



Introduction to Preservation Planning

The Why and the How

Christoph Becker

Vienna University of Technology

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



Bern, November 2009



Outline

- Why preservation planning?
 - Trusted digital repositories
 - Criteria for trustworthiness
 - Policies vs. plans
- Preservation Planning
 - What is a preservation plan?
 - Outlook: How to create a preservation plan



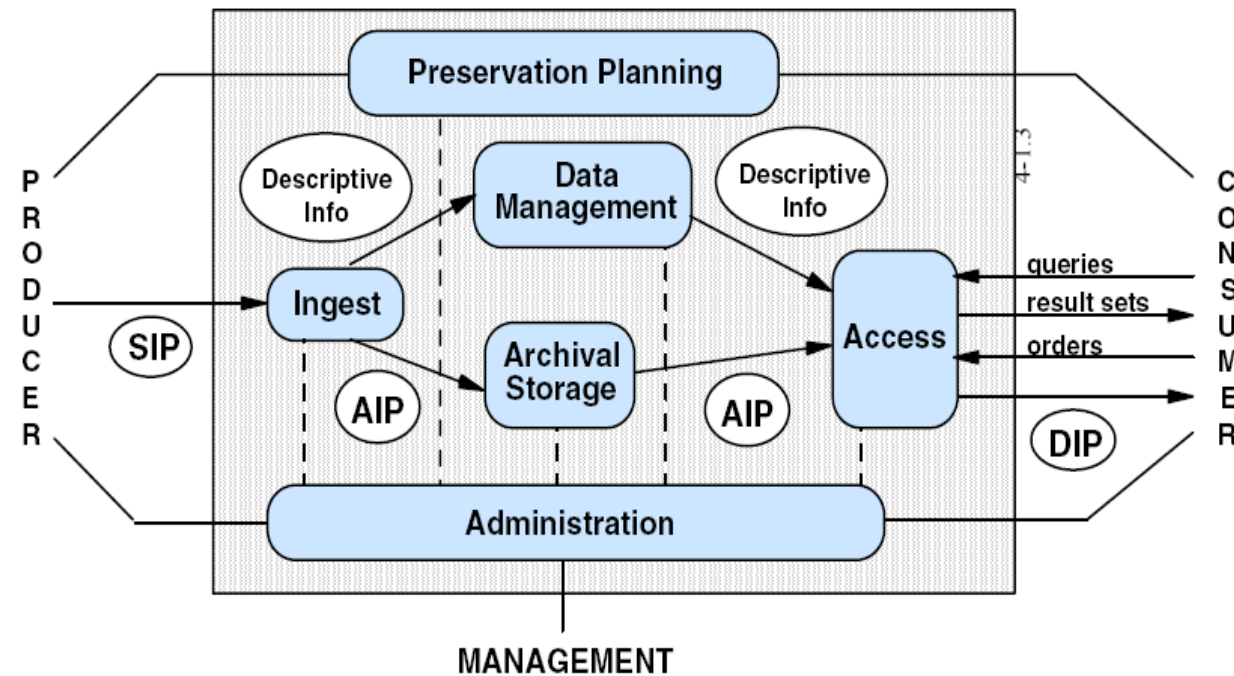
The challenge

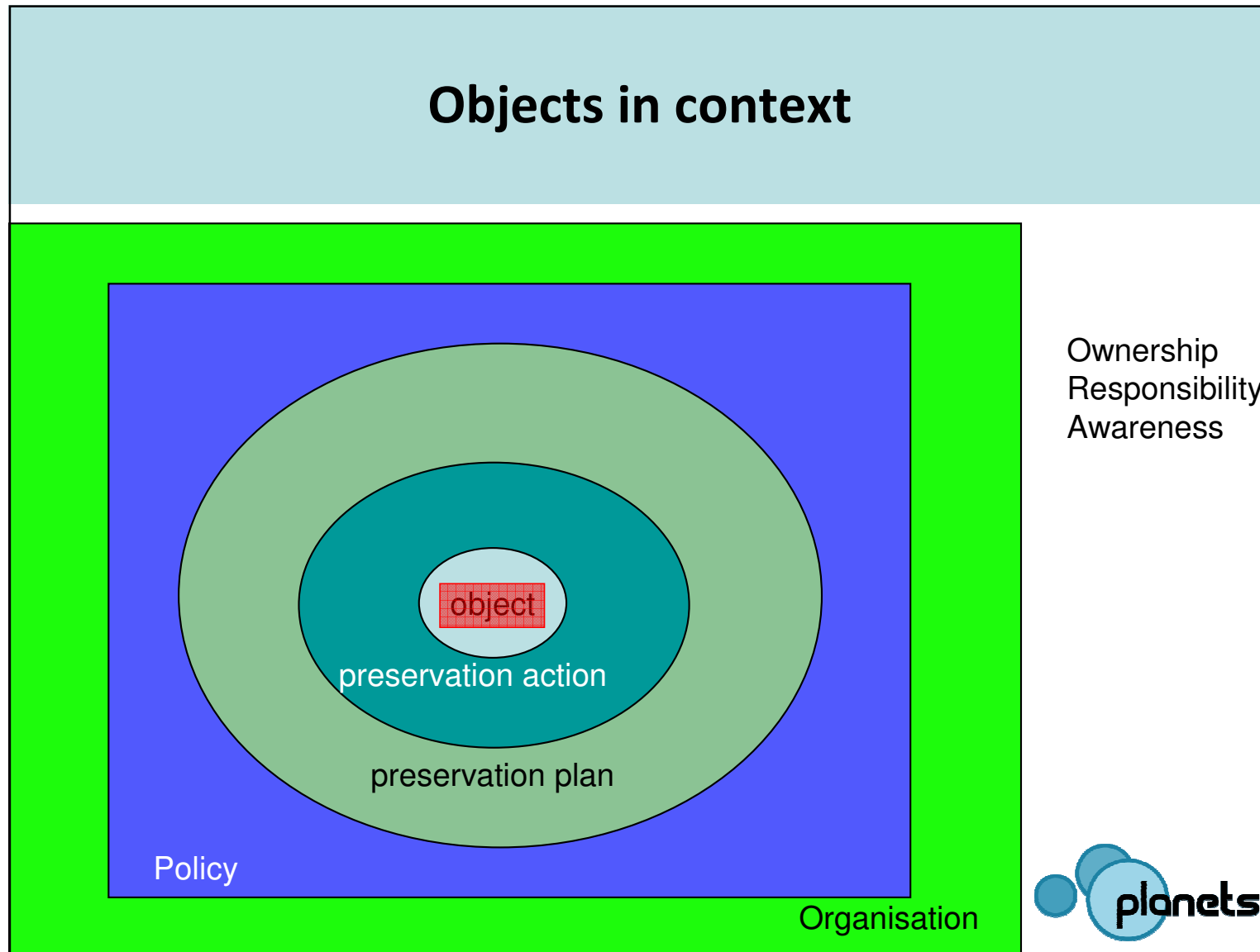
- Issues
 - Enormous and rapidly increasing amount of digital information
 - Fragile resources and rapid evolution of technology
 - Obsolescence, corruption, loss of valuable information
 - *Pro-active* and ongoing attention / maintenance required
 - Potential solutions still fragmented
- Stakeholders
 - Memory institutions ('content holders'): archives, libraries
 - (Scientific) data centres
 - Government organisations (record creators)
 - Business companies (record creators, intellectual capital)
 - Individuals (e.g. family pictures)
- Trust in digital resources



Trustworthiness in digital repositories

- Producers need trust in digital repositories
- Consumers need trust in digital repositories
- Repositories need trust in external providers





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:
Repository identifies properties it will preserve for digital objects



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.

