

The Planets Project

September 2008

Peter Bright



- A 4-year research and technology development project co-funded by the European Union to address core digital preservation challenges
 - Improve decision-making about long term preservation
 - Ensure long-term access to valued digital content
 - Control the costs of preservation actions through increased automation, scalable infrastructure
 - Ensure wide adoption across the user community and establish market place for preservation services and tools
- Started June 2006 with €15m budget
- Builds on strong digital archiving and preservation programmes at European, National and institutional levels
- Coordinated by the British Library and involves 16 partners including national libraries and archives, leading technology companies and research universities



The British Library
 National Library, Netherlands
 Austrian National Library
 State and University Library, Denmark
 Royal Library, Denmark

 National Archives, UK
 Swiss Federal Archives
 National Archives, Netherlands

3



Tessella Plc
 IBM Netherlands
 Microsoft Research, Cambridge
 Austrian Research Centers

 Hatii at University of Glasgow
 University of Freiburg
 Technical University of Vienna
 University at Cologne

4



Motivation & benefits

National Libraries & Archives

- Have the legal responsibility and the legislative framework to safeguard digital information
- Have been collecting digital documents and records since 1982
- Realize that meeting the challenge of preserving access goes beyond the capabilities of any single institution
- Have limited ability to ensure that today's digital information will be accessible for future generations
- Collaboration with research & IT is a must
- Need pragmatic solutions here and now

What's in it for us?

"Planets will provide the technology component of our digital preservation solution"

Richard Boulderstone, 15/06/07

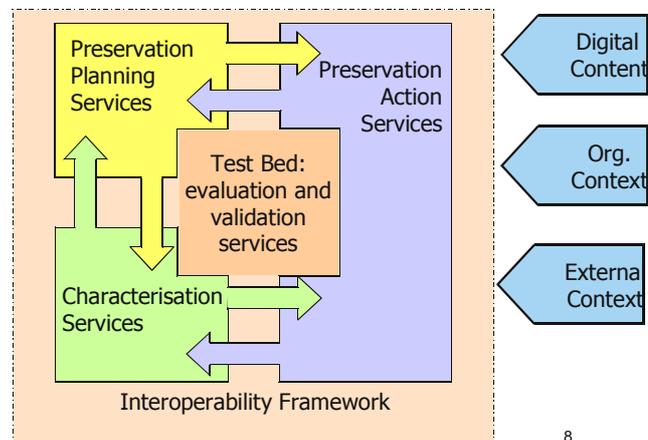
Planets will enable us to

- Profile our digital collections against our policies
- Identify and diagnose problems in our digital collections
- Compare different treatment plans
- Select and implement treatment for a wide range of problems
- Verify that the treatment was successful
- Know how solutions work through empirical evidence

... and encourage vendors and service providers to provide these capabilities to us

7

Project architecture reflects problem structure



8

Preservation Actions

- There are two main strategies to provide long-term access to digital objects:
 - Change the object (migration)
 - Change the environment (emulation)
- Planets is developing:
 - Emulation tools
 - Dioscuri emulator, off-the-shelf virtualization
 - Migration tools
 - Wrapping existing software where possible
 - Creating new software where necessary

Preservation Characterisation

- To provide access to digital objects we need to know about the objects:
 - What kind of object they are
 - What their significant properties are
- Extracting this information is “characterisation”
- The Preservation Characterisation work package will produce:
 - A standard way of describing object characteristics (XCDL)
 - Tools and services to characterise objects
 - e.g. droid, XCEL

Preservation Planning

- The key to Preservation Planning is
 - making the right decisions...
 - to take the right action...
 - at the right time
- Preservation decisions depend on content characteristics, technical infrastructure, tool availability, and organizational policy
- The Planets Preservation Planning work package will produce technology to enable
 - creation,
 - evaluation,
 - and execution of Preservation Plans

11

Interoperability Framework

- The Interoperability Framework joins the Preservation Action, Preservation Characterisation, and Preservation Planning software together
- It provides:
 - Standard interfaces so that new Actions/Characterisations can be developed and seamlessly integrated
 - Registries of existing tools to enable discovery and tool selection
 - Orchestration of tools into workflows

