Preservation and Long-term access through Networked Services

Hans Hofman
Nationaal Archief Netherlands

DigCCurr
Chapel Hill, 19 April 2007

planets

digital preservation research and technology

EU Support for digital preservation

- Major initiative in the Information Science and Technology (IST) Framework Programme 6 Call 5
  - Two Integrated Projects funded: **Planets** (BL), **Caspar** (CCLRC)
  - Coordinated action: **DPE** (HATII at Glasgow)
  - Research projects
- Planets builds on strong digital archiving and preservation programmes at European, National and institutional levels
  - 16 partners: 5 libraries, 3 archives, 4 IT-companies and 4 universities
  - Four year project starting June 2006 with 15me budget

Planets goals

- Planets
  - Addresses core digital preservation challenges
  - Uses an empirical approach to learn what works and why
- Increase Europe's ability to ensure long-term access to its cultural and scientific heritage
  - Improve decision-making about long term preservation
  - Ensure long-term access to valued digital content
  - Control the costs of preservation actions through increased automation, scaleable infrastructure
  - Ensure wide adoption across the user community and establish market place for preservation services and tools
- Build practical solutions

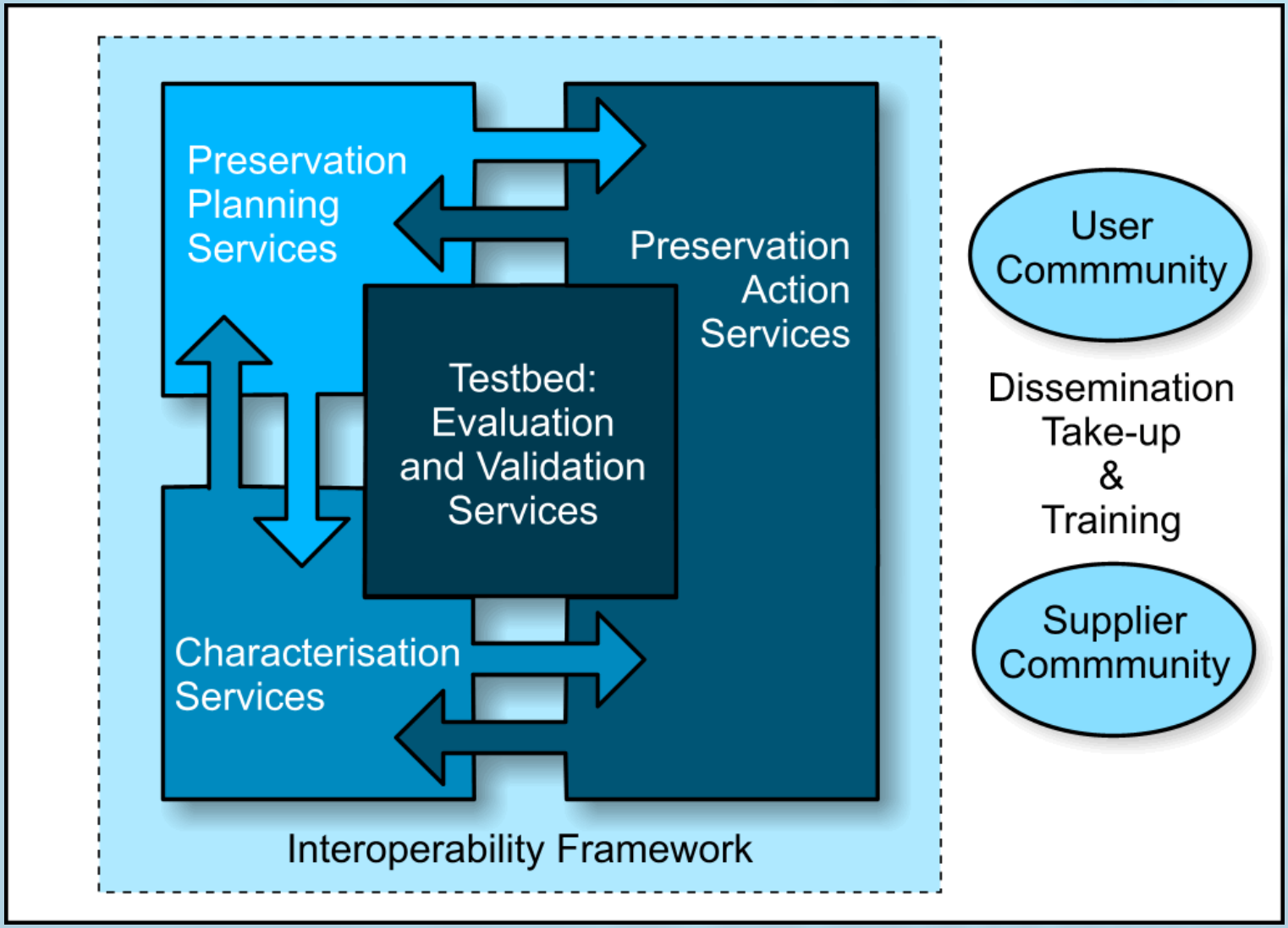
Approach

- Develop a practical and proven methodology for preservation planning
- Build practical solutions by
  - Integrating existing expertise, designs and tools
    - e.g. Delos digital preservation cluster, TNA/Pronom, Dutch KB/ NA emulation project, …
  - Develop methodologies for decision making, evaluation and testing
  - Build or enhance registries on file formats, tools and services
  - Sharing and if necessary building new tools
- Provide an interoperable distributed open source environment, which will also enable
    - Third-parties to provide tools and services
    - Vendors to integrate preservation services
    - Content owners to ensure long-term access to their digital content

Planets Project Components

- **Planning services** that empower organisations to define, evaluate, and execute preservation plans
- Methodologies, tools and **services for Characterisation** of digital objects
- Innovative solutions for **Preservation Actions**
- An **Interoperability Framework** provides services distributed services
- A **Testbed** enables objective evaluation of protocols, tools, services and plans
- **Outreach**, workshops and training to engage the user and vendor communities

