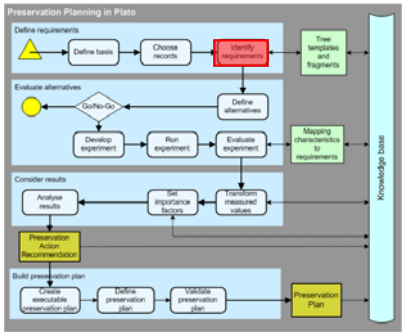
Datei Bearbeiten Ansicht Einfügen Format Navigieren Extras Maps Modi Hilfe

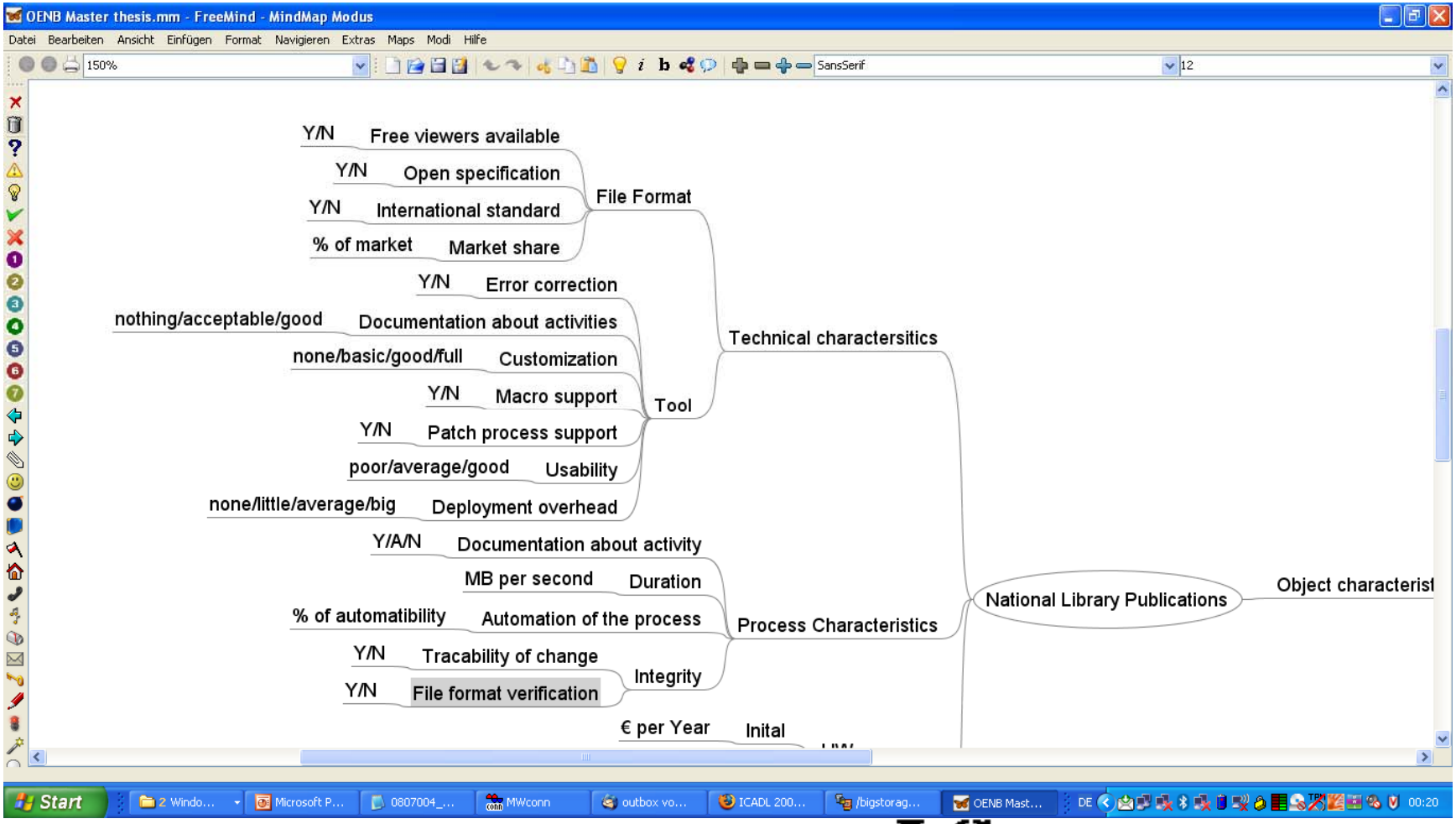
150% SansSerif 12

- osts
- Content
  - Video
    - Resolution: higher/unchanged/lower
    - Free viewers available: Y/N
    - Open specificaiton: Y/N
    - International standard: Y/N
    - Market share: % of market
  - Audio
    - Bitrate: higher/unchanged/lower
    - Free viewers available: Y/N
    - Open specificaiton: Y/N
    - International standard: Y/N
    - Market share: % of market
- Behaviour
  - Script blocking: Y/N
  - Deactivation of security meachnism: Y/N
  - URLs clickable: Y/N
  - Jump References: Y/N
  - Input Forms working: Y/N
  - Video playable: Internal/External/No
  - Audio playable: Internal/External/No
  - Content machine readable: Y/N