12

Testbed

- Evaluation of tools is an integral part of the decision-making process, e.g.:
 - Does the migration retain significant properties?
 - How long will a migration take?
- The Testbed provides a “laboratory” for testing Planets tools and workflows

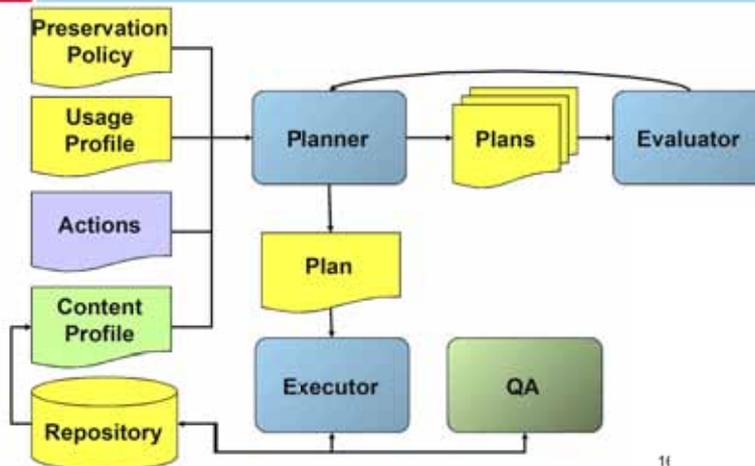
How might it work – donation scenario

- A scientist donates her on-line research repository
 - Stretching back thirty years
 - Papers, technical reports, notes in many formats
 - Original research data
 - Software tools that implement research ideas
- Step 1: Ingest original contribution into repository
- Step 2: Characterise files – acceptable; unacceptable; unknown
- Step 3: Convert unacceptable files into acceptable formats
 - Select best conversion tools, parameters
 - Run tools to convert
 - Perform automated QA on results, identifying problems
 - Address problems by hand
- Step 4: Ingest conversions into repository
 - Record relationship to originals

Scenario: Notice change in user community

- Sample policy: 90% of users can access all published reports
- Content profile: 5% of published reports in dvi format
- Usage profile: 98% of users can not view dvi files
- Identify possible plans (using PLATO) including
 - Convert to PDF
 - Convert to tiff
 - On access conversion to PDF
 - Provide users with viewer plug-in
 - Provide emulation environment for 1990 workstation ☺
- Select plan (using PLATO, testbed empirical data)
 - Such as convert to PDF using dvi2ps | ps2pdf
- Convert content (using data registry)
- QA results (using comparison services)
- Ingest results into repository (using adaptor)

1E

How might it work?

1E

Progress to date

Year 1:

- Established the project team and reached consensus about the project's goals and structure
- Moved from requirements gathering into specification and implementation

Year 2:

- Completed internal releases of all major software components
- Launched training courses for preservation planning modules
- Started to look at sustainability options

Year 3:

- Complete external releases of all major software components
- Increase level of integration
- Case studies (including at BL) with repository integration
- Conduct experiments on tools and methods
- Develop and launch sustainability plan

Comments from Year 2 External Review, July 2008

“Overall, the Review Panel remains impressed with the deliverables of the PLANETS project through this second review period including the clear commitment of the partners. The nature and extent of the project overall is articulated to a very high standard and the quality and clarity of the deliverables continues to be of a very high standard.

The panel is impressed with the progression in both depth and breadth of the overall programme from last year, especially with regard to the number and quality of the demonstrations, and the uniqueness of the research output from the XC*L activity is seen by the panel as a central plank of the PLANETS outcomes for validation/quality assurance.”

Planets Preservation Planning in detail

- PP work package is producing:
 - Assessment of existing PP practices
 - PP tool (Plato) to enable development/creation of plans
 - OAIS Extension to incorporate Planets PP model into OAIS

15

Planets Preservation Planning

A number of (partial) DP solutions

- different strategies
- different tools for each strategy
- different parameters for each tool

Which solution to select?

- to meet preservation requirements (what are these?)
- cost? stability? future use?

