Planets: Integrated Digital Preservation Services
Planets overview

- 4-year project co-funded by the European Union
- Under the same programme as CASPAR and DPE
- Started in June 2006 with €15m budget
- Coordinated by the British Library
- 16 partners including national libraries and archives, technology companies and research universities
- Focuses on the needs of libraries and archives
Aims and objectives

- Help ensure long-term access to cultural and scientific heritage
  - Improve decision-making
  - Ensure long-term access
  - Control costs
  - Ensure wide adoption across user community
  - 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
Planets architecture

Interoperability Framework

Preservation Planning Services

Preservation Action Services

Test Bed: evaluation and validation services

Characterisation Services

Digital Content

Org. Context

External Context

Technical Environment
Planets architecture: key components

- **Preservation planning**
  - tools and services for formulating and selecting preservation plans
- **Preservation characterisation**
  - tools and services for automatic analysis of digital objects’ technical and intellectual characteristics
  - supporting registry of characterisation information
- **Preservation action**
  - methodology for describing preservation action tools
  - supporting registry
  - migration and emulation tools
- **Testbed**
  - hardware and software environment for comparing digital preservation tools and assessing their effectiveness
- **Interoperability framework**
  - service-oriented architecture
  - provides shared functions
  - integrates the Planets tools and services
Preservation planning

- Preservation Policy
- Content Profile
- Usage Profile
- Actions

Preservation Planner → Plans → Plan Evaluator

Sample Content
Planning for a desired future

- Parameters for successful preservation strategy
  - Organisational policy
  - Usage pattern of digital collection
- Methodology and workflow for identifying preservation requirements, defining potential solutions and selecting the most appropriate
  - Plato, workflow-based decision support tool
- Validation framework for measuring and comparing characteristics (in a typical before and after scenario)
- Proactive preservation planning including collection profiling, technology watch and advice services
- Implementing the OAIS Preservation Planning function
Preservation Planning using Plato

Planets Interoperability framework
## Planets Preservation Planning Tool (Plato)

### Identify Requirements

Expand All | Collapse All

**Website**
- **Record characteristics**
  - **Appearance**
  - **Content**
  - **Structure**
  - **Behaviour**
    - **Deactivate**
      - **Mailto:**
      - **Preserve**
        - **Menus**
        - **Pop-ups**
      - **Freeze**
        - **Current date/time**
        - **Visitor counter**
        - **Newsfeeds**
      - **Context**
        - **Technical characteristics**
          - **Ubiquity**
          - **Tool Support**
          - **Documentation**
            - **Quality**
            - **Disclosure**
            - **Openness**

### Table

<table>
<thead>
<tr>
<th>Focus</th>
<th>Node</th>
<th>Single</th>
<th>Scale</th>
<th>Restriction</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Website</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Record characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Appearance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Behaviour</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deactivate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mailto:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Preserve</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Menus</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pop-ups</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Freeze</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Current date/time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Visitor counter</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Newsfeeds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2007 Institute of Software Technology and Interactive Systems: office bears
### Analyse Results

<table>
<thead>
<tr>
<th>Sum</th>
<th>PDF/A (Tool A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PDF/A (Tool B)</td>
</tr>
</tbody>
</table>

### Minimalist root node

<table>
<thead>
<tr>
<th>Focus</th>
<th>Name</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimalist root node</td>
<td>PDF/A (Tool A): 2.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDF/A (Tool B): 3.19</td>
</tr>
<tr>
<td>X</td>
<td>Image properties</td>
<td>PDF/A (Tool A): 0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDF/A (Tool B): 0.60</td>
</tr>
<tr>
<td>X</td>
<td>Karma</td>
<td>PDF/A (Tool A): 0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDF/A (Tool B): 0.00</td>
</tr>
<tr>
<td>X</td>
<td>Filesize (in Relation to Original)</td>
<td>PDF/A (Tool A): 0.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDF/A (Tool B): 0.99</td>
</tr>
<tr>
<td>X</td>
<td>A Single-Leaf</td>
<td>PDF/A (Tool A): 0.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDF/A (Tool B): 0.80</td>
</tr>
<tr>
<td>X</td>
<td>IntRange 0-10</td>
<td>PDF/A (Tool A): 0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PDF/A (Tool B): 0.60</td>
</tr>
</tbody>
</table>
Content characterisation

- Automating the process of identifying, describing and extracting the essential characteristics of digital objects
- Supporting registry with technical information about significant properties of digital object types
- Generic XML language supporting automatic validation of document conversions and evaluation of migration quality
  - *eXtensible characterisation definition language (XCDL)* and *eXtensible characterisation extraction language (XCEL)*
  - Hierarchically decompose and represent documents in an abstract XML language
- Extractor and Comparator software
XCEL extractor