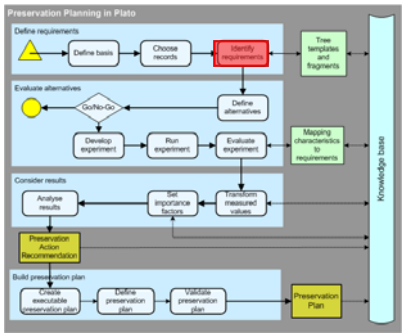
Start slides Microsoft Pow... 0807004_guta... MWconn outbox von ra... ICADL 2008 - ... OENB Master t... DE 23:14

Identify requirements

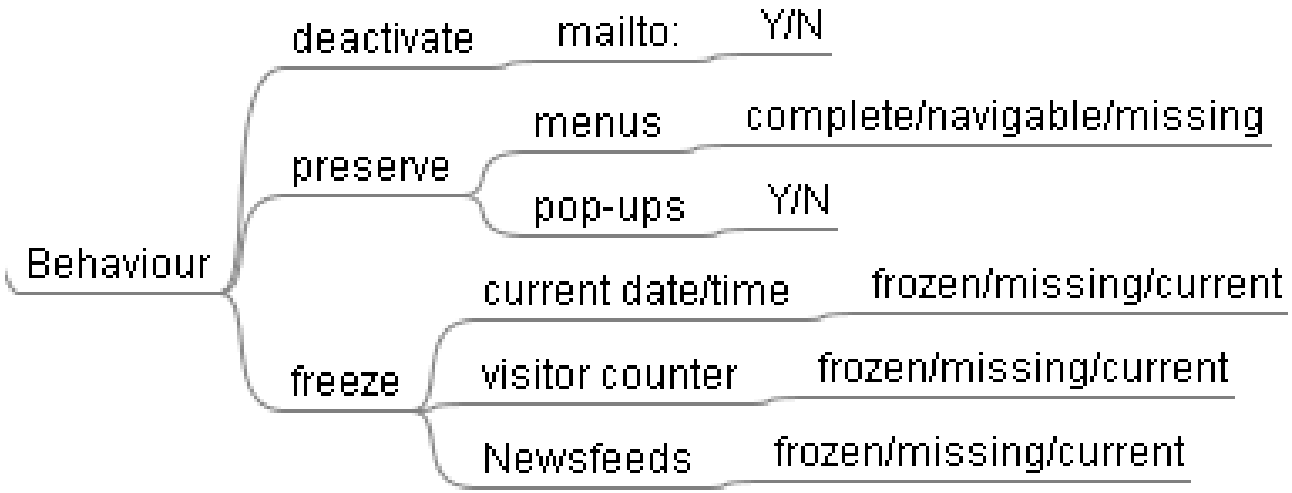




Identify requirements

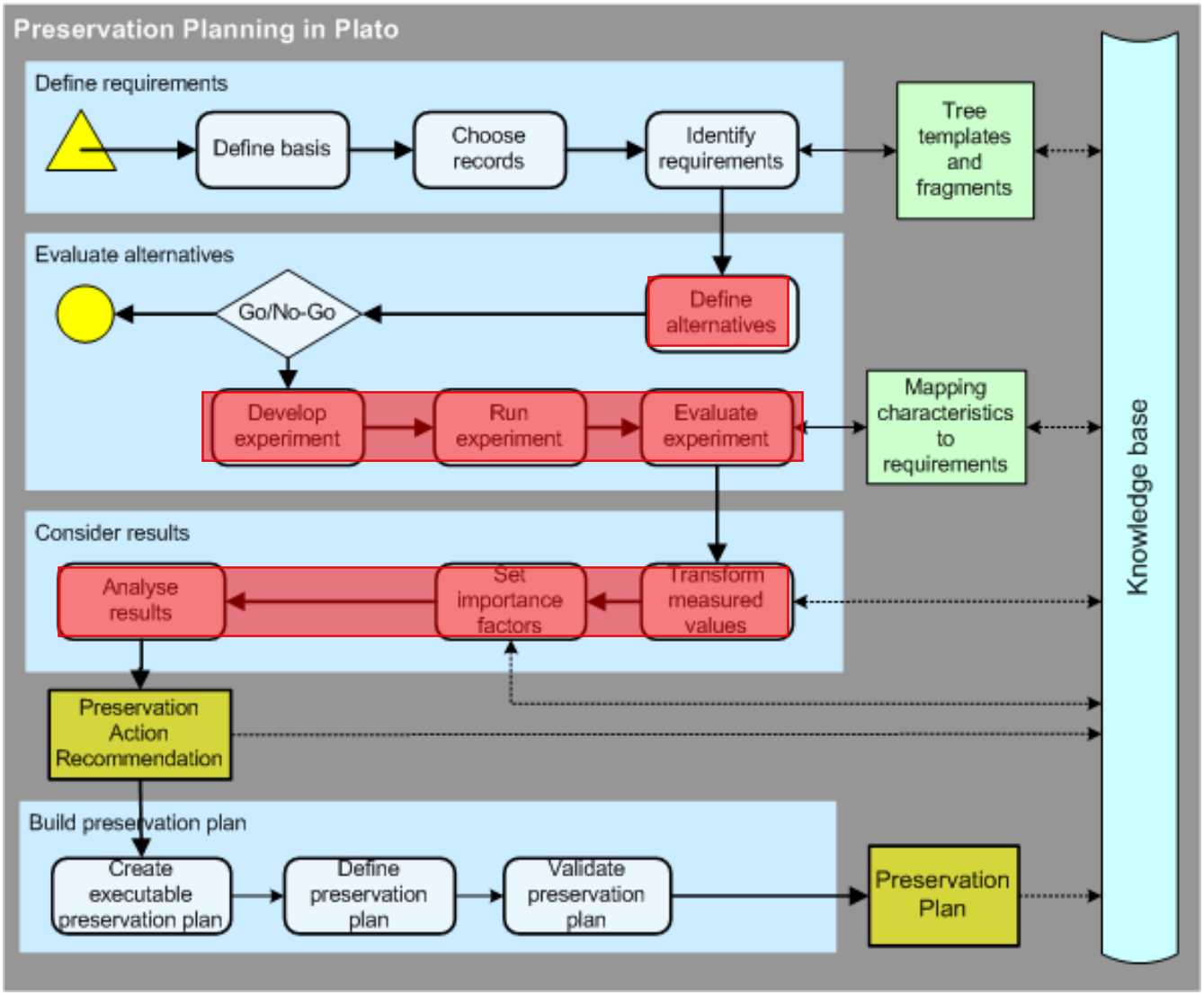


Behaviour

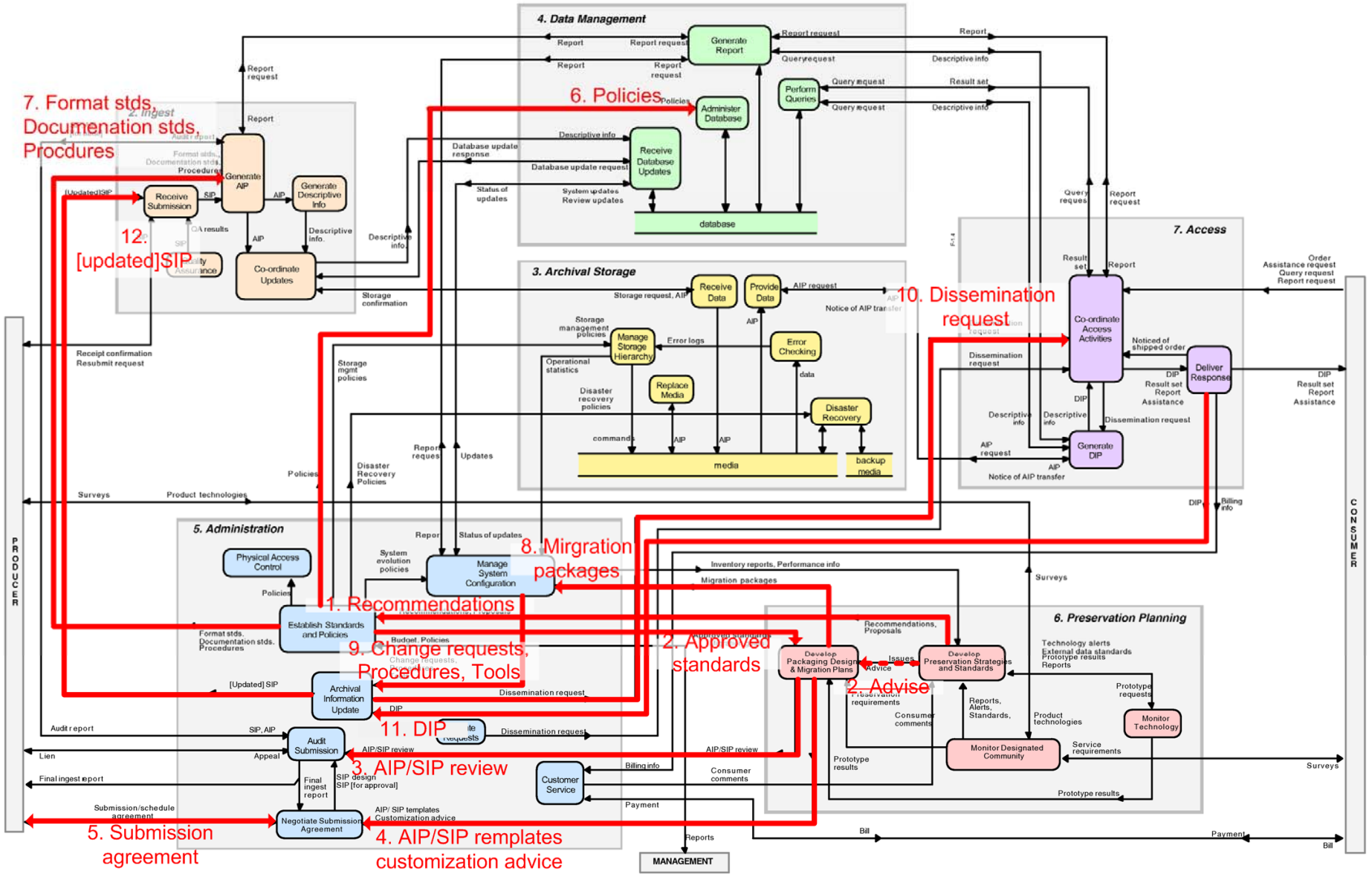


- Visitor counter and similar functionalities can be
  - Frozen at harvesting time
  - Omitted
  - Remain operational, i.e. the counter will be increased upon archival calls
    (is this desired? count? demonstrate functionality?)