Hence: Digital preservation planning

20

Planets Preservation Planning: Goals

- motivate and allow holders of digital collections to precisely specify their preservation requirements
- provide structured model to describe and document these
- create defined setting to evaluate preservation strategies
- document outcome of evaluations to allow informed, accountable decision

21

Planets Preservation Planning: Previous Work

- Planets approach to Preservation Planning is based on DELOS DP Testbed
- DELOS Network of Excellence on Digital Libraries; Digital Preservation Cluster
- Joint efforts by
 - Dutch National Archives,
 - Vienna University of Technology
 - Austrian National Library
 - State and University Library Göttingen
 - The Austrian Audiovisual Research Archive
 - National Research Council (CNR), Pisa, Italy
 - and other Cluster 6 members

22

Planets Preservation Planning: Previous Work

DELOS Digital Preservation Testbed

- requires clear specification of all goals/requirements for preservation
- provides structured view of these objectives
- offers platform for repeatable, documented evaluation
- framework and methodology

It allows

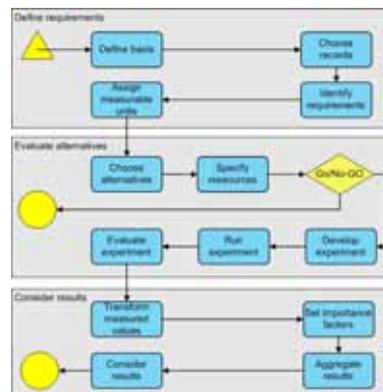
- evaluation against well defined criteria
- comparison of different preservation strategies

It is the basis for

- informed,
 - documented, and
 - accountable decisions
- on which preservation plan to adopt

26

Planets Preservation Planning: Workflow

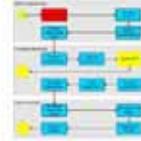


27

Planets Preservation Planning: Define Basis

Description of the setting

- Identify a consistent collection
- What is the institutional setting? Tasks? Obligations?
- What are the objects?
- Who are their users?
- Where does the preservation process take place?
- How many objects? How many new ones per year?
- How well-controlled is the environment?
- ...



21

Planets Preservation Planning: Choose objects/records

Select a small number of representative sample objects from the identified collection

Different object types

- Text documents, audio, video, e-mail, multimedia, databases, data sets, ...

Object characteristics

- different versions, embedded objects, relations between objects

Choice of objects affects the evaluation



21

Planets Preservation Planning: Identify objectives -> Objective Tree

Define all relevant goals and characteristics (high-level, detail) with respect to a given application domain



What are the requirements/goals

- Essential characteristics of objects
- Content, context, structure, form and behaviour
- Authenticity, reliability, integrity, usability, costs

Put the objectives in relation to each other (hierarchical)

Objective tree approaches:

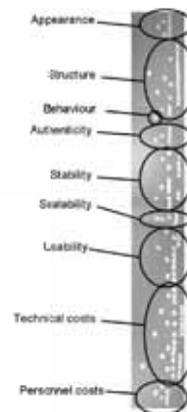
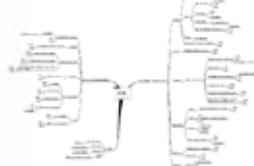
- bottom-up
- top-down

21

Planets Preservation Planning: Identify objectives -> Objective Tree

Four major groups

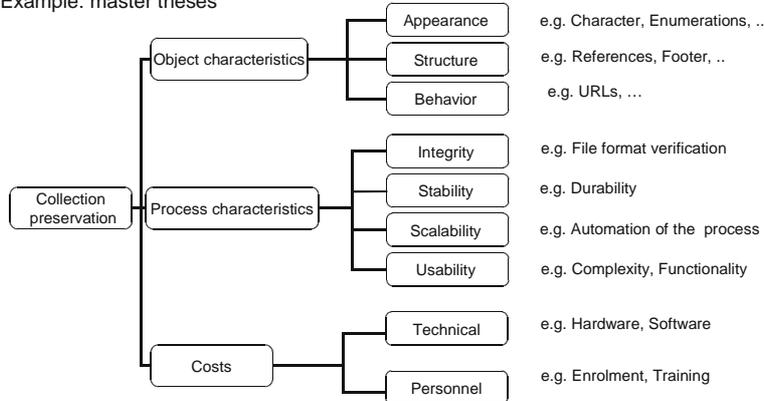
- object characteristics (content, metadata ...)
- record characteristics (context, relations, ...)
- process characteristics (scalability, error detection, ...)
- costs (set-up, per object. HW/SW. personnel, ...)



