

Why do we have to plan digital preservation?

Christoph Becker

Vienna University of Technology
<http://www.ifs.tuwien.ac.at/~becker>



Copenhagen, June 2009



Overview

- What is the problem? Defining the issues
- Trustworthiness in repositories
- Preservation policies vs. Preservation plans
- What is a preservation plan?
- Preservation Planning Process: Scope and role
- What is needed? How to identify requirements?
 - The organisational/ business context, usage context, etc.



The issue / challenge

- The enormous and rapidly increasing amount of digital information
 - Fragile resources
- Rapid evolution in technology
- Obsolescence, corruption, loss of valuable information
- *Pro-active* and ongoing attention / maintenance required
- Potential solutions still fragmented
- Consumers need trust in digital content

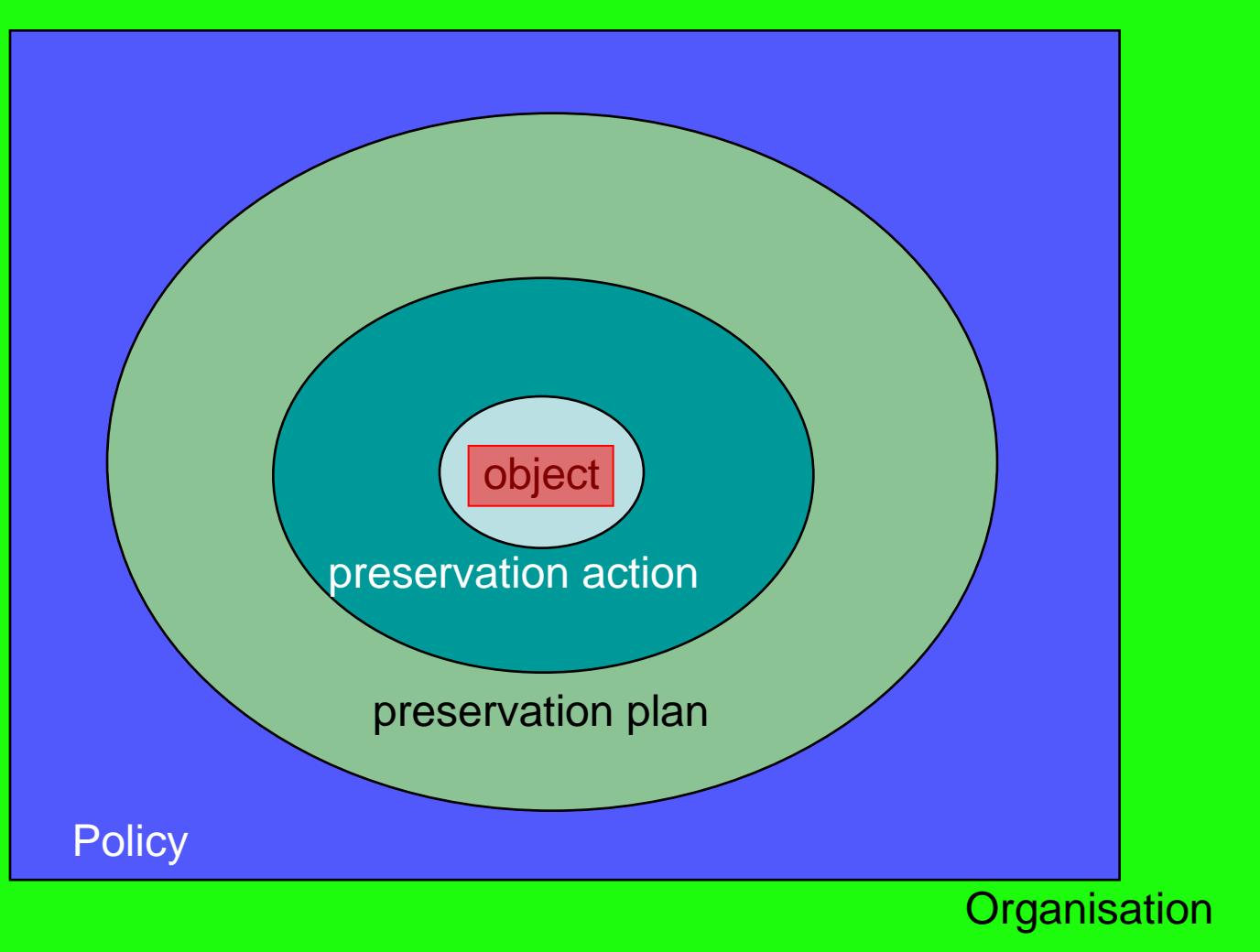


Stakeholders

- Memory institutions ('content holders')
 - archives, libraries
- (Scientific) data centres
- Government organisations (record creators)
- Business companies (record creators, intellectual capital)
- Individuals (e.g. family pictures)
- Trustworthiness



Objects in context



Trustworthiness in digital repositories

- RLG- National Archives and Records Administration Digital Repository Certification Task Force
 - Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
- NESTOR
 - Catalogue of Criteria of Trusted Digital Repositories
- DRAMBORA: Self-assessment



TRAC and Preservation Planning I

A 3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.

- Policies, plans, monitoring

A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.

- Preservation plans need traceability



TRAC and Preservation Planning II

B3.1 Repository has documented preservation strategies.

- Preservation Plan

B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.

- Monitor environment
- Update preservation plans



Nestor Criteria & Preservation Planning

- 8. The digital repository has a strategic plan for its technical preservation measures.
- 9.2 The digital repository identifies which characteristics of the digital objects are significant for information preservation.
 - Cf. TRAC B1.1



How to prepare?

- Understand the organisational context
 - mandate/ legislation
 - the organisational policy
 - user community
- Understand the objects
 - (collection of) digital objects: characteristics
- Understand the infrastructure
 - technology (past, present, future), infrastructure
 - people, knowledge, skills
- Available options
 - potential methods/ strategies
- Decision making process: preservation planning



How to prepare?

- **Understand the organisational context**
 - mandate/ legislation
 - the organisational policy