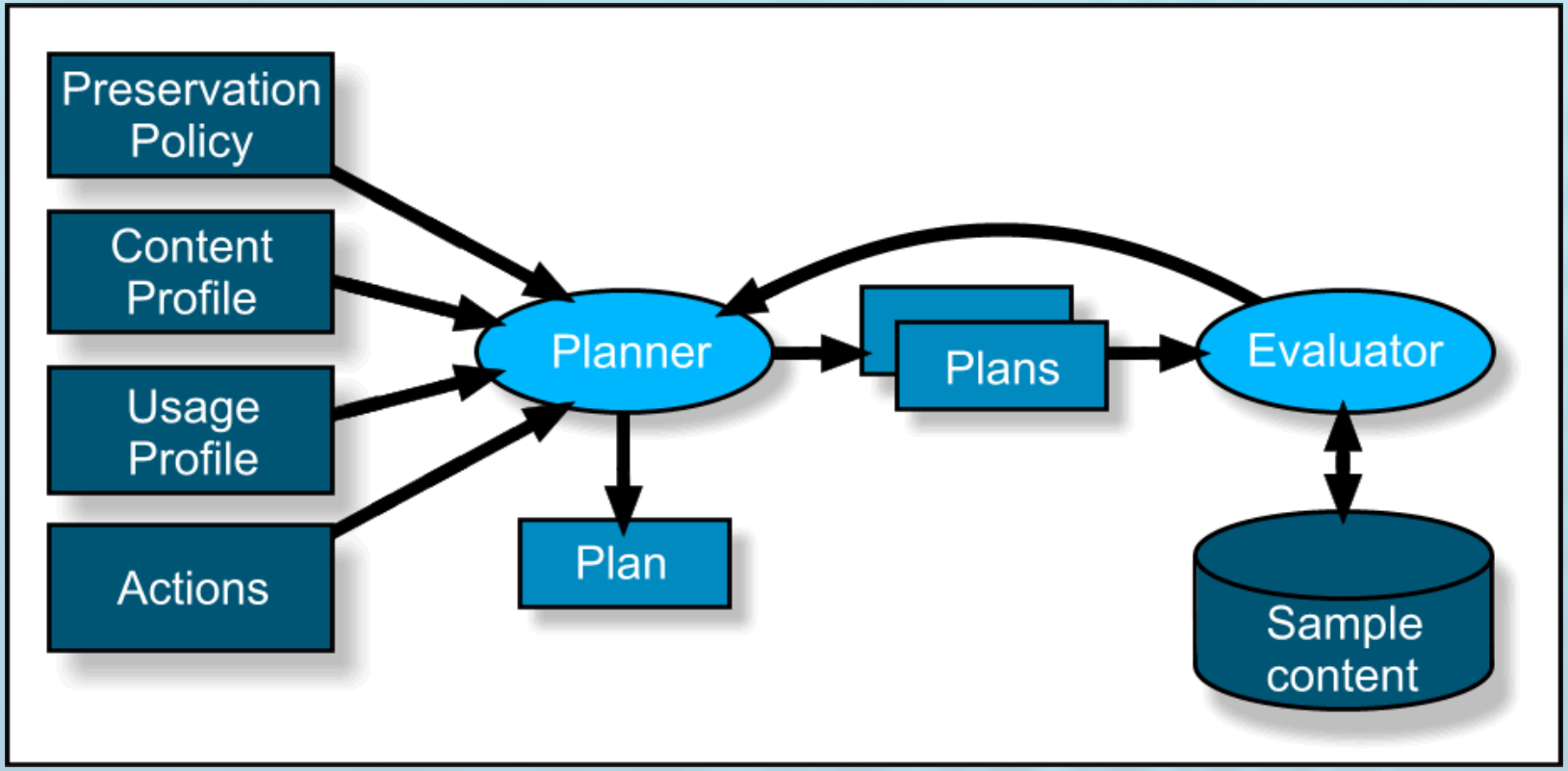
Project architecture reflects problem structure



