



# Digital Preservation

**Ao.univ.Prof. Andreas Rauber**

Institut für Softwaretechnik und Interaktive Systeme

TU Wien

<http://www.ifs.tuwien.ac.at/~andi>

.....



FACULTY OF **INFORMATICS**

# Overview

- 
- What is Digital Preservation?
  - A glimpse at the OAIS model
  - Preservation Planning
  - Plato: Preservation Planning Tool
-

# 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.

# Why do we need Digital Preservation?

- Essential for all digital objects
  - Office documents, accounting, emails, ...
  - Scientific datasets, sensor data, metadata, databases, ...
  - Applications, simulations,...
  
- All application domains
  - Cultural heritage data
  - eGovernment, public administration
  - Science / Research
  - Industry
  - Health, pharmaceutical industry
  - Aviation, control systems, construction, ...
  - Private data
  - ...

## 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

## 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

# 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 hole“
- 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
- Needs to be defined in a preservation plan

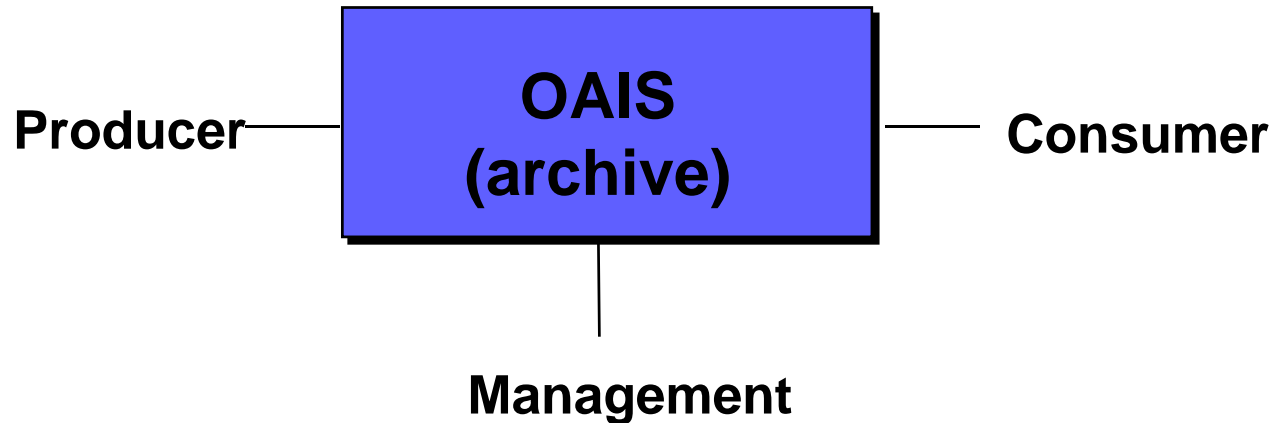
# Overview

- 
- What is Digital Preservation?
  - A glimpse at the OAIS model
  - Preservation Planning
  - Plato: Preservation Planning Tool
-

# OAIS

- NASA: National Space Science Data Center
  - NASA's first digital archive
  - Experienced many technological changes since 1966
- Consultative Committee for Space Data Systems
  - International group of space agencies
  - Developed range of discipline-independent standards
  - Evolved into ISO TC 20/ SC 13 working group around 1990
- Reference Model for an Open Archival Information System (OAIS), Blue Book, CCSDS 650.0-B-1, January 2002
- ISO 14721:2003
- <http://ssdoo.gsfc.nasa.gov/nost/isoas/overview.html>

