

Preservation Planning Sub-Project

Hans Hofman

Nationaal Archief Netherlands

16th ICA Congress, Kuala Lumpur, 23 July 2008



nationaal archief



Overview

- What is the problem? Defining the issues
- The scope and role of Preservation Planning
- The Planets approach
 - The organisational/ business context
 - Usage requirements and collection profiles
 - The planning process and preservation plan



The issue / challenge

- The enormous and rapidly increasing amount of digital information
 - Fragile resources
- The rapid evolution in technology
- The risk of obsolescence and therefore corruption and/or loss of valuable information
- **(Pro-)active** and ongoing attention / maintenance required
- Potential solutions: still fragmented
 - infrastructure
 - not comprehensive

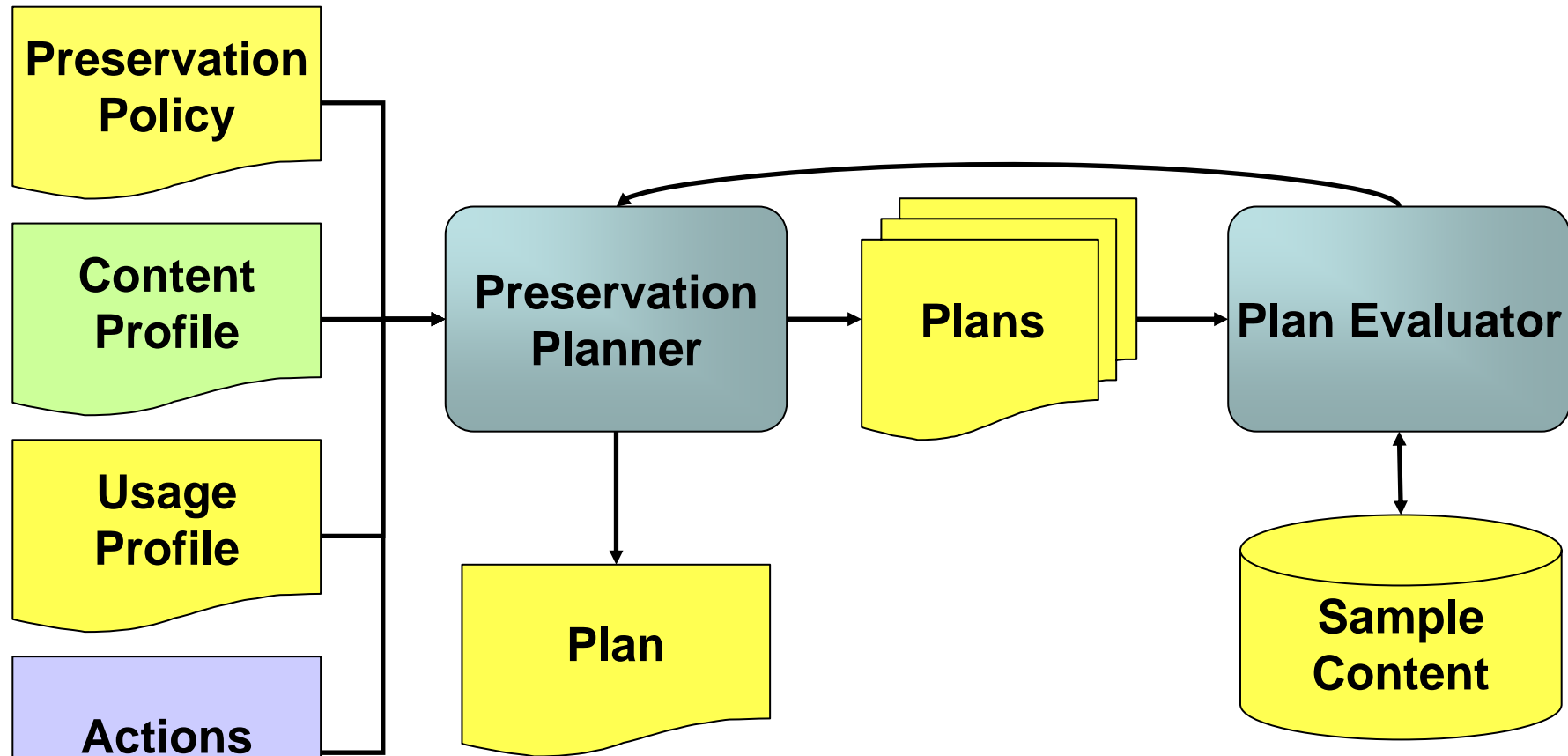


Stakeholders

- Memory institutions ('content holders')
 - archives, libraries
- (Scientific) data centres
- Government organisations (record creators)
- Business companies (record creators, intellectual capital)
- Individuals