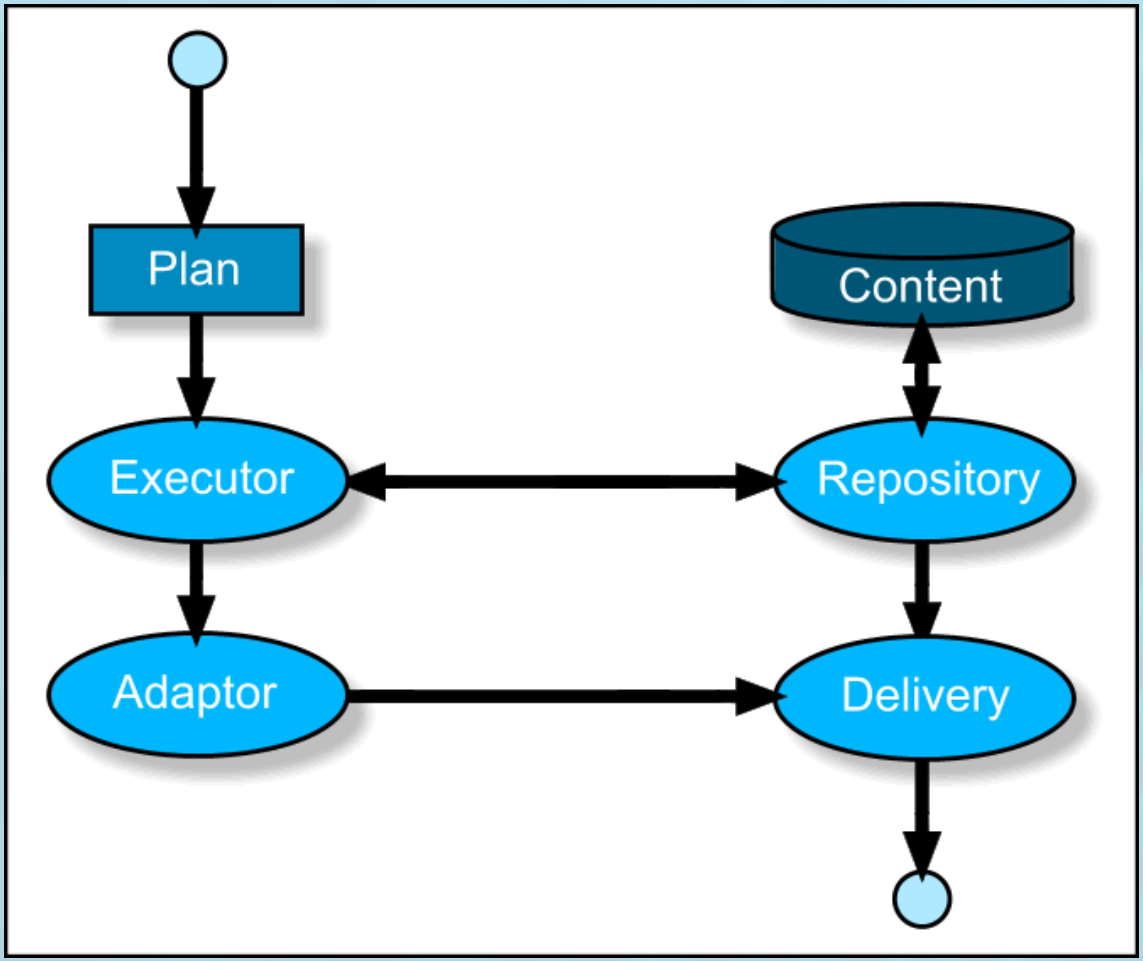
Preservation planning decision making

- Define a well-defined, consistent, proven and documented processes of preservation planning and execution
- Identify criteria for preservation, based upon
  - Organisational policies
  - ‘Collection’ profile
  - Provenance of digital objects (authenticity !)
- Identify and evaluate potential preservation actions
- Develop preservation plan(s)
- Requirements/ intentions:
  - To make it a pro-active process
  - To be automated as much as possible

Preservation planning decision-making process



Preservation plan execution



Characterise objects/ content

- Characterise content to check on/ evaluate preservation actions
- Two associated approaches
  1. Intellectual approach: building objectives trees (based upon ‘utility analysis’)
  2. Extraction of intrinsic file (format) information
- Build on TNA’s PRONOM for file-format identification
  - Define a characterisation language
  - Define an extraction language
  - Define an pluggable interpreter
- Extend to measure loss due to actions
- Leverage understanding to improve file formats
  - Address a root cause of digital obsolescence

Preservation actions

Transform content/objects

- Wrap third-party transformation tools
- Fill gaps with new tools
- Preserve relational databases
  - Build on Swiss Federal Archives' work
- Preserve Office content
  - Build on MSFT tools