# OAIS



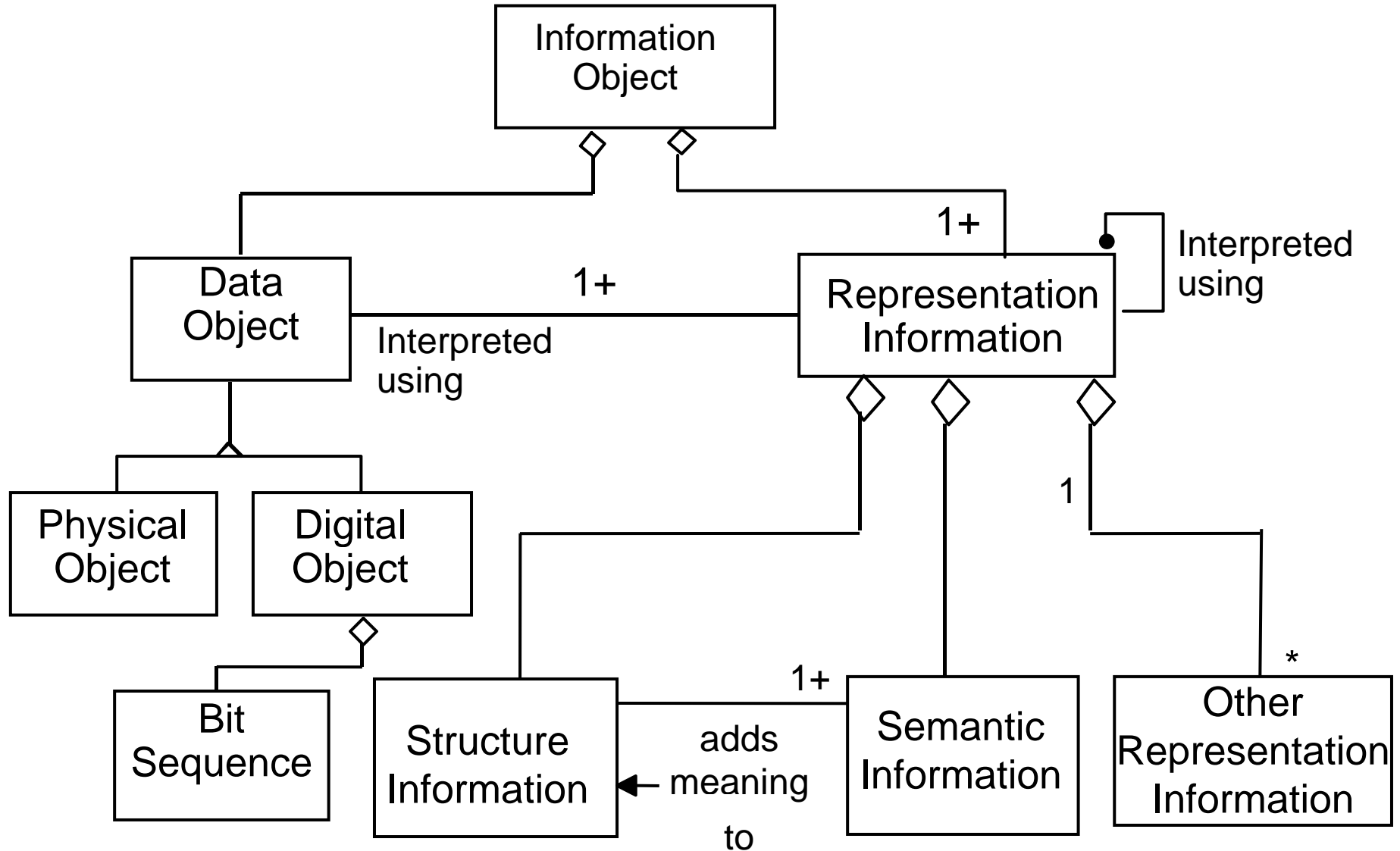
- Producer is the role played by those persons, or client systems, who provide the information to be preserved
- Management is the role played by those who set overall OAIS policy as one component in a broader policy domain
- Consumer is the role played by those persons, or client systems, who interact with OAIS services to find and acquire preserved information of interest

## OAIS Information Definition

- Information is always expressed (i.e., represented) by some type of data
- Data interpreted using its Representation Information yields Information
- Information Object preservation requires clear identification and understanding of the Data Object and its associated Representation Information



