Characterisation in Planets

Adrian Brown The National Archives, UK





Planets overview

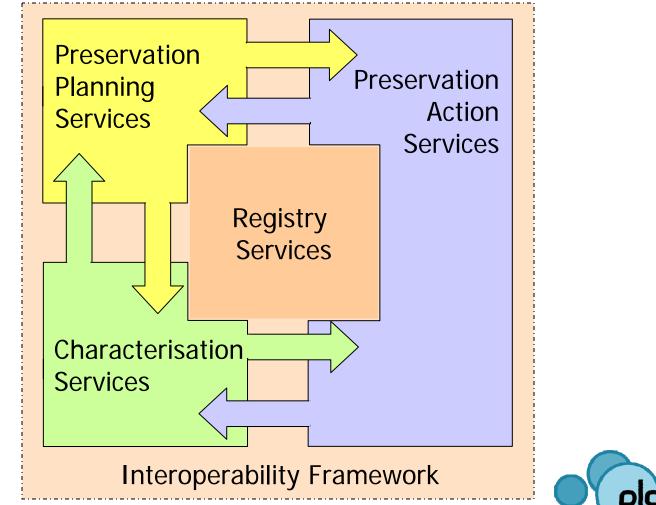
- 4 year project, started in 2006
- Brings together libraries, archives, research institutions and technology providers
- Developing a scaleable digital preservation infrastructure
- Ensuring wide adoption across user communities and establishing a market place for preservation services and tools

www.planets-project.eu





Planets architecture







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





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





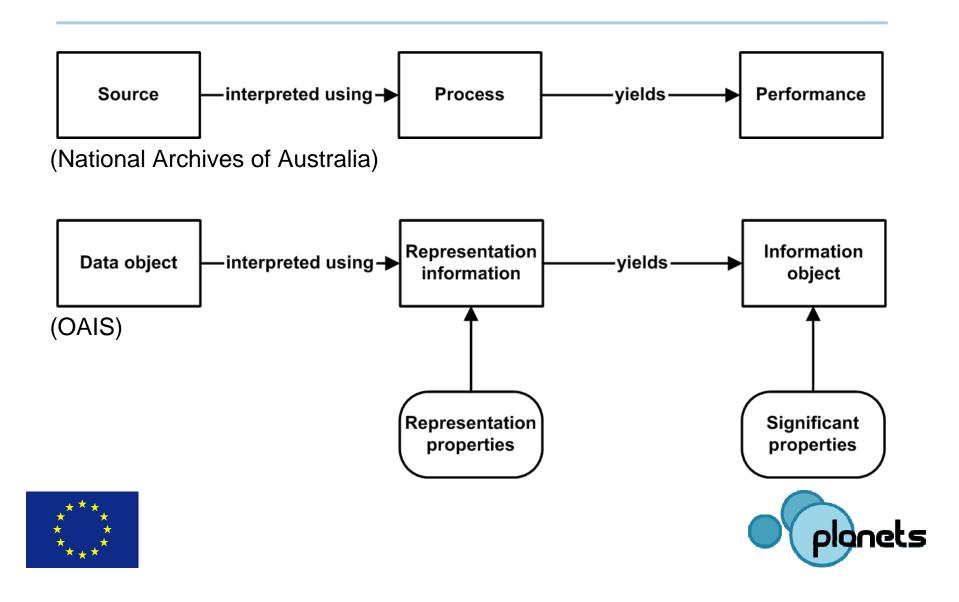
Characterisation processes

- 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



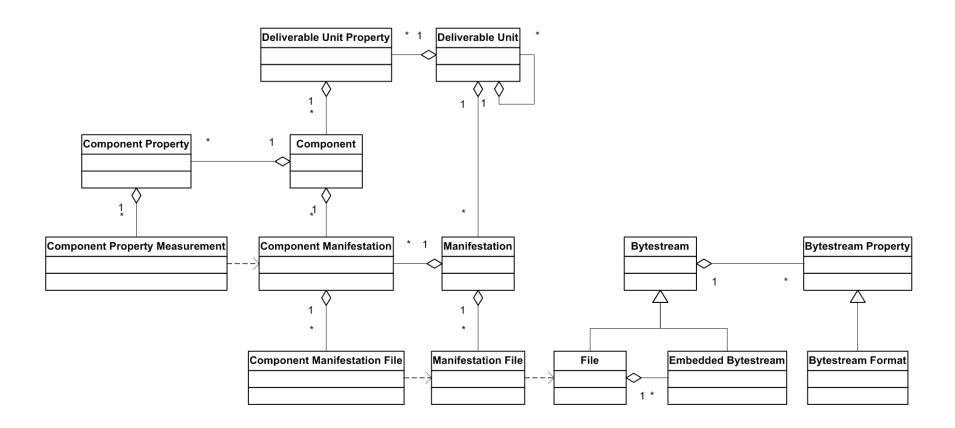
Models

- 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





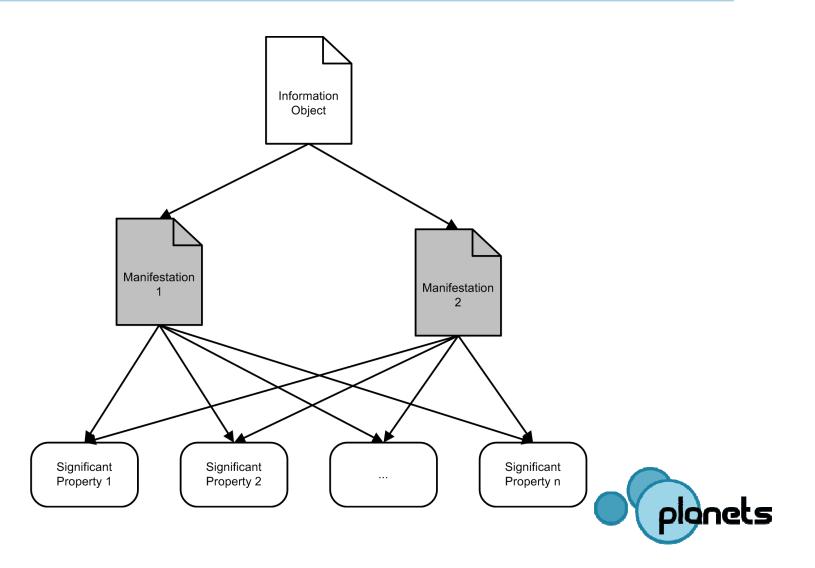
Significant properties

- 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







? Help : detailed report on file format

Q Details: File format summary

format PRONO	M Unique Identifier Software Vendor Lifecycles
Details for: Po	ertable Document Format 1.6 🛛 🕒 Save as XML CSV 🖉 Print
Go to: summary	Documentation > Signatures > Compression > Character encoding > Rights > Reference files >
Summa ry	
Name	Portable Document Format
Version	1.6
Other names	PDF (1.6)
Identifiers	PUID: fmt/20 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)

www.nationalarchives.gov.uk/pronom





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





XCDL extractor

output/basi2c16.png.xcdl outp	put/ArcTriomphe-cHRM-red-	green-swap.png.xco	di output/Ar	rcTriomphe-c	olortest-goo	d.png.xcdl	output/	asi0g02.png.×	cd outp	ut/basi0g04.png.xcd	dl output/basi0g	g08.png.xcdl	output/ba 🔨	>
xml version='1.</th <th>0' encoding='</th> <th>UTF-8*?></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>11</th> <th></th> <th></th> <th>~</th>	0' encoding='	UTF-8*?>									11			~
<xcdl <="" td="" xmlns:xsi="</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>