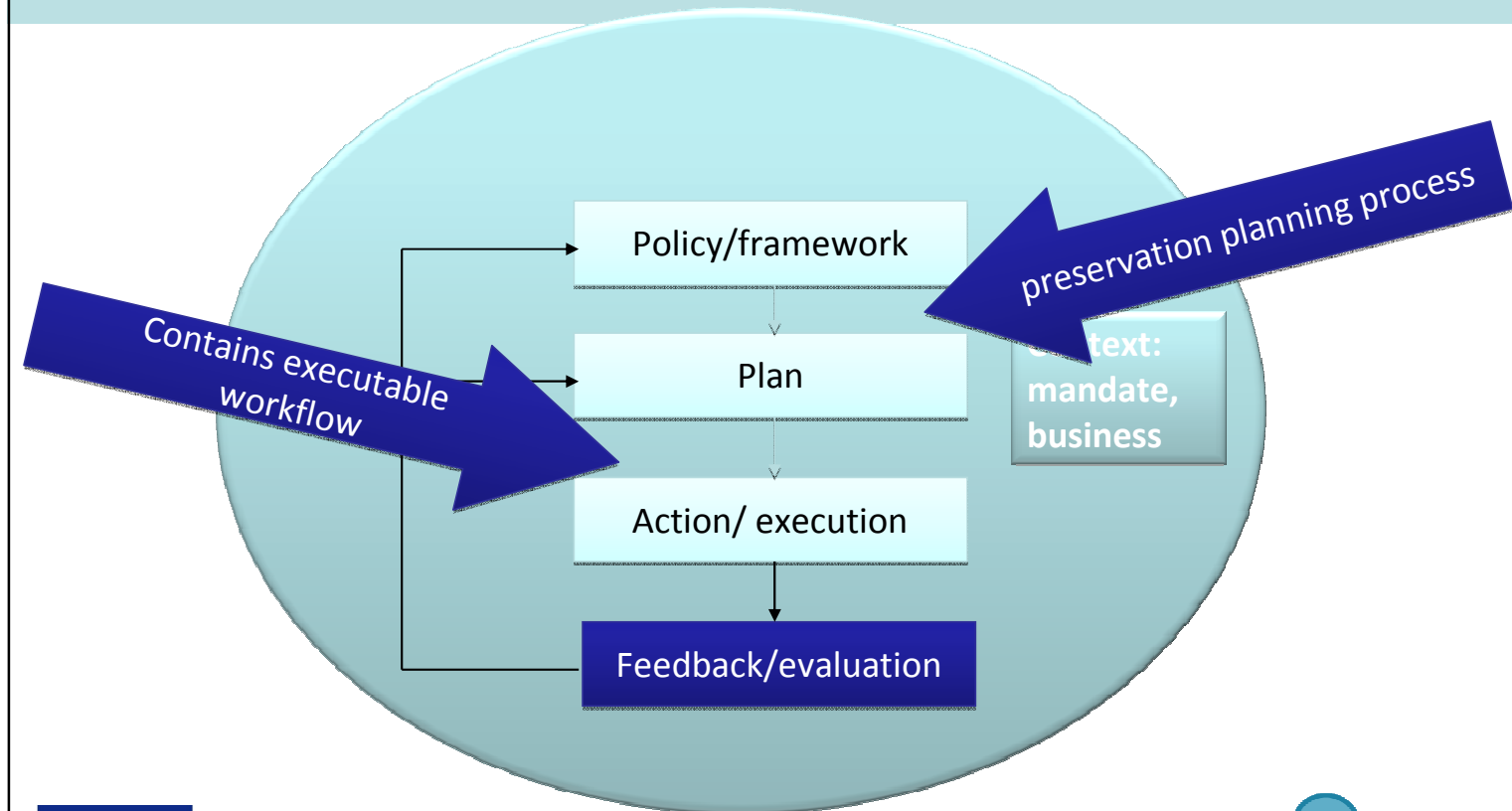


Policies

- 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



From preservation policy to action

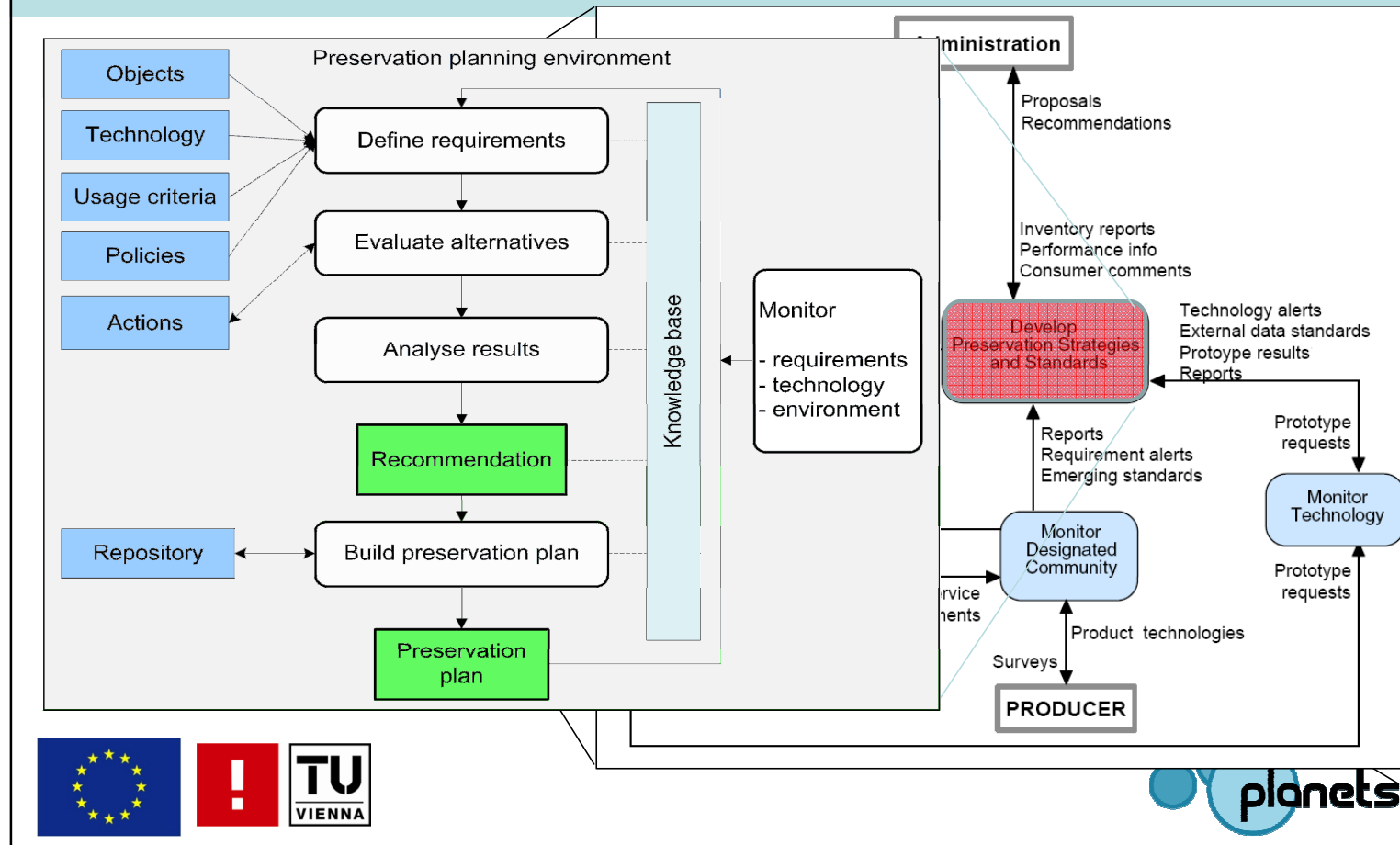


Objectives of Preservation Planning