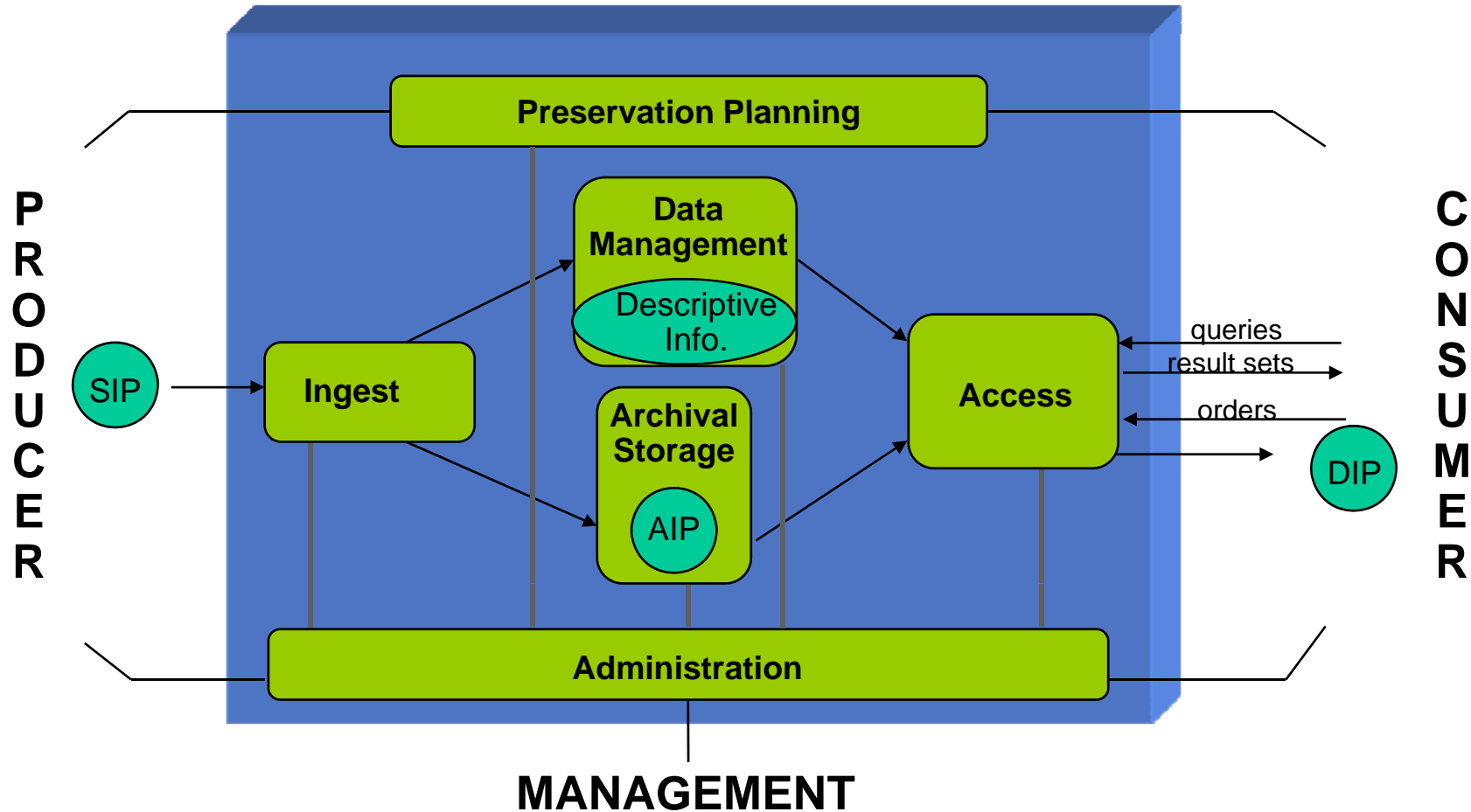
# OAIS



## Information Package Variants

- **SIP:** Submission Information Package
  - Negotiated between Producer and OAIS
  - Sent to OAIS by a Producer
- **AIP:** Archival Information Package
  - Information Package used for preservation
  - Includes complete set of Preservation Description Information (PDI) for the Content Information
- **DIP:** Dissemination Information Package
  - Includes part or all of one or more Archival Information Packages
  - Sent to a Consumer by the OAIS

