iPRES 2007, 11-12 October, Beijing, China

A practical approach to digital preservation: updates from Planets

Helen Hockx-Yu, Adam Farquhar
The British Library

Outline of presentation

- Brief introduction to Planets
- Motivations for involvement
- Key components of Planets architecture
- How does Planets relate to OAIS?
- Typical preservation scenarios using Planets tools and services
- Progress to date
- What's next?

Plants overview

- A 4-year research and technology development project co-funded by the European Union to address core digital preservation challenges.
- CASPAR and DPE
- Started June 2006 with €15m budget
- Coordinated by the British Library
- Involves 16 partners including national libraries and archives, leading technology companies and research universities
- Builds on strong digital archiving and preservation programmes
- Focuses on the needs of libraries and archives

Aims and objectives

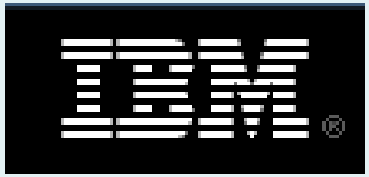
- **Increase Europe's ability to ensure long-term access to its cultural and scientific heritage**
  - Improve decision-making about long term preservation
  - Ensure long-term access to valued digital content
  - Control the costs of preservation actions through increased automation, scaleable infrastructure
  - Ensure wide adoption across the user community and establish market place for preservation services and tools
- **Build practical solutions**
  - Integrate existing expertise, designs and tools
  - Deliver tools and services that can be used in an operational environment

Planets partners

- 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

Planets partners



- Tessella Plc
- IBM Netherlands
- Microsoft Research
- Austrian Research Centers GmbH

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

The Planets team

All Staff Meeting, Feb 2007

Motivations

- For national libraries & archives
  - Have the legal responsibility and the legislative framework to safeguard digital information
  - Have been collecting digital documents and records since 1982
  - Realise 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 & ICT is a must
  - Need pragmatic solutions here and now
- Preservation and access over the long term is their **primary mission**
- A solution that fails for content holders fails for everyone