Preservation Planning Workflow



Preservation Planning & OAIS Model



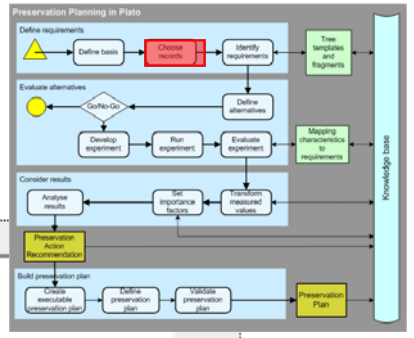
Overview

- What is Digital Preservation?
- Preservation Planning
- Plato: Preservation Planning Tool

Plato

- Preservation Planning Tool
- Reference implementation of planning workflow
- Web-based application, release 2.0 Nov. 12 2008
- Documents the process and ensures that all steps are considered
- Automates several steps, integrates services
- Creates a preservation plan (XML, PDF)
- Technical basis:
  - Java Enterprise Beans, EJB 3 (Hibernate)
  - Based on JBoss Application Server
  - XML Import/Export (XStream)
- http://www.ifs.tuwien.ac.at/dp/plato

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

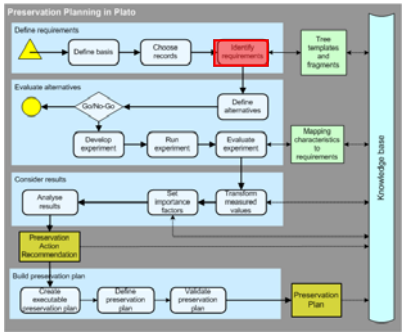
Add record

Upload new record

Durchsuchen... | Upload File

- Upload objects
- Identify using DROID & Pronom

Plato



- Create Objective Tree: Mindmaps &Tree-Editor

PLANETS Preservation Planning Tool *(Plato)*

[logout becker] [help]

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

Identify Requirements

Objective Tree
Descriptive Information

How can I define the objective tree?