# OAIS

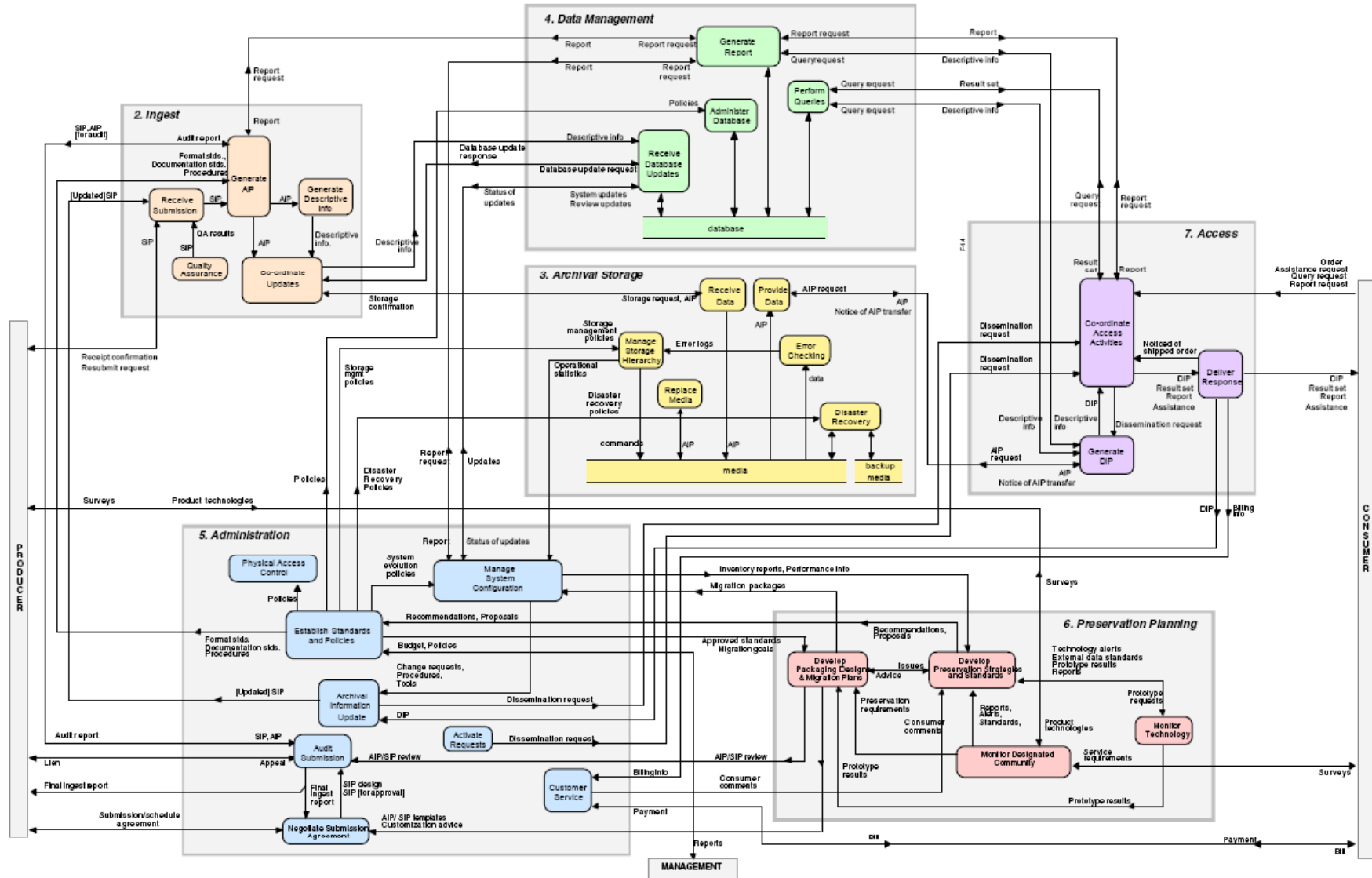


SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package

# OAIS



# Overview

- 
- What is Digital Preservation?
  - A glimpse at the OAIS model
  - 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?