- 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



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

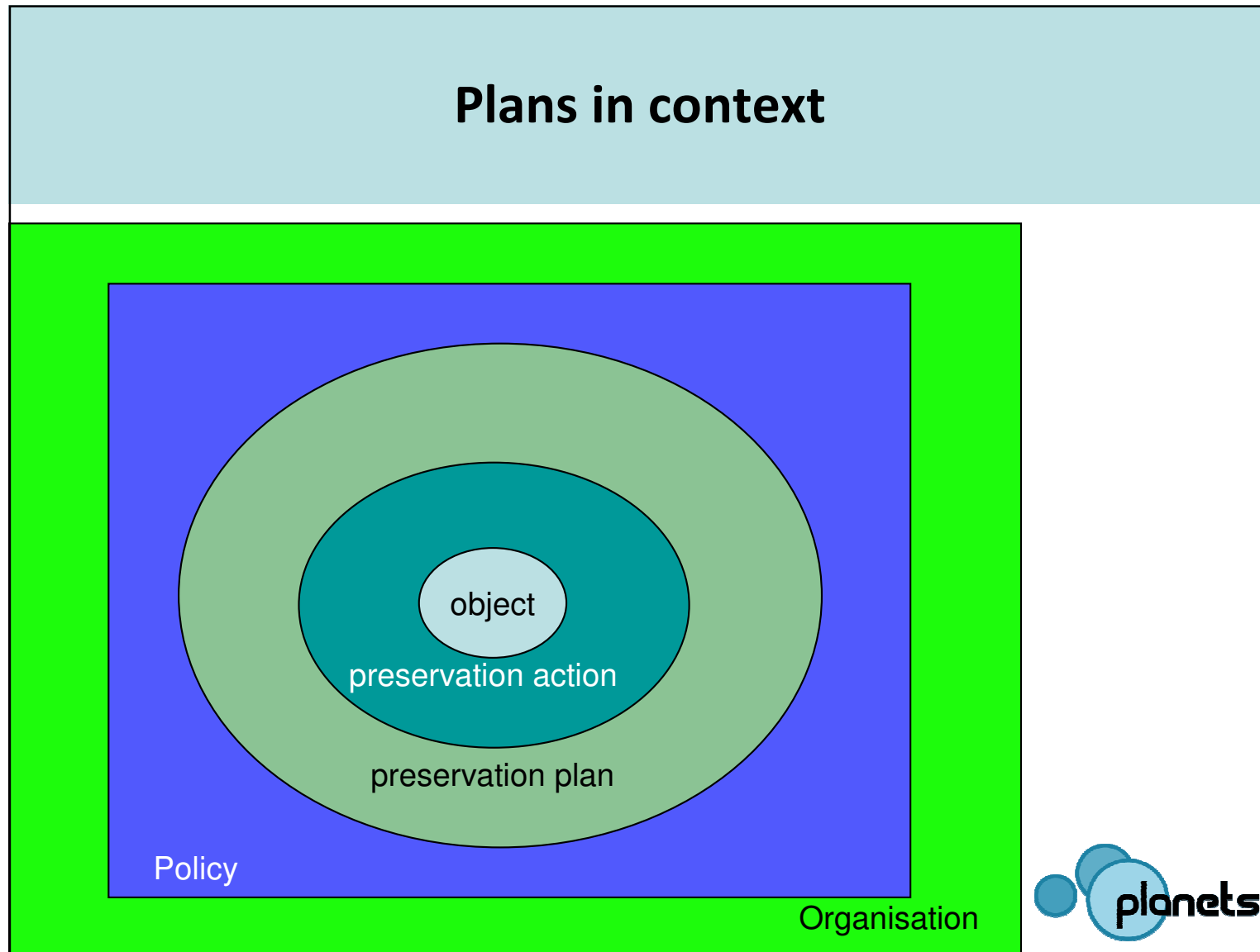
- Translation of a preservation policy
- Specification of how to treat a collection in a given institutional setting
- Monitored for
 - ✓ changes in technology
 - ✓ changes in organisational setting
 - ✓ changes in user requirements
 - ✓ changes in available tools
 - ✓ changes in preservation methods
- Species 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

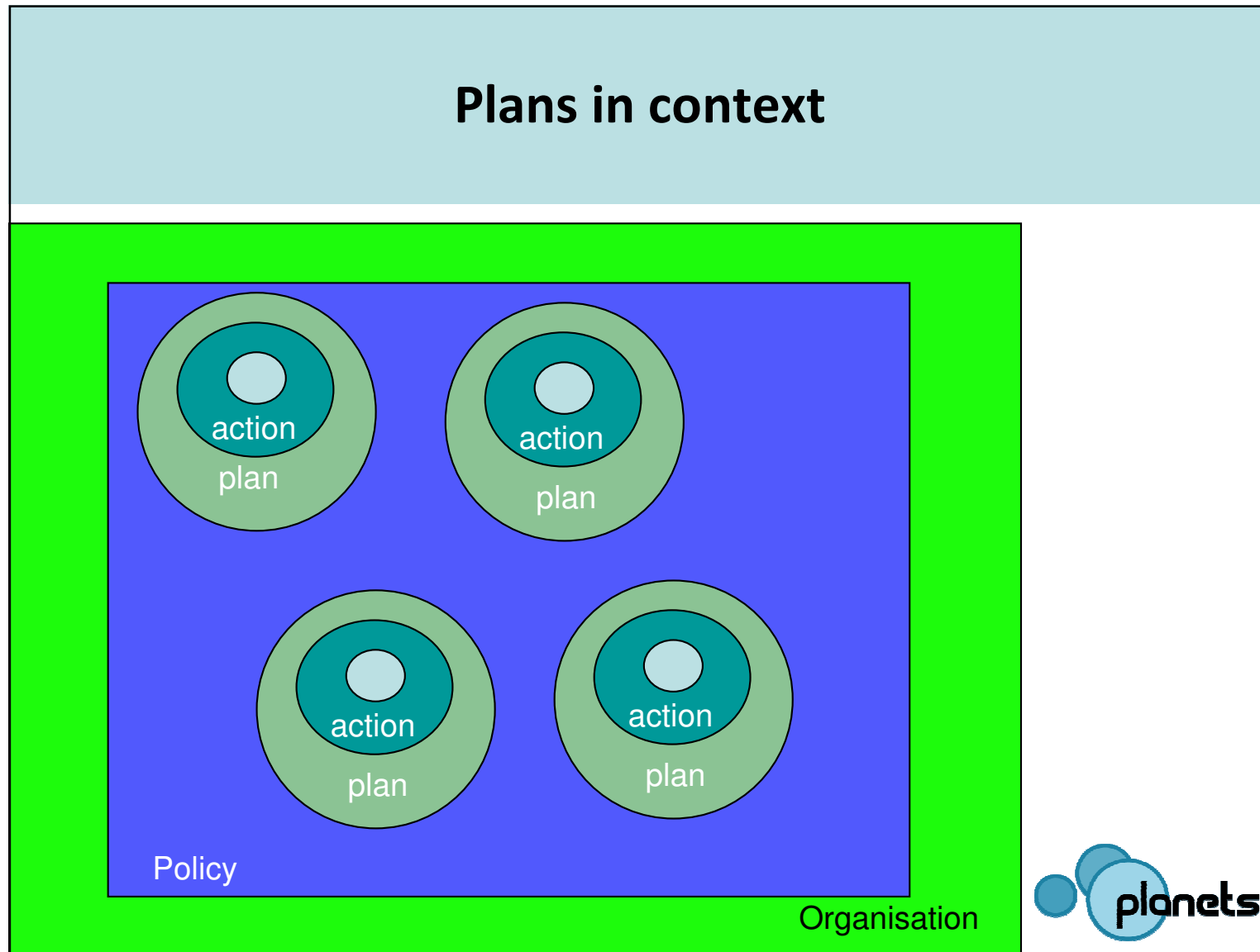


The content of a preservation plan

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



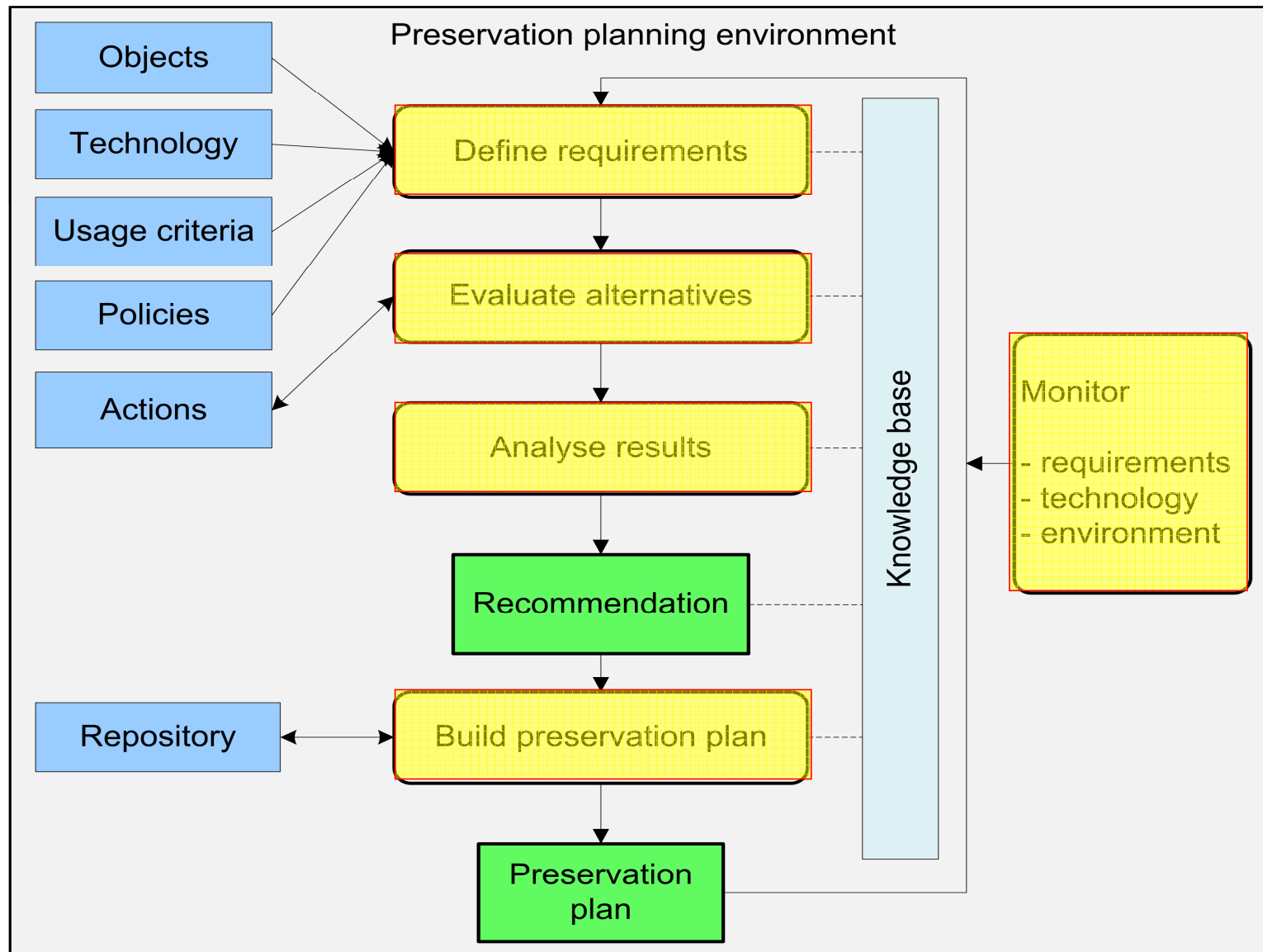




Evaluating preservation strategies

- Variety of solutions and tools exist
- Each strategy has unique strengths and weaknesses
- Requirements vary across settings
- Decision on which solution to adopt is complex
- Documentation and accountability is essential
- Preservation planning process assists in decision making
- Evaluating preservation strategies on representative samples according to specific requirements and criteria





Preservation Planning in Plato

- Web based planning tool implementing the Planets preservation planning workflow
- Publicly available
- Automation of the planning process
 - Integration of registries and services for
 - File format identification
 - Preservation action (migration, emulation...)
 - Characterisation and comparison
- Knowledge base to support planning
- New release 2.1!
- <http://www.ifs.tuwien.ac.at/dp/plato>



Summary and Outlook

- Why preservation planning?
- Trustworthy repository criteria
- What is a preservation plan?
- Understanding the context
 - organisational needs, user needs, legal requirements
- Evaluate available strategies (actions) against criteria
 - identify best strategy
 - well-founded and documented decision
 - define preservation plan
- Execute plan
- Continuous monitoring
- Automated decision process support: PLATO



Thank you for your attention!

Questions?

www.planets-project.eu
www.ifs.tuwien.ac.at/~becker
becker@ifs.tuwien.ac.at