[↑] Objective Tree

Expand All | Collapse All

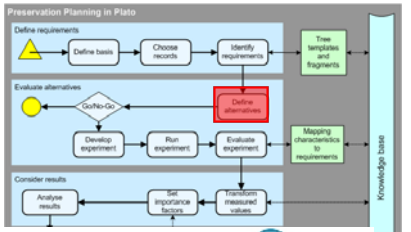
Website

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

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

Quick Access:

Plato



PLANETS Preservation Planning Tool *(Plato)*

[logout kulovits] [help]

Project | Define Requirements ::: Evaluate Requirements ::: Consider Results | Polar bear image archive

Define the alternatives of the Project

ID	Name	Description	Remove
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

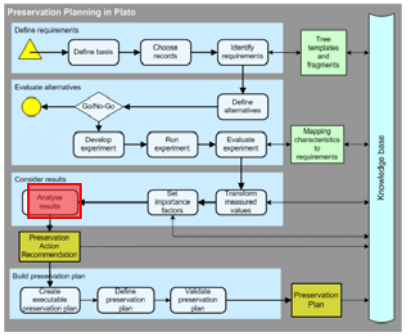
Show Preservation Services

	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

- Select preservation actions
- Planets registry, Crib, Minimi
- Migration & Emulation services

Plato



PLANETS Preservation Planning Tool - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://localhost:8080/plato/workflow/analyseresults.seam

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

- [x] PDF/A (Tool A)
- [x] PDF/A (Tool B)

Show

Expand All | Collapse All

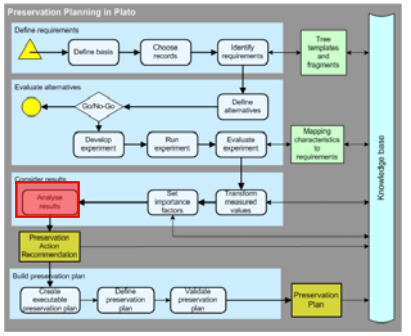
Minimalist root node

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

Fertig

- Automatic evaluation of properties
- Analyze & compare performance



Case study: electronic theses

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

Preservation Planning

What we have after this presentation

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

Conclusions

- Preservation Planning to ensure "optimal" preservation
- A simple, methodologically sound model to specify and document requirements
- Repeatable and documented evaluation
- Basis for well-informed, accountable decisions
- Concretization of OAIS model
- Follows recommendations of TRAC and nestor
- Generic workflow that can easily be integrated in different institutional settings
- **Plato:**
  - Tool support to perform solid, well-documented analyses
  - Creates core preservation plan
  - 241 users who have created 135 preservation planning projects

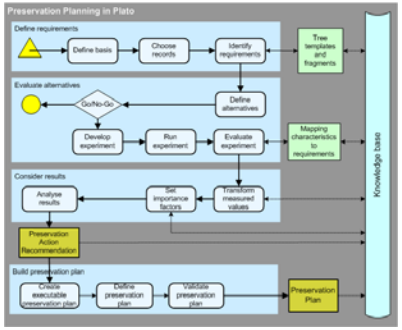
http://www.ifs.tuwien.ac.at/dp

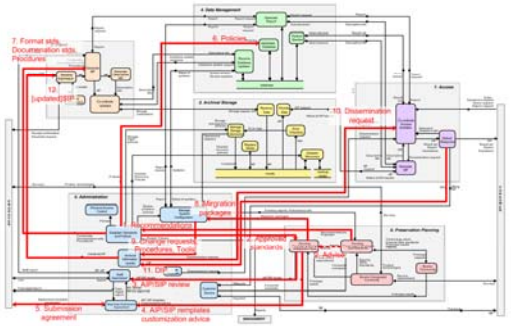
http://www.ifs.tuwien.ac.at/dp/plato

Tutorials / Training Events

- ECDL 2009, Sep. 27 – Oct. 2, Corfu, Greece
- Planets training, Sep. 21-23, Sofia, Bulgaria
- Planets training, June 22-24, Copenhagen, DK
- JCDL 2009, June 15-19 , Austin/TX
- IST Africa 2009, May 6-8, Uganda
- DigCCurr 2009, April 1-3, Chapel Hill, USA
- wePreserve Training, March 23-27, Barcelona
- Nestor Spring School 2009, March 16-20, Stauffen, DE
- CeBIT 2009, Mar. 3-7, Hannover, Germany
- DPC repository day, Dec. 12, London
- ICADL 2008, Dec. 2-5, Bali, Indonesia
- RCDL 2008, Oct. 7-11, Dubna, Russia
- ECDL 2008, Sep. 14-19, Aarhus, Denmark
- DELOS Summer School 2008, June 8-11,Pisa, Italy
- Sun PASIG 2008, May 27-29, San Francisco, USA

Thank you!





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

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