Preservation planning



Objectives of Preservation Planning

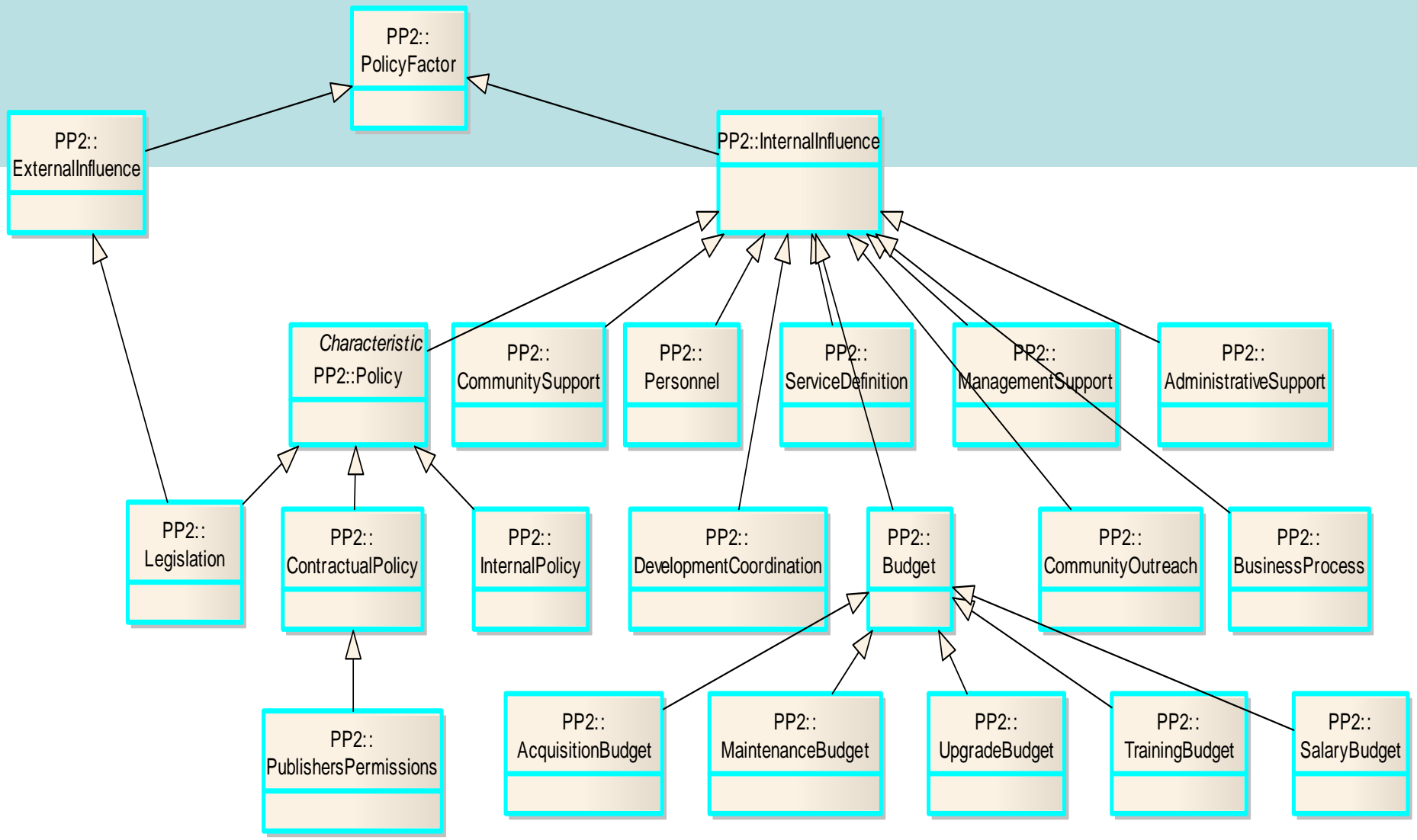
- Identify and analyse the organisational context
 - including a risk assessment
 - define a framework for preservation / policy
- Support decision-making about digital preservation including
 - Identifying criteria for preservation within that context
 - Defining workflow for evaluating/ defining preservation plans
 - Developing methodologies for assessing the risks of applying different preservation strategies for different types of digital objects
- Enable formulation, evaluation and execution of high-quality and cost-effective preservation plans that suit the organisational (e.g. repository) needs
- Support the on-going evaluation of the results of executing preservation plans and provide a feedback mechanism
- Document the planning process carefully