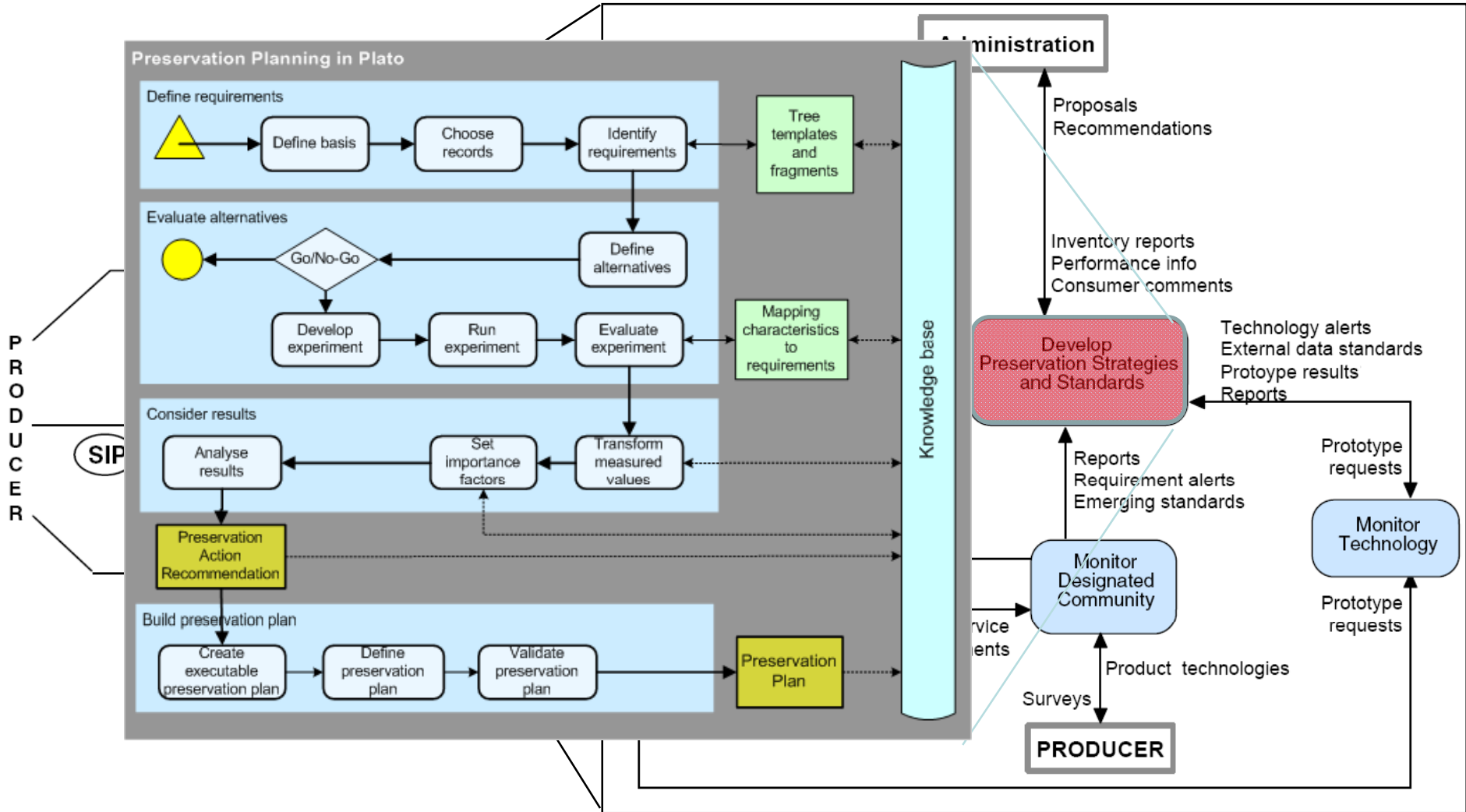
# Preservation Planning

## Preservation Planning Workflow

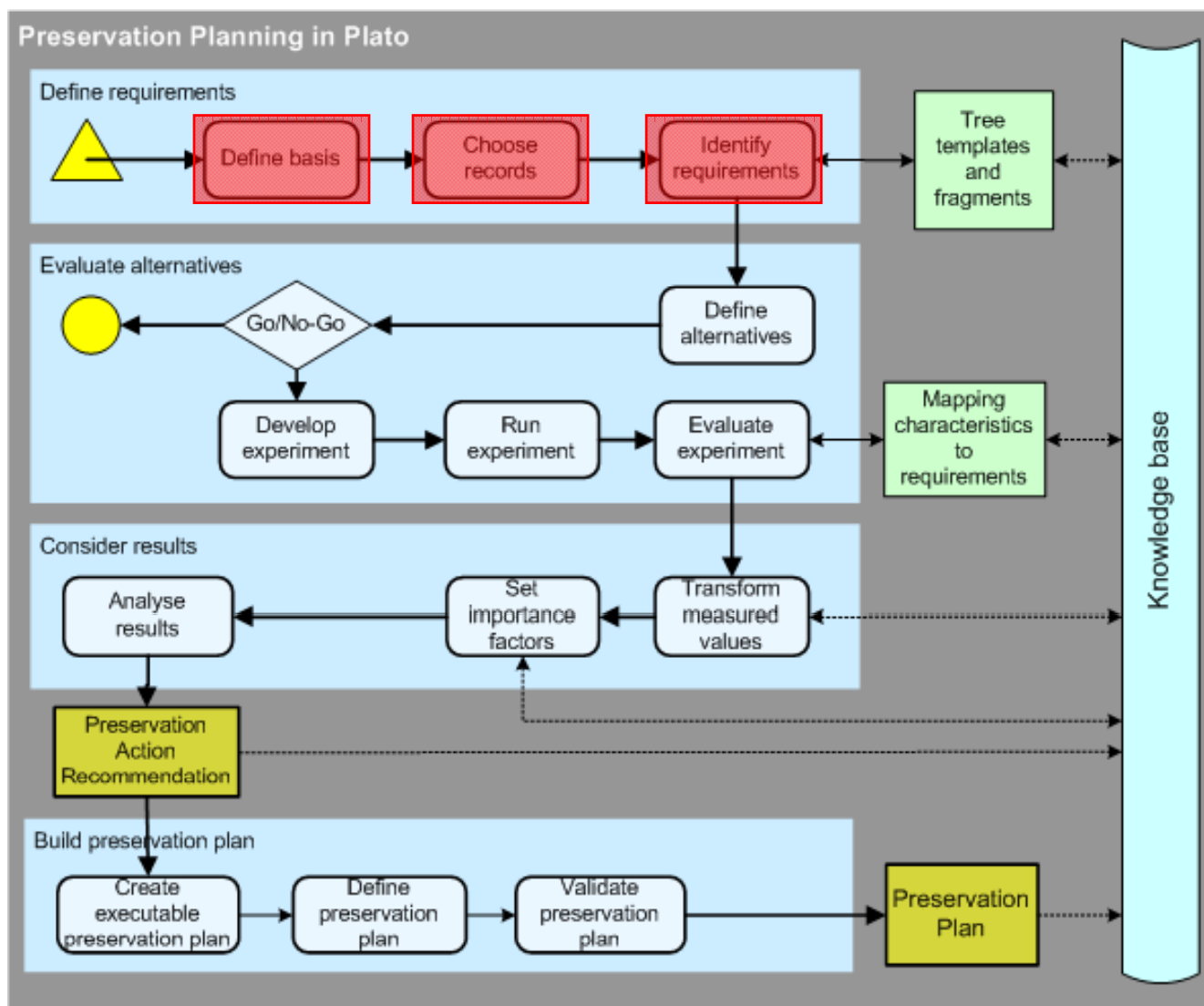
- Originally developed within the DELOS DP Cluster now refined and integrated within PLANETS
- Based on
  - Preservation Planning approach using Utility Analysis, developed at Vienna University of Technology