 - user community
- Understand the objects
 - (collection of) digital objects: characteristics
- Understand the infrastructure
 - technology (past, present, future), infrastructure
 - people, knowledge, skills
- Available options
 - potential methods/ strategies
- Decision making process: preservation planning



Preservation policy: situation now

- Trying to understand what organisations do in this area:
 - Large institutions are accumulating expertise and are building trusted digital repositories
 - Small institutions generally lack expertise and funding to build a digital repository
 - Large institutions have formulated various *requirements* – as can be discovered in different types of documents
- High level and abstract
- Developed a model for capturing organisational policies



Potential resources

- Mandate/ vision / mission statements
- Policy documents (if they exist)
- Project plans
- Guidelines
- Procedures/ rules



Preservation policy: examples

- Example policy statements of institutions with a digital preservation programme
 - UK Data Archive
 - National Archives of Australia
 - ISO/TR 18492:2005
Long-term preservation of electronic document-based information



UK Data Archive

- < UK Data Archive Preservation Policy
- <http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0308.pdf>
- p. 11: “The UKDA has chosen to implement a preservation strategy based upon open and available file formats, data migration and media refreshment.”
- What does this choice mean in practice? Two examples:
 - Emulation is –apparently– not a preservation strategy that will be chosen; all obsolete files will be migrated.
 - Migration to open file formats will be preferred.



National Archives of Australia

- < An Approach to the Preservation of Digital Records
- http://www.naa.gov.au/images/an-approach-green-paper_tcm2-888.pdf
- p. 14: “The digital preservation program must be able to preserve any digital record that is brought into National Archives’ custody regardless of the application or system it is from or data format it is stored in.”
- What does this choice mean in practice? One example:
 - all records that are accepted, should be preserved, regardless file format, medium, application, etc.
 - transform to open standard + keep ‘original’ format



ISO/TR 18492:2005

- International standard: Long-term preservation of electronic document-based information
- p. 12: Migration to standard formats

Storage repositories should consider **migrating** electronic document-based information from the wide variety of formats used by creators or recipients to a smaller number of “standardized” formats upon their transfer to the custody of the repository.