Model for organisational requirements (policy model)

- Approach:
 - Literature study and interviews with decision makers
 - Extraction of requirements and building of conceptual model
 - Create model from first principles
- First version of model
 - describes and positions policy requirements in wider context of Planets data model
 - indicates potential policy requirements on different levels and in relation to various types of preservation objects (e.g. collection, deliverable unit, manifestation, byte stream, etc.)
 - (thus) is able to cover policy requirements that are relevant for all types of preservation actions in different organisational settings
- Translate organisational constraints and requirements into a machine interpretable model
 - an organisation can choose from them according to its needs





Usage model

- Preliminary model with user requirements has been created on the basis of results of probe approach
- Model shows some variances between different types of users
- Usage requirements
 - Performance, Usability, Presentation, Authenticity
 - Understandability, Rights, Costs
- These requirements are in first instance relevant for presentation files, but will have an impact on preservation files
- The PP/3 user requirements model will be used as input into PLATO
- Article on methodology published in D-Lib Magazine article, May/June 2008.



Collection profile

- What types of objects (both technical and intellectual aspects)?
- Technical: file formats
 - registries (e.g. PRONOM, ...)
- Intellectual: for instance documentary form, structure, look and feel, ‘behaviour’
 - objective tree ‘templates’
 - an (intellectual) object may consist of different computer files
 - what strategy then?