XCEL processing software which extracts content of a file and express it in XCDL
XCDL comparator

Identify degrees of equality between XCDL documents

ODT

Migration

PDF

Extractor

ODT XCEL PDF XCEL

? XCEL ? XCEL

ODT XCDL

Comparer

PDF XCDL

93%
Preservation action

- Inventory of commonly used file formats
- Gap analysis of current tools
- Formal language for describing preservation action tools
- Developing the most wanted tools for digital objects
- Supporting registry with technical descriptions of tools
- Migration tools to convert digital objects into newer formats
- Emulation tools enabling interaction with digital objects using original software applications
  - Dioscuri – the modular emulator
  - Universal Virtual Computer (UVC)
Dioscuri – the modular emulator
The Universal Virtual Machine (UVC)
Planets Testbed: learn what works

- “A controlled environment with metrics and benchmark content that allows the objective comparison of preservation tools and strategies through experimentation”
- A dedicated scientific research environment
- Consists of
  - Data storage, hardware, Planets software, Testbed application
  - Benchmark and other content
- Use well known and widely spread technologies
- Available for external organisations in early 2009
Testbed experiment process

Design Experiment
- Step 1: Define Basic Properties
- Step 2: Design Experiment
- Step 3: Specify Resources

Run Experiment
- Step 4: Go / No go decision
- Step 5: Run Experiment

Evaluate Experiment
- Step 6: Evaluate Experiment
Prototype Testbed
Planets Software: Vision

- A single downloadable software package
- Simple to install, configure and administer
- When deployed, a Planets instance is created, in which
  - an administrator can
    - create user accounts
    - deploy and browse services
    - browse registries
  - a preservation expert can
    - define service workflows (Workflow Design Tool)
    - define and evaluate preservation plans (Preservation Planning Application)
    - define and run experiments (Testbed)
  - a librarian or archivist can
    - define and test preservation plans
    - execute preservation processes on a repository (Online Design Tool)
Planets Software: motivation

- Provide commonly required functions
  - Data persistence
  - User management
  - Authentication and Authorisation
  - Monitoring, Logging, and Messaging

- Meet non-functional requirements on the infrastructure
  - Robust
  - Scalable
  - Distributed
Planets software: benefits

- **Efficiency**
  - Components can be shared and only need to be developed once
  - Optimise the number of components: IF provides a single database for all components, so only one database need be installed

- **Interoperability**
  - Common components help ensure interoperability between various applications
  - Support access to remote and distributed 3rd party characterisation and migration services by enforcing Web Service standards
Planets software: architecture

Planets Interoperability Framework

PLANETS Applications
- Workflow Designer
- Administration Tool
- Testbed
- Preservation Planner
- Characterization Tools

Internal Repositories
- Planets/Service
- External Services

External Repositories

Core Elements
- Security/Authentication Authorization
- Monitoring/Logging/Auditing
- Workflow Execution Engine
- Transaction Manager
- Work Space
- Error/Exception Handling

Service Bus
- Service Registry
- Registry Services

External Registries
Interoperability Framework components

- Application server (JBoss)
- Relational Database (Apache Derby)
- Service Registry (jUDDI)
- Data Registry (JCR implementation on Jackrabbit)
- Workflow Design Tool
  - Workflow expressed in BPEL
  - Expert design tool based on the Eclipse BPEL Plugin
Single sign-on service

Initial login directed to the Single Sign On Service
Administration interface

Main administration interface linking to various applications...
Planets Web Service registry

Admin interface linking to Planets Web Service registry, where services can be manually registered and searched
Workflow design tool
Workflow design tool

Allowing graphically orchestration of complex workflows, including loops and branches, and production of deployable BPEL code.
Workflow execution

Executing service request...
Workflow execution

The completed Workflow provides an XML response, including a link to a report page generated by the Workflow.
Workflow execution

Results of the workflow for each submitted file. Migrated files are written to the data registry.
Data registry

It is possible to directly browse the node structure of the JCR-based Data Repository.
One can also execute Xpath or standard SQL queries on the repository
Data registry

... and observe the content (in this case a TIFF image migrated to the JPEG format)
Data registry
Conclusion

- Planets will help organisations diagnose and treat problems with their digital collections
- High levels of automation and scalable components will reduce costs and improve quality
- Empirical data will enable improved decision making
Find out more about Planets

- **Contact us:** info@planets-project.eu
- **Website:** www.planets-project.eu
- **Planets news:** http://www.planets-project.eu/news/
- **Workshops, conferences and training events:** http://www.planets-project.eu/events/
- **Quarterly newsletters:**
  http://www.planets-project.eu/publications/?search[0]=10
- **Scientific publications:**
  http://www.planets-project.eu/publications/?search[0]=12