Planets overview

- 4 year project, started in 2006
- Brings together libraries, archives, research institutions and technology providers
- Developing a scaleable digital preservation infrastructure
- Ensuring wide adoption across user communities and establishing a market place for preservation services and tools

www.planets-project.eu
Planets architecture

- Preservation Planning Services
- Characterisation Services
- Registry Services
- Preservation Action Services
- Interoperability Framework
Characterisation sub-project

• To understand the properties of digital objects which are significant to their long-term accessibility
• To develop methods for expressing and measuring these properties to:
  • inform preservation planning
  • validate the results of preservation actions
  • support the needs of user communities
• To promote the development of object types with ‘preservation-friendly’ characteristics

[European Union flag] [Planets logo]
Areas of work

• Characterisation strategies and methodologies
• Characterisation tools
  – Measure properties of objects
• Characterisation registry
  – Describes tools and their capabilities
  – Describes object types and properties
Characterisation processes

- Identification
  - *The object is in PNG 1.0 format*
- Validation
  - *The object is valid against the PNG 1.0 specification*
- Property extraction
  - *The object uses 24-bit colour and has a resolution of 300 dpi*
Models

Source $\xrightarrow{\text{interpreted using}}$ Process $\xrightarrow{\text{yields}}$ Performance

(National Archives of Australia)

Data object $\xrightarrow{\text{interpreted using}}$ Representation information $\xrightarrow{\text{yields}}$ Information object

(OAIS)

Representation properties $\xrightarrow{\text{significant properties}}$

EU flag and Planets logo
Models

• Representation properties:
  – *Describe the data object and its representation network*

• Significant properties:
  – *Describe the information object*
Planets conceptual model
Representation properties

• Identified through characterisation
• Described with reference to the characterisation registry
• Use cases:
  – Recursive selection of characterisation tools
  – Populating decision trees for preservation planning using PLATO
White Paper

- Discusses concept of Representation Information Registries
- Surveys current state of the art
- Discusses roles of registries in Planets
- Makes recommendations for future research

www.planets-project.eu/docs/reports/Planets_PC3-D7_RepInformationRegistries.pdf
Significant properties

- Identified through characterisation
- Described with reference to the characterisation registry
- Use cases:
  - Populating decision trees for preservation planning using PLATO
  - Validation of preservation actions
Validation
Categories of significant property

- Content
- Context
- Structure
- Appearance
- Behaviour
Details for: Portable Document Format 1.6

Summary

Name: Portable Document Format
Version: 1.6
Other names: PDF (1.6)
Identifiers: PLID: f1nt20
MIME: application/pdf

Family

Classification: Page Description
Disclosure: Full

Description: Portable Document Format is a platform-independent format for representing formatted documents, developed by Adobe Systems Incorporated. It is the native format of Adobe's Acrobat family of software products, version 1.6 corresponding to the release of Acrobat 7.0. PDF is based on, and shares the same imaging model as, the PostScript page description language. A PDF file comprises a Header section, a Body section containing the objects which make up the document, a Cross Reference Table, and a Trailer section. PDF files can contain a wide variety of content, including text, images, video and audio.

Orientation: Binary
Byte order: Big-endian (Motorola)
eXtensible Characterisation Languages

- Description Language (XCDL)
  - Expresses the properties of an object in neutral form
- Extraction Language (XCEL)
  - Expresses how to extract those properties for a given object format
- XCDL extraction tool
Third-party tools
# Active Preservation Management Console

**Command:** CHTP  
**Machine:** potent.essella.co.uk  
**Date Received:** 2006-Dec-16 04:50:12  
**Working Area:** /home/carj/tmp/F/  
**Date Completed:** /home/carj/tmp/F/  
**Client Web Service:** http://localhost:8080/jcrQueueManager/services/JobCompleteService

<table>
<thead>
<tr>
<th>Name</th>
<th>Tool Id</th>
<th>Date Started</th>
<th>Date Completed</th>
<th>Priority</th>
<th>% Completed</th>
<th>Status</th>
<th>Action</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterize File Set</td>
<td>100</td>
<td>2006-Dec-15 04:51:24</td>
<td></td>
<td>Normal Priority</td>
<td>0</td>
<td>Running</td>
<td>Characterise File Set started...</td>
<td>Tool Completed Normally</td>
</tr>
<tr>
<td>Jhove-Tiff</td>
<td>100</td>
<td>2006-Dec-15 04:51:54</td>
<td>2006-Dec-15 04:51:54</td>
<td>Normal Priority</td>
<td>100</td>
<td>Completed</td>
<td>Tool Completed Normally</td>
<td></td>
</tr>
<tr>
<td>Jhove-Wave</td>
<td>100</td>
<td>2006-Dec-15 04:51:54</td>
<td>2006-Dec-15 04:51:54</td>
<td>Normal Priority</td>
<td>100</td>
<td>Completed</td>
<td>Tool Completed Normally</td>
<td></td>
</tr>
</tbody>
</table>
Questions?