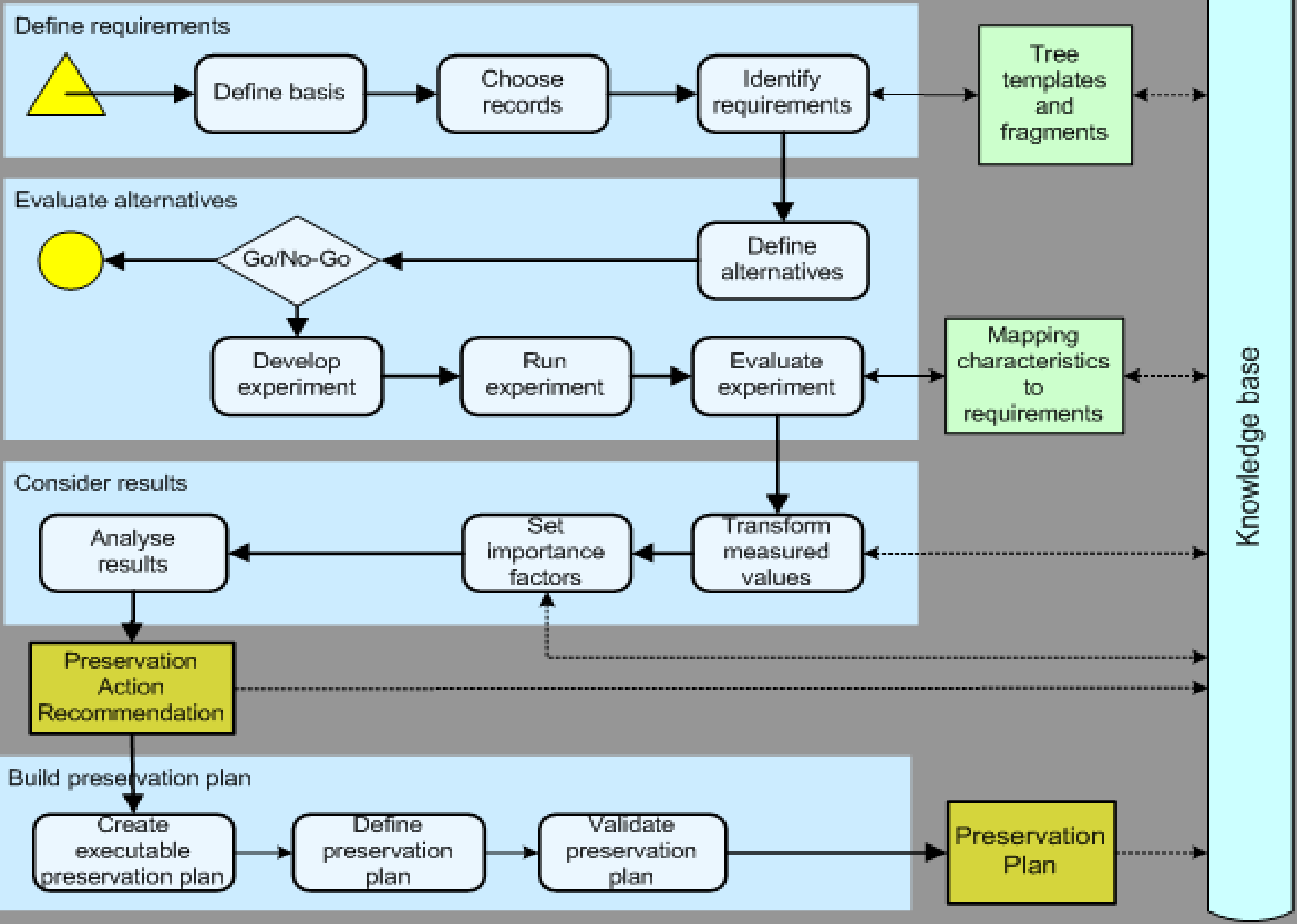


Requirements for objects

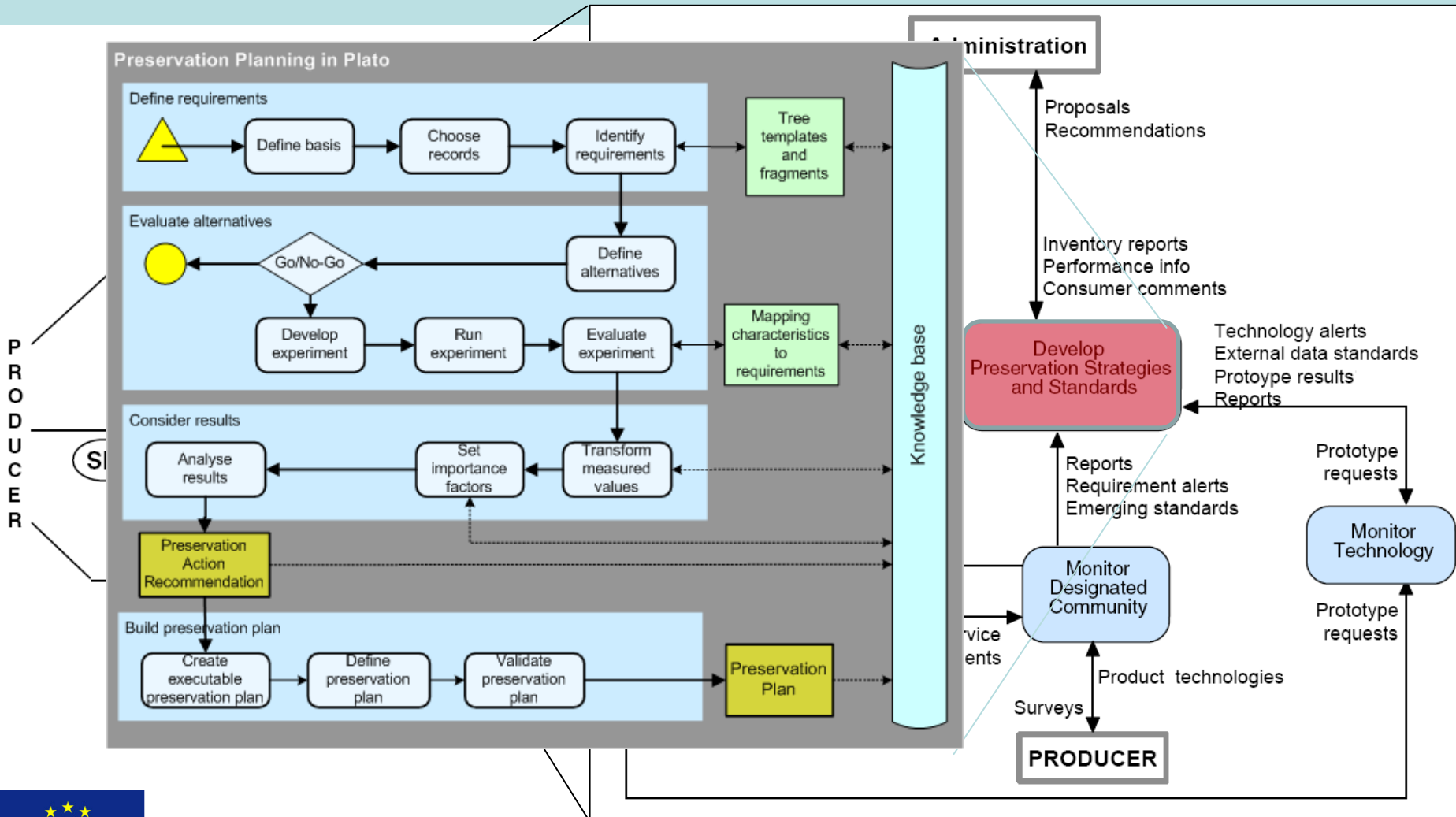
- **Authenticity**
 - to be what it purports to be,
 - to have been created or sent by the person purported to have created or sent it, and
 - to have been created or sent at the time purported
- **Reliability**
 - contents can be trusted as a full and accurate representation of the transactions, activities or facts to which they attest and can be depended upon in the course of subsequent transactions or activities
- **Integrity**
 - being complete and unaltered
- **Usability**
 - can be located, retrieved, presented and interpreted, so retrievable, readable, interpretable
- **Accuracy**
 - the degree to which data, information, documents or records are precise, correct, truthful, free of error or distortion or pertinent to the matter.