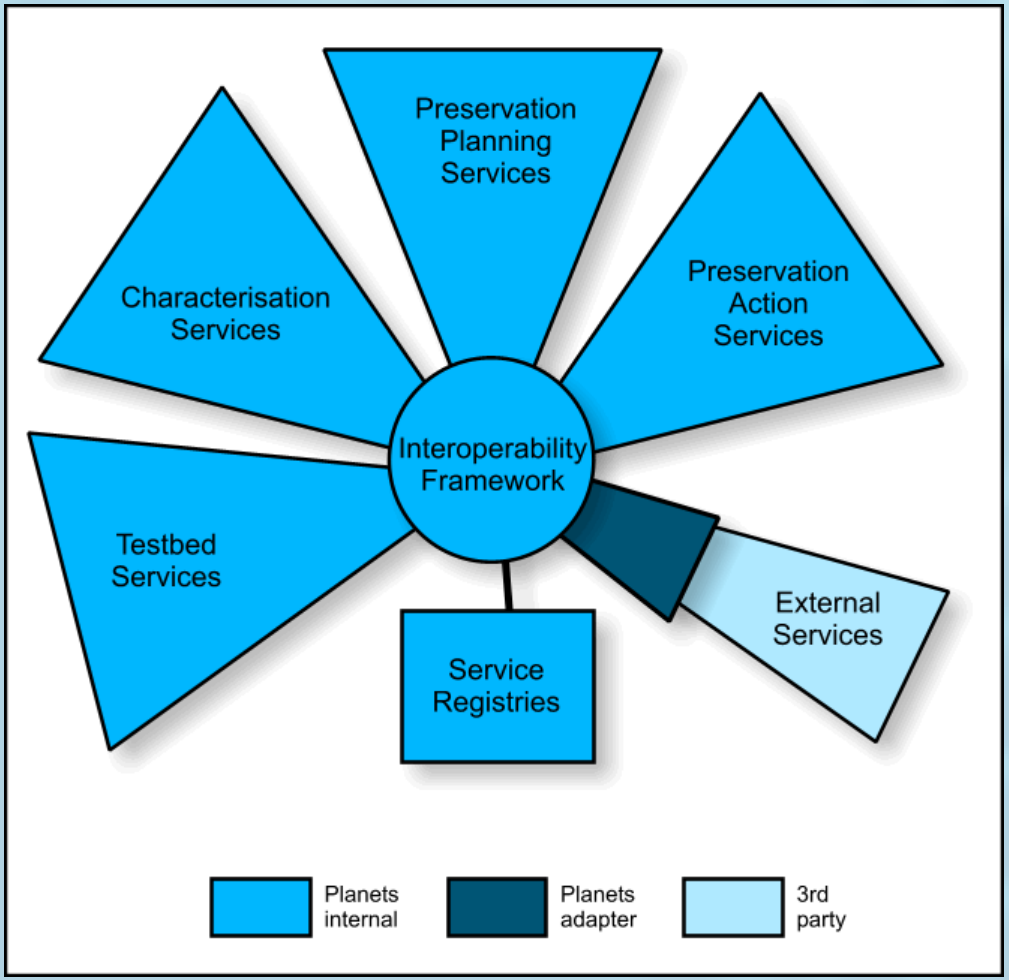
Transform environments

- Modular emulation of the full hardware/software environment
  - Provides full look-and feel
  - Superb for highly dynamic content
- Layered durable emulation
  - Build on IBM Universal Virtual Computer (UVC)
  - Establish abstract device drivers

Testbed

- Provides a foundation for objective evaluation
  - Design experiments and tests
    - includes well-defined corpora of digital objects
  - Experiment: collect data, evaluate results, compare
  - Validation framework for evaluating preservation plans
  - Benchmark tools and services
- Consists of
  - Data storage, hardware, Planets software, testbed software
  - Templates for experiment design, testing and documenting
  - Benchmark and other content
- Provides resources for
  - The project partners
  - The preservation community
  - External organisations
  - Tool and service certification

Interoperability Framework



- Planets provides an interoperability framework including
  - Interoperable distributed services
  - Service registries and shared data-stores
  - Encapsulate tools as services
  - Orchestration capability to combine services

What digital curators do and what they need to know?

- Understand the preservation processes, especially
  - preservation planning: how to take the best decisions to execute the appropriate and tested preservation actions
  - know how to identify what criteria should inform those decisions in different contexts
  - know how to apply those criteria on digital objects: what is it we need to preserve?
  - to test and evaluate available preservation strategies with respect to a given type of objects
  - how to do this in an effective and efficient way
  - …
- Training programme on Planets results, but
  - co-ordination of efforts with other initiatives such as DPE, Caspar, Nestor, DPC and others
    - modular
    - bring together course materials
    - building upon work by ERPANET

- Questions?

- **For more information:**
  - http://www.planets-project.eu
  - Planets-info@bl.uk
  - hans.hofman@nationaalarchief.nl