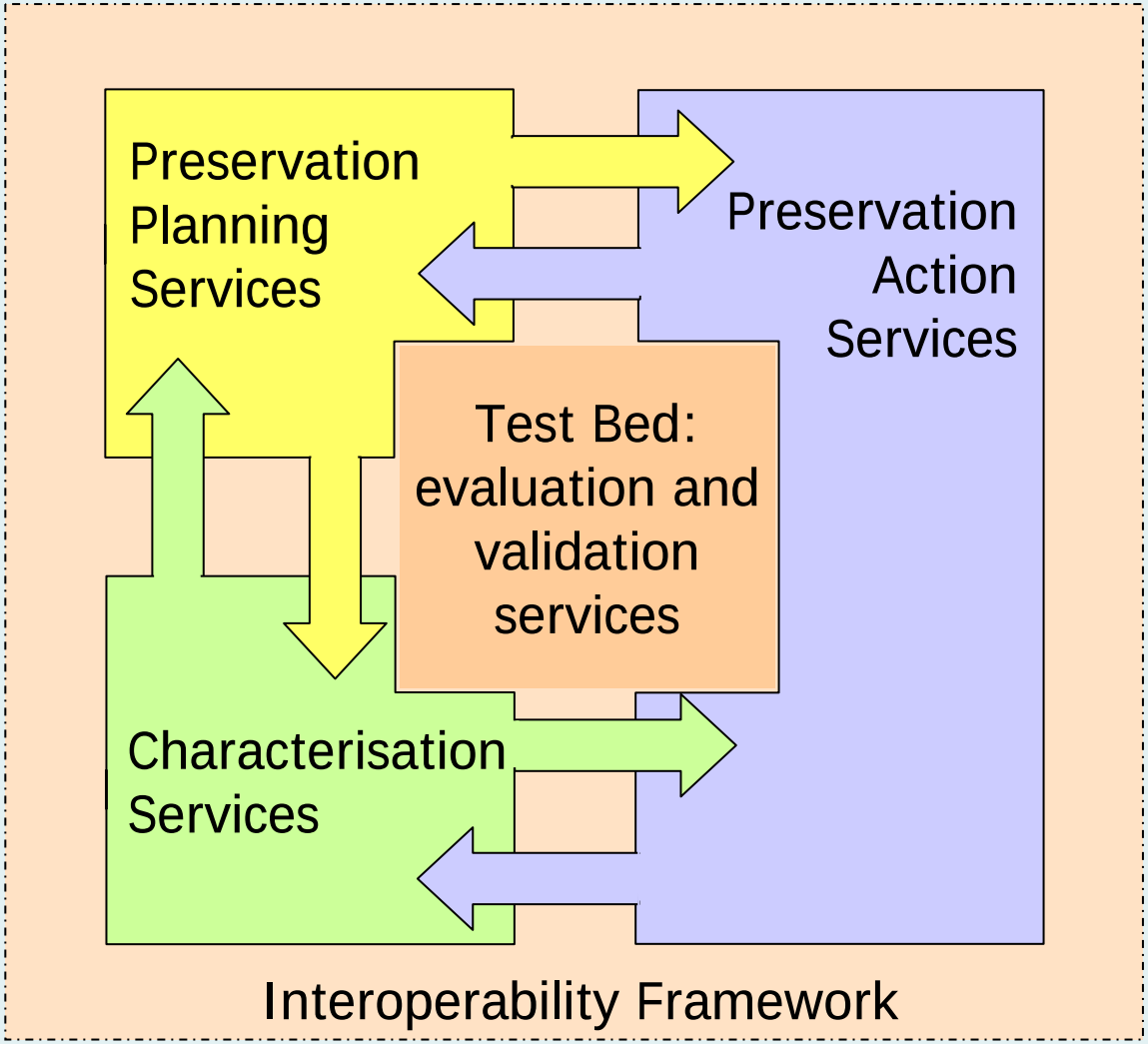
Motivations

- For researchers
  - Complex cross-disciplinary issues
  - Fundamental frameworks still unclear
  - Huge potential impact for a broad range of society
- For technology companies
  - Different types: Content creation application vendors; System integrators; Product vendors
  - Opportunity to introduce innovative services and products
  - Opportunity to increase competitiveness
  - The market is emerging – personal and corporate
  - Few vendors with the capability

What's in it for the British Library?

- "Planets will provide **the** technology component of our digital preservation solution"
  Richard Boulderstone, BL Director, 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

Planets Architecture



Preservation Planning in OAIS

- OAIS breaks preservation planning into 4 functions:
  - Monitor designated community
  - Monitor technology
  - Develop preservation strategy and standards
  - Develop packaging designs and migration plans
- Includes only high-level descriptions; no details for practical implementation
- Important preservation functions seem to be implicit or missing in OAIS

Preservation Planning in Planets

- High level functions: Preservation Watch, Preservation Planning, Action and Characterisation
- Broadly map to the OAIS preservation planning functions but also provide added functionality and practical implementation
- Planets focus on preservation of digital content
  - OAIS also deals with repository longevity
- Planets does not yet fully address packaging design
- Planets experience may lead to refinement / extension of the OAIS reference model

Progress to date

- Established the project team and reached consensus about the project's goals and structure
- Moved from requirements gathering into specification and implementation
- Developed prototype tools and services for preservation planning, preservation action and preservation characterisation
- Released first prototypes of the Interoperability Framework (IF) and the Testbed
- Started deploying tools and services in the IF and the Testbed

What's next (by end 2008)?

- Preservation Planning tools (PLATO®) including decision support and risk assessment modules
- Integrated preservation planning services including an automated collection profiling service, a technology watch service, and an advice service
- A description language for preservation action tools
- Planets-compliant migration tools for digital objects
- Emulation tools for specific environments
- Final specifications of a characterisation description and extraction language

What's next (by end 2008)?

- Characterisation tools which extract significant properties from digital objects
- Characterisation and preservation action tool registries
- A Testbed offering preservation plan assessment service to organisations outside Planets
- Planets Interoperability Framework as downloadable "click-and-install" software package
- A dissemination and take-up programme including workshops and training events to engage with suppliers and the library and archive communities

Scenario 1: Donation

- A scientist donates her research repository
  - Stretching back thirty years
  - Papers, technical reports, notes in many formats
  - Original research data
  - Software tools that implement research ideas
- Many possible uses
  - A university IP officer wants to defend a patent challenge
  - A biographer wants review the unpublished work
  - A former student wants to revive a line of research

Scenario 1

- Step 1: Ingest original contributions into repository
  - Using repository capabilities
- Step 2: Characterise objects according to policies
  - Using XCDL/XCEL, policy language
- Step 3: Convert undesirable objects into desirable forms
  - Using PLATO to build preservation plans
  - Using action registry to identify conversion services
  - Using testbed experimental data to inform selection
  - Using workflow, data registry, run services to convert objects
- Step 4: Perform automated QA on results
  - Using XCDL/XCEL, comparison services
  - Address residual problems
- Step 5: Ingest conversions into repository
  - Using repository adaptor
  - Record relationship to originals, workflow details

Scenario 2: Changes in user community

- Sample policy: 90% of users can access all published reports
- Usage profile: 98% of users can not view dvi files
- Content profile: 5% of published reports in dvi format
- Identify possible plans (using PLATO) including
  - Convert to PDF
  - Convert to tiff
  - Provide users with viewer plug-in
  - Provide on-the-fly conversion to PDF
- 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)

Conclusion

- Planets methods, tools, and services will help organisations diagnose and treat problems with their digital objects
- High levels of automation and scalable components will reduce costs and improve quality
- Empirical data will enable improved decision making
- Find out more: http://www.planets-project.eu