21

Planets Preservation Planning: Identify objectives -> Objective Tree

Example: master theses



25

Planets Preservation Planning: Assign measurable units

Assign measurable effect to each leaf

- ensure that leaf criteria are objectively (and automatically) measurable
 - seconds/Euro per object
 - bits color depth
 - ...
- subjective scales where necessary
 - diffusion of file format
 - amount of (expected) support
 - ...

No limitations on the use of scale



31

Planets Preservation Planning: Choose alternatives

Given the type of object and requirements what strategies would be best: suitable/are possible?

- a (set of) migration tool(s)
- a (set of) emulation tool(s)
- both
- other?

For each alternative strategy precise definition of

- Which tool (OS, version,...)
- Which functions of the tool in which order
- Which parameters



31

Planets Preservation Planning: Specify resources

Detailed design and overview of the resources

- human resources (qualification, roles, responsibility, ...)
- technical requirements (hardware and software components)
- time (time to run experiment,...)
- cost (costs of the experiments,...)



32

Planets Preservation Planning: Go/No-Go

deliberate step for taking a decision whether it will be useful and cost-effective to continue the procedure, given

- the resources to be spent (people, money)
- the expected results

review of the experiment/ evaluation process design so far

- e.g. is the design correct and optimal?
- is the design complete (given the objectives)

go / no-go / deferred-go (condition)



31

Planets Preservation Planning: Develop experiment

Formulate for each experiment in detail

- Development plan (includes steps to build and test software components, definition of procedures and preparation of testbed environment)
- Test plan (mechanisms how to)
- Evaluation/experiment plan (workflow/sequence of activities)



31

Planets Preservation Planning: Run experiment

Run experiment with the previously defined sample records and specified tools

The results will be evaluated in the next stage



31

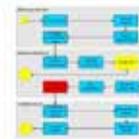
Planets Preservation Planning: Evaluate experiment

Evaluate how successfully the requirements are met

Measure performance with respect to leaf criteria in the objective tree

The evaluation will also identify

- need for repeating the process
- unexpected results (or undesired)



31

Planets Preservation Planning: Transform measured values

Measures come in seconds, euro, bits, goodness values,...

Need to make them comparable

Transform measured values to uniform scale

Transformation tables for each leaf criterion

Linear transformation, logarithmic, special scale

e.g. scale 1-5 plus "not-acceptable"



31

Planets Preservation Planning: Set importance factors

Definition which criteria are more important

Depends on individual preferences and requirements

Adaptation for each implementation

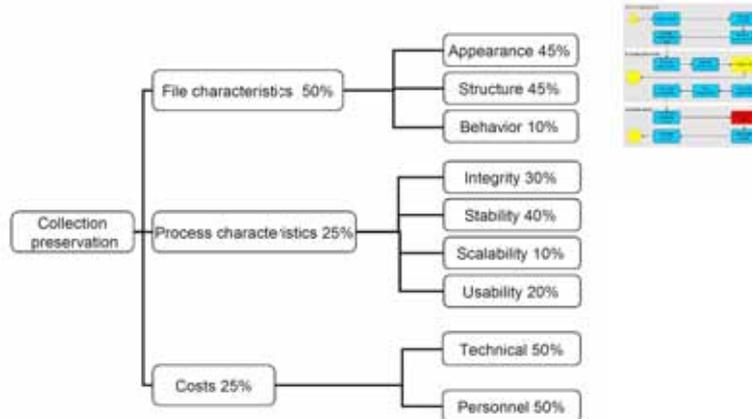
Influence on the final ranking

Aggregation of weights



31

Planets Preservation Planning: Set importance factors



31

Planets Preservation Planning: Aggregate values

Multiply the transformed measured values in the leaf nodes with the leaf weights

Sum up the transformed weighted values over all branches of the tree

Creates performance values for each alternative on each of the sub-criteria identified



41

Planets Preservation Planning: Aggregate values



Planets Preservation Planning: Aggregate values