“**Standardized**” formats could be a consensus on formats that are widely used and are likely to cover a majority of a particular class of electronic document-based information. Proprietary file formats should be avoided. Among the technology neutral formats that merit consideration are PDF/A-1, XML, TIFF and JPEG.



Policies

- Framework (requirements) are identified at a high level
- Some requirements can already be explicit. Examples:
 - choice for one strategy
(e.g. migration to open document format)
 - choice that some types of records/documents can be denied because e.g. an *exotic* file format is used

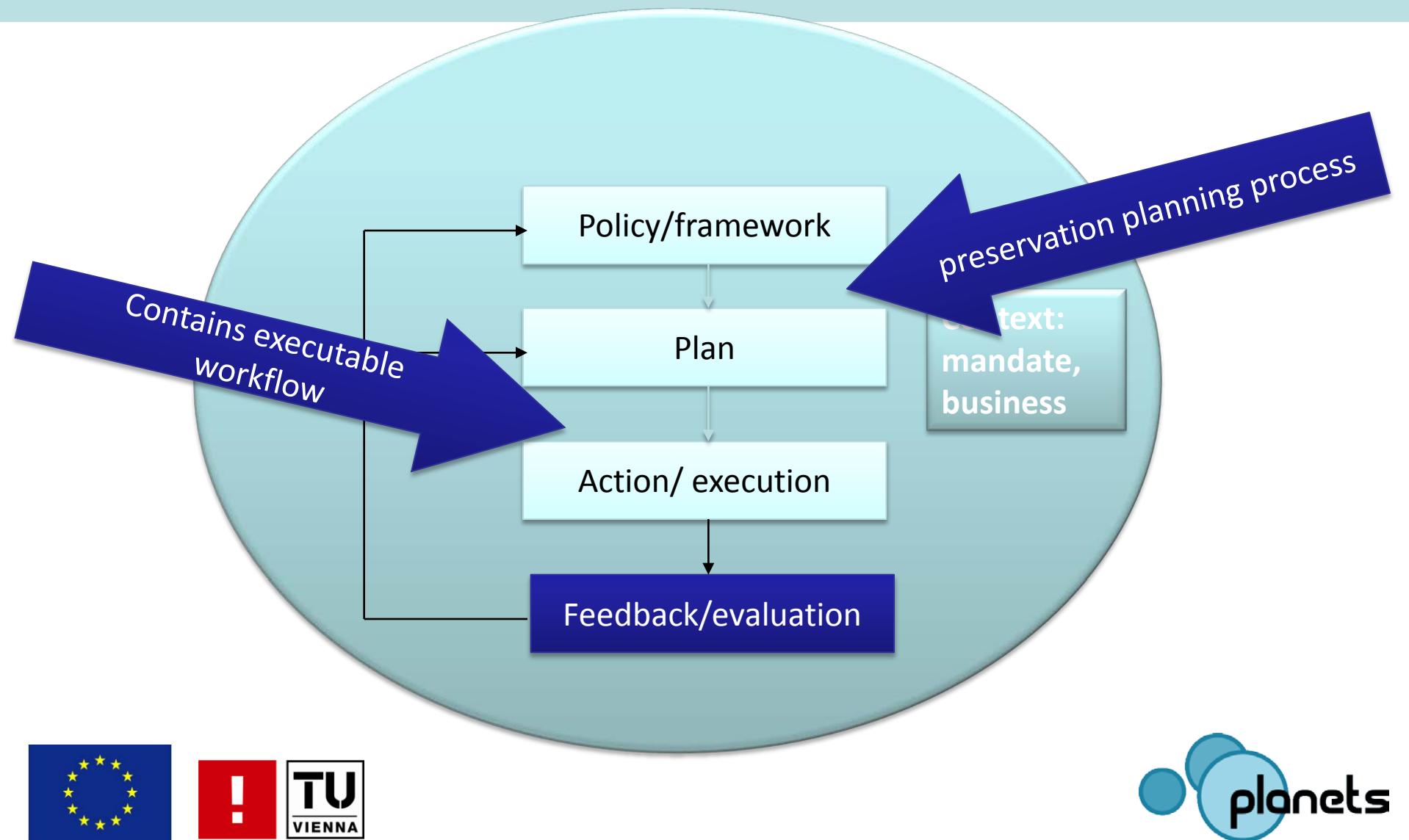


Policy model

- First version of model
 - describes policy requirements and organisational constraints
 - 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.)
 - covers policy requirements that are relevant for all types of preservation actions in different organisational settings



From preservation policy to action



Objectives of Preservation Planning

- Support decision-making about digital preservation
 - Identify criteria for preservation
 - Workflow for evaluating alternatives and defining preservation plans
 - Develop methodologies for assessing the risks of applying different preservation strategies for different types of digital objects
- Create, run and monitor preservation plans
 - Enable formulation, evaluation and execution of high-quality cost-effective preservation plans that suit the organisational needs
 - Support the on-going evaluation of the results of executing preservation plans and provide a feedback mechanism
 - Document the planning process carefully

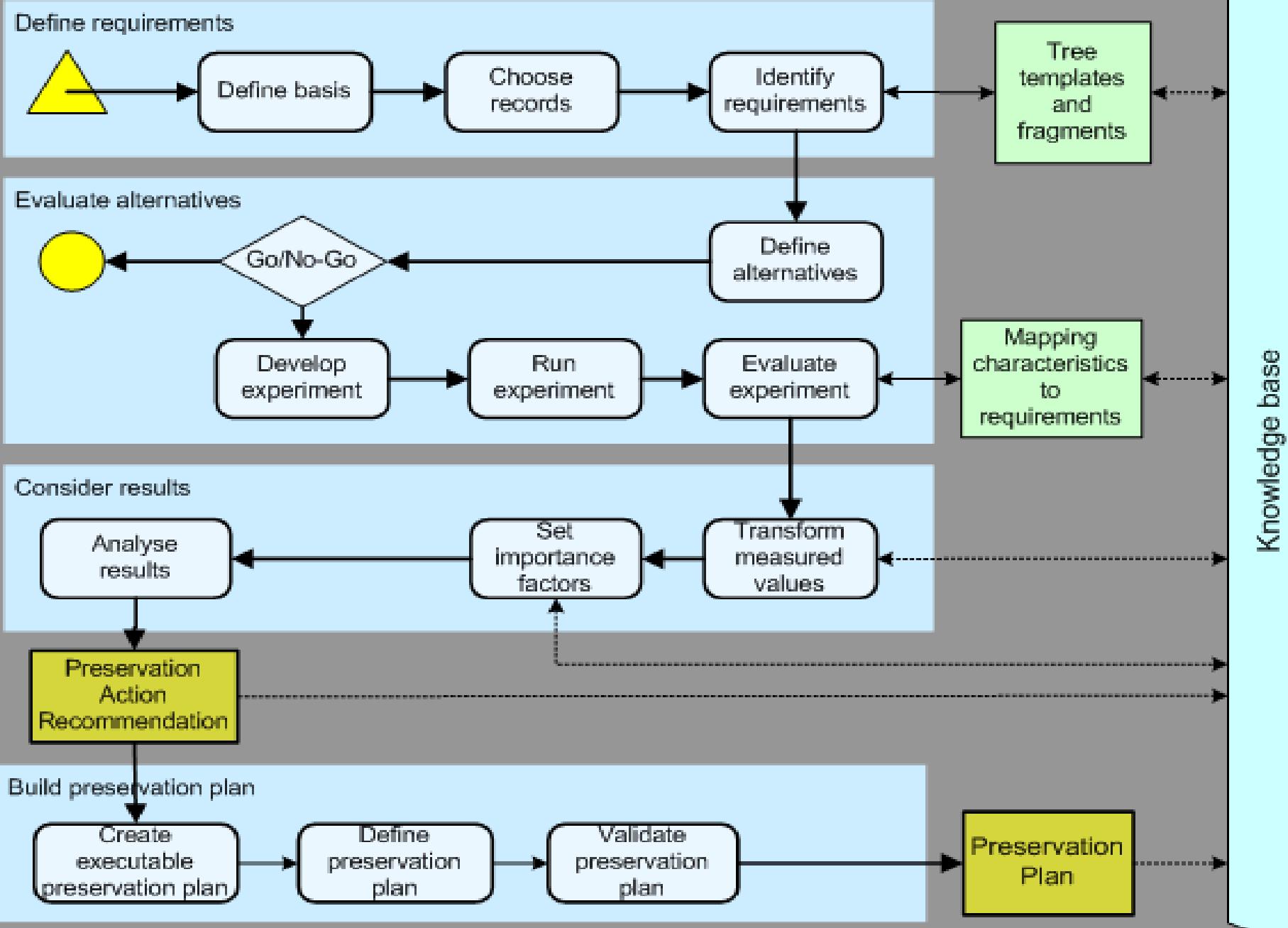


Issues

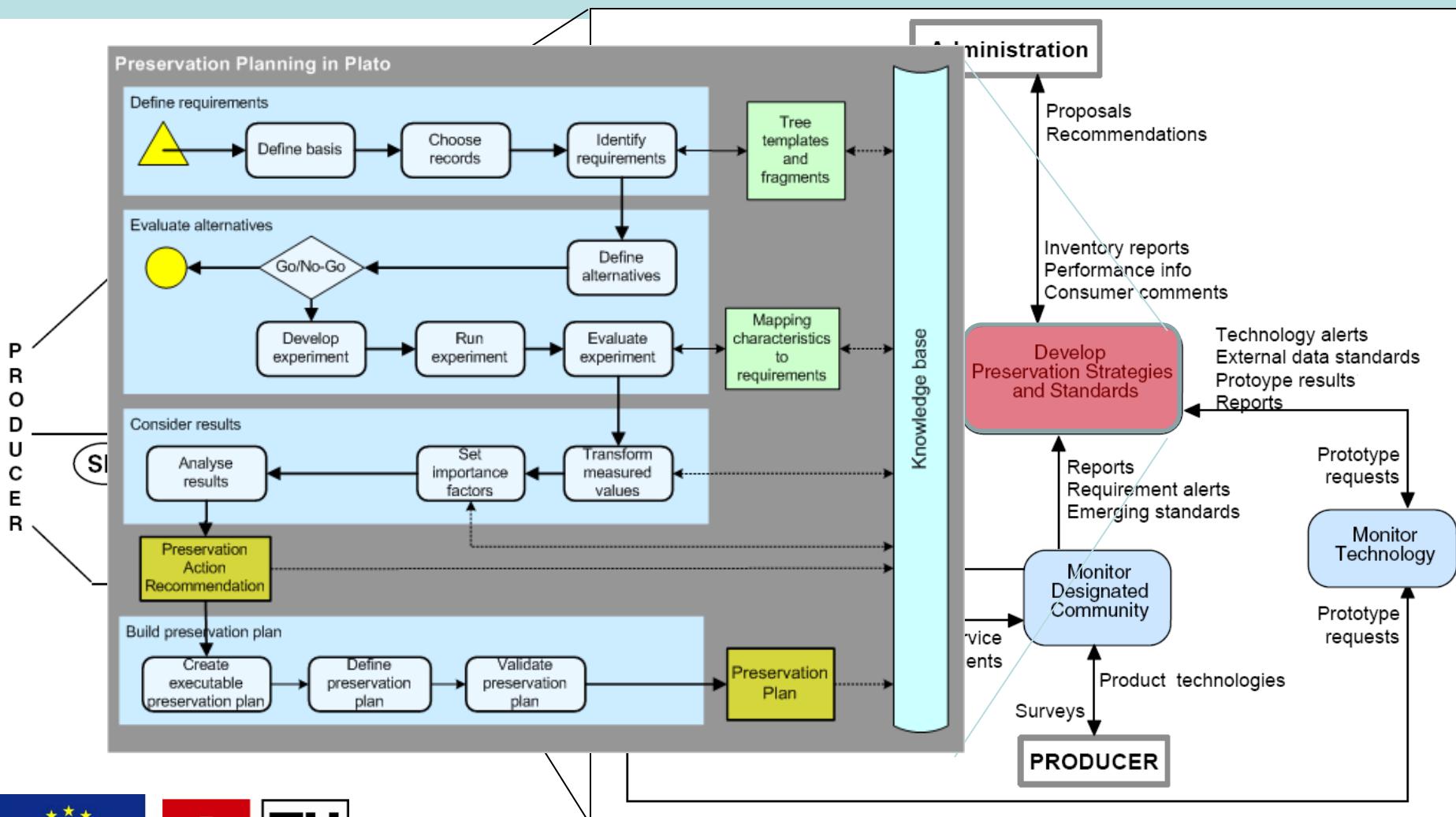
- How to develop a natural and logical flow of questions to be answered?
- How to translate that into a decision, documented in a preservation plan?
 - what is a preservation plan?
- How to enable the automated execution of the plan?
- How to evaluate the result of the execution of the plan?



Preservation Planning in Plato



Preservation Planning and OAIS



Definition of a Preservation Plan

- ‘A *preservation plan* defines a series of preservation actions to be taken by a responsible institution to address an identified risk for a given set of digital objects or records (called collection).’
- The Preservation Plan takes into account the preservation policies, legal obligations, organisational and technical constraints, user requirements and preservation goal. It also describes the preservation context, the evaluated alternative preservation strategies and the resulting decision for one strategy, including the rationale of the decision.



Characteristics of a preservation plan

- Translation of a preservation policy
- Specification of how to treat a collection in a given institutional setting
- Monitored for
 - ✓ changes in technology
 - ✓ changes in organisational setting
 - ✓ changes in user requirements
 - ✓ changes in available tools
 - ✓ changes in preservation methods
- Specifies concrete action
 - ✓ The **preservation action plan** can be an executable workflow definition, detailing actions and required technical environment
 - ✓ The preservation plan provides the context/background of the preservation action plan



The content of a preservation plan

1. Identification
2. Status
 - ✓ What was the immediate reason for this plan?
 - ✓ Has it been approved and if so, when and by whom
 - ✓ How does it relate to other plans related to a specific type of objects?
3. Description of institutional setting
4. Description of the collection (digital objects)
5. Purpose and requirements
6. Evidence of decision for a specific preservation action
 - ✓ what is the foundation of the decision
 - ✓ description of evaluation of possible actions
7. Costs considerations
8. Trigger for re-evaluation
9. Roles and responsibilities
10. Preservation action plan
 - ✓ executable program



Models and tools in Planets

- Conceptual model of potential preservation requirements (characteristics), both organisational and object-oriented
- Conceptual model of PP-process
- Validation framework (how to measure whether an action has been successful?)
- Definition of preservation (action) plan
- Tools:
 - Machine interpretable models for usage, collection profile, policy requirements/constraints
 - Plato tool for decision making: v.2 (November 2008)
 - Upcoming release in summer
 - Validation Framework - comparator - metrics
 - Technology watch service
 - Collection profiling and risk assessment (related to PRONOM)



A note on risk management

- Risks vs. trustworthiness
- TRAC, nestor: criteria checklists
- DRAMBORA: dynamic self-assessment
- DRAMBORA can be applied to analyse preservation planning
 - Activities
 - Assets
 - Risks
- Complementary



Summary and Outlook

- Understanding of context
 - organisational needs, user needs, legal requirements
- Identify criteria for preservation
- Determine what to keep/maintain
 - essential characteristics, objective trees
- Evaluate available strategies (actions) against criteria
 - identify best strategy
 - well-founded and documented decision
 - define preservation plan
- Execute plan when needed
- Evaluate what happened
- Re-iterate when changes happen in technologies, policies, usage, or the collection
- Automated decision process support



Thank you for your attention!

Questions?

www.planets-project.eu

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

becker@ifs.tuwien.ac.at