Preservation Planning in Plato



Preservation Planning and OAIS

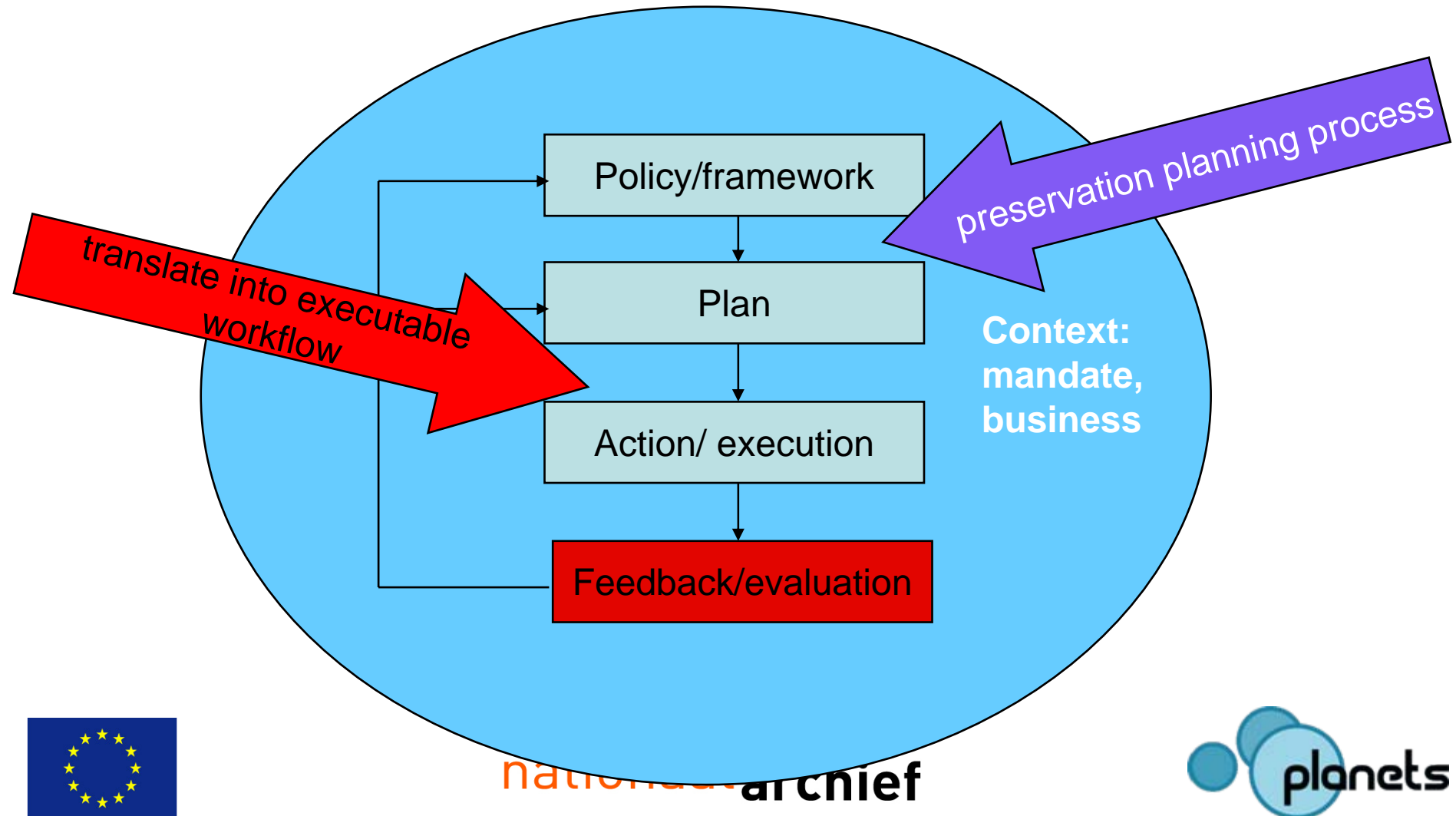


Characteristics of a preservation plan

- It is a concrete translation of a preservation policy how to handle/treat a certain type of digital objects in a given institutional setting
- New plans will be needed over time due to
 - ✓ changes in technology
 - ✓ changes in organisational setting
 - ✓ changes in user requirements
 - ✓ changes in available tools
 - ✓ changes in preservation methods
- It also specifies a series of steps or actions along with responsibilities and rules and conditions for execution
 - ✓ This is called **preservation action plan**. It is in the form of an executable workflow definition, detailing the actions and the required technical environment
 - ✓ Relationship with a specific action
 - ✓ The preservation plan provides the context/ background of the preservation action plan



From preservation policy to action



Planets functions

- **Technology watch**
 - Risk assessment service
 - Recommender services
 - Trigger for adapting preservation plans
- **Comparison of (available) preservation strategies based on organisational policies, usage information and collection profile(s)**
 - ‘Utility analysis’: identifying essential characteristics (tomorrow)
- **Developing and updating preservation plans according to (new) monitoring information and the available strategies evaluation**
 - The plans will trigger preservation actions (preferably in an automated fashion)
 - Tools and services registry
- **Validation framework (+ metrics) for evaluating the results of preservation actions**
- **Testbed**
 - Experimenting, documented and comparable results.
- **Characterisation**
 - File format registry

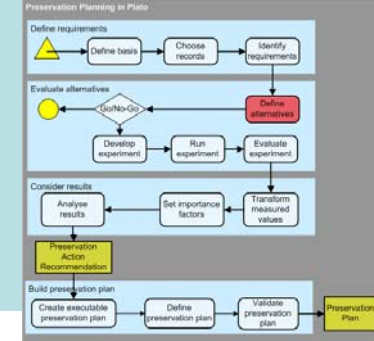


Planets approach: supporting tools

- Tool to support the decision making process (workflow) and a mechanism capable of translating those parameters into best available strategy
 - **Plato**, publicly available
 - leading to creation of a preservation plan
- Tool for making ‘collection profiles’
 - based upon technical and intellectual characteristics
- Tool for comparing and validating results of preservation actions (migration + emulation)
- Registries with information about available services/tools and their performance



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 |
|-------------------------------------|---------------------|--|-------------|
| <input type="checkbox"/> | JPG > BMP | Windows Bitmap, version 3.0 | JPG>BMP |
| <input checked="" type="checkbox"/> | JPG > TIF | Tagged Image File Format, version 3 | JPG>BMP>TIF |
| <input type="checkbox"/> | JPG > TIF #2 | Tagged Image File Format, version 3 | JPG>TIF |
| <input checked="" type="checkbox"/> | JPG > TIF_2 | Tagged Image File Format, version 3 | JPG>TIF_2 |
| <input type="checkbox"/> | JPG > PNG | Portable Network Graphics, version 1.0 | JPG>PNG |
| <input type="checkbox"/> | JPG > JP2 | JPEG 2000 | JPG>JP2 |

Create alternatives for selected services



Analyse Results

Aggregation method:

| Select | Alternative |
|-------------------------------------|-------------|
| <input checked="" type="checkbox"/> | PDF/A ToolA |
| <input checked="" type="checkbox"/> | PDF/A ToolB |

Show

[Expand All](#) | [Collapse All](#)

Minimalist root node

| Focus | 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 | ▼ 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:

Save recommendation

Generate final report

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

Thank you for your attention!

Questions?

www.planets-project.eu

hans.hofman@nationaalarchief.nl



nationaalarchief