  - Testbed for evaluation developed at Nationalarchief, The Netherlands
- Follows the OAIS model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

.....

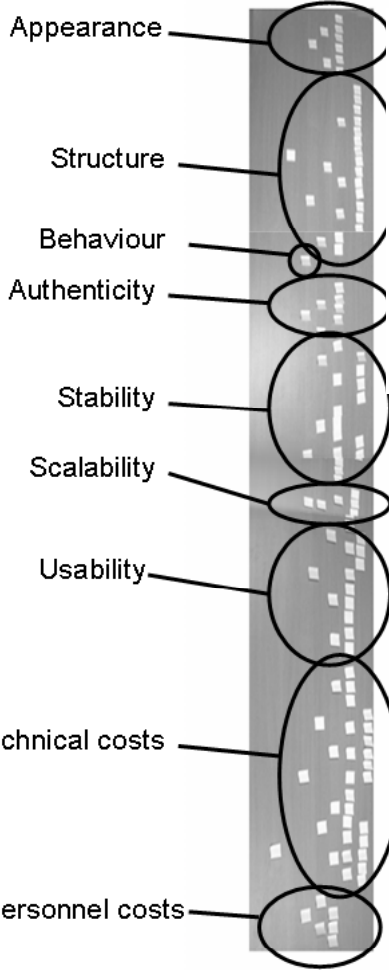
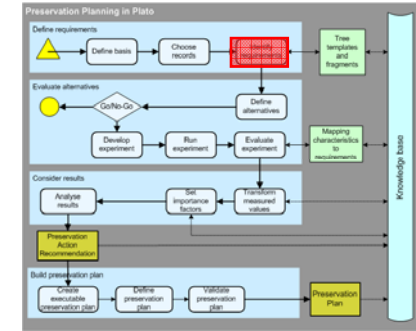
# Preservation Planning



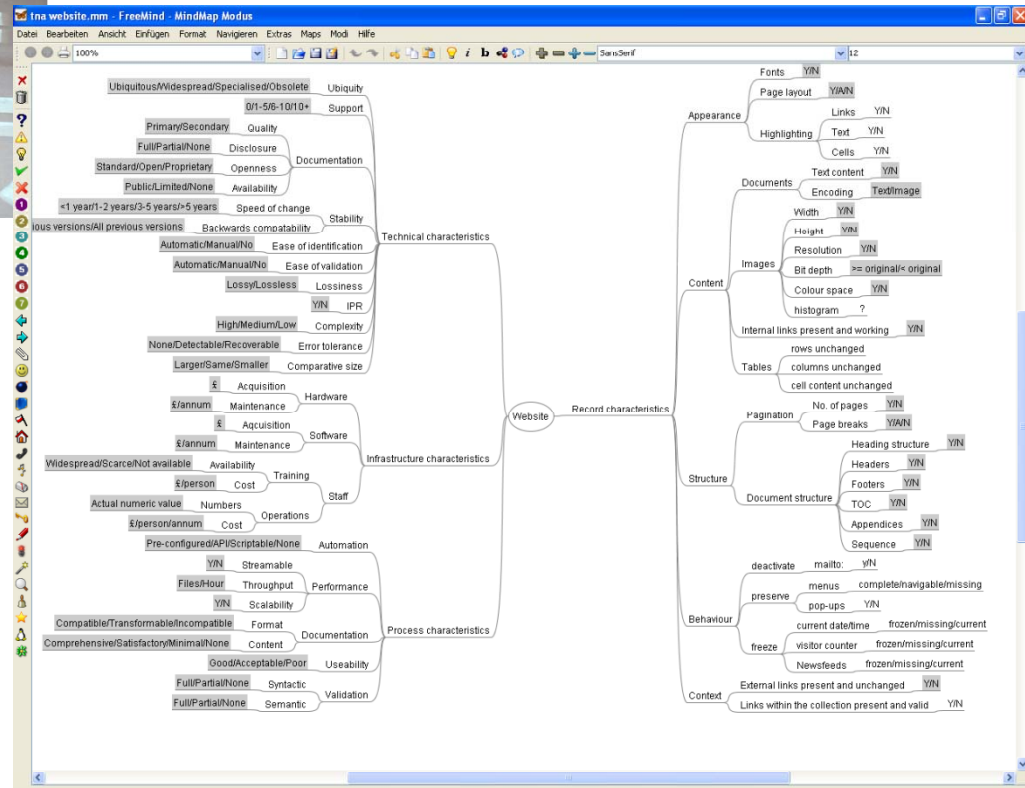
# Preservation Planning Workflow



# Identify requirements

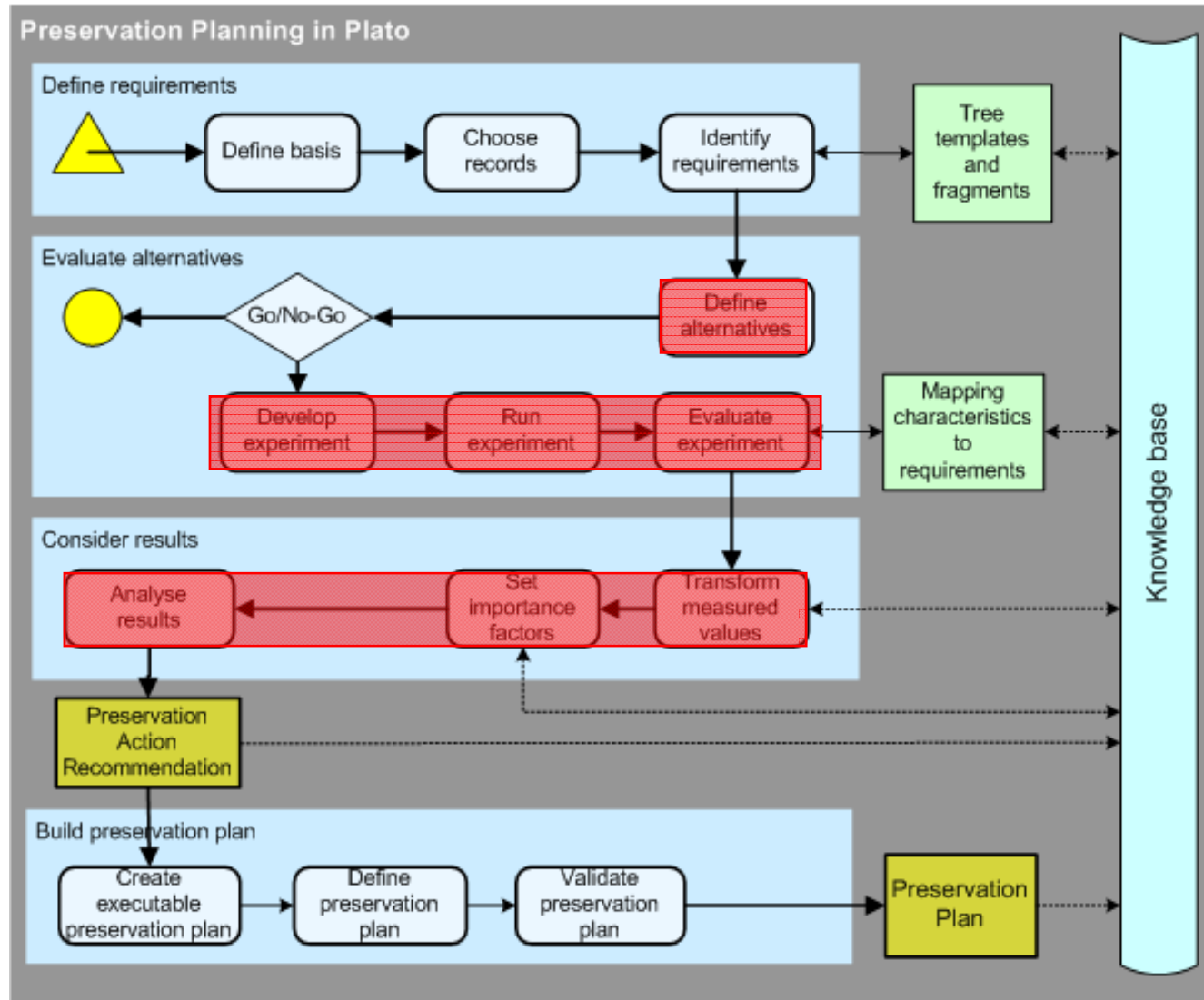


Analog...



... or  
born  
digital

# Preservation Planning Workflow



# Overview

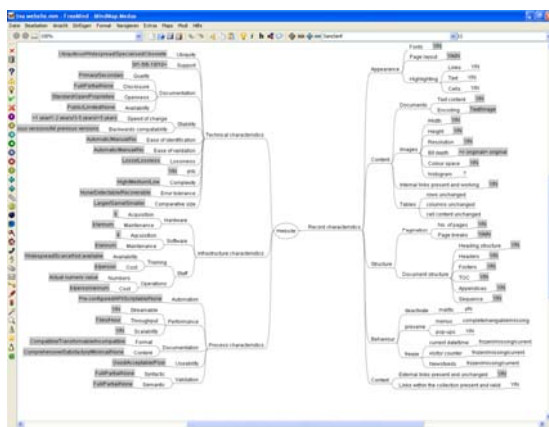
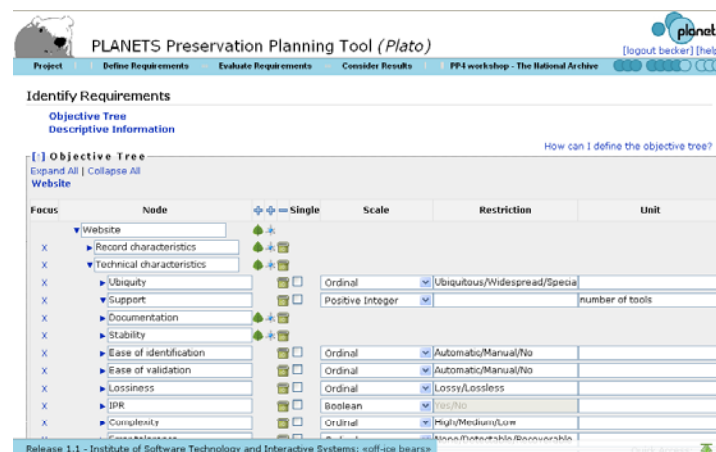
- 
- What is Digital Preservation?
  - A glimpse at the OAIS model
  - Preservation Planning
  - Plato: Preservation Planning Tool
-

## Plato

- Preservation Planning Tool
- Reference implementation of planning workflow
- Web-based application, release 2.0 Nov. 12 2009
- Documents the process and ensures that all steps are considered
- Automates several steps
- Creates a preservation plan (XML, PDF)
- Technical basis:
  - Java Enterprise Beans, EJB 3 (Hibernate)
  - Based on JBoss Application Server
  - JBoss Seam Integration Framework
  - Java Server Faces with Facelets
  - XML Import/Export (XStream)

## Plato

- Assists in analyzing the collection
  - Profiling, analysis of sample objects via Pronom and other services
- Allows creation of objective tree
  - Within application or via import of mindmaps
- Allows the selection of Preservation action tools

Identify Requirements

Objective Tree

Descriptive Information

How can I define the objective tree?

[-] Objective Tree

Expand All | Collapse All

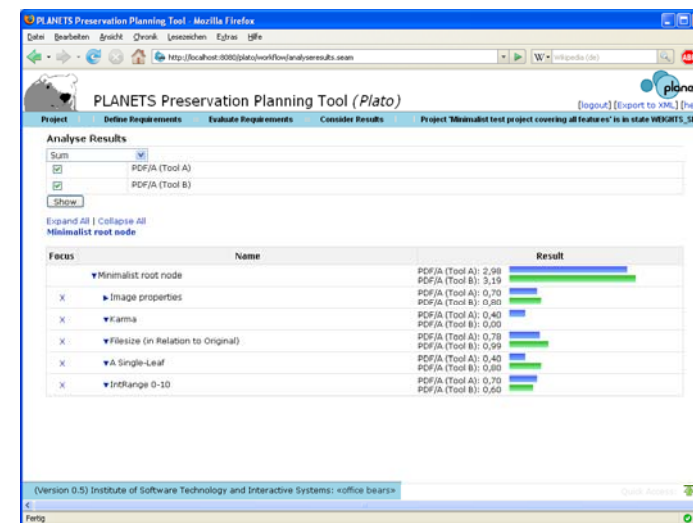
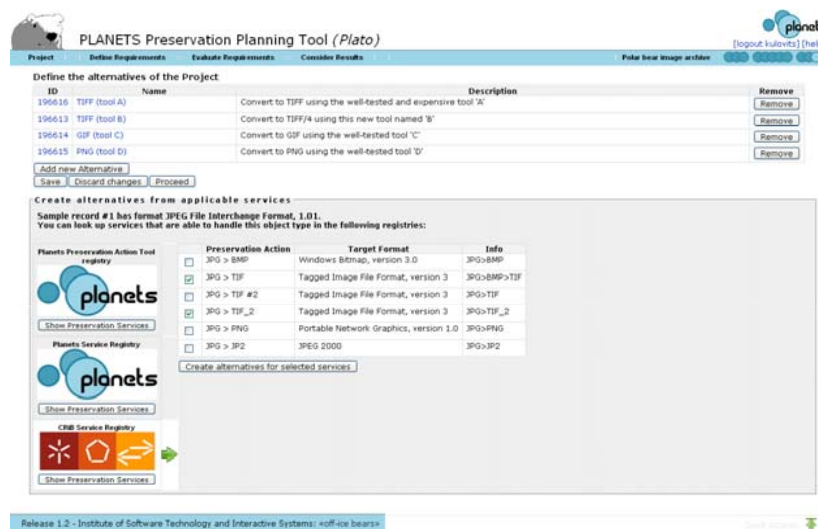
Website

Focus	Node	Single	Scale	Restriction	Unit
X	Website				
X	Record characteristics				
X	Technical characteristics				
X	Ubiquity	<input type="checkbox"/>	Ordinal	Ubiquitous/Widespread/Special	
X	Support	<input type="checkbox"/>	Positive Integer		number of tools
X	Documentation				
X	Stability				
X	Ease of identification	<input type="checkbox"/>	Ordinal	Automatic/Manual/No	
X	Ease of validation	<input type="checkbox"/>	Ordinal	Automatic/Manual/No	
X	Lossiness	<input type="checkbox"/>	Ordinal	Lossy/Lossless	
X	IPR	<input type="checkbox"/>	Boolean	Yes/No	
X	Complexity	<input type="checkbox"/>	Ordinal	Highly/Medium/Low	

Release 1.1 - Institute of Software Technology and Interactive Systems; off-ice bears

## Plato

- Runs experiments and documents results
- Allows definition of transformation rules, weightings
- Performs evaluation, sensitivity analysis,
- Provides recommendation (ranks solutions)



## What Preservation Planning produces:

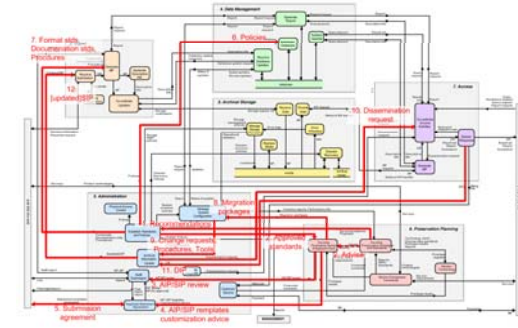
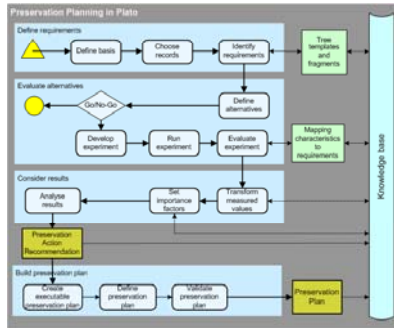
- Basic Preservation Plan:
  - PDF: [Preservation Plan.pdf](#)
  - XML: [Preservation Plan.xml](#)
- That was developed in a solid, repeatable and documented process
- That is optimal for the needs of a given institution and for the data at hand

## 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

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

# Thank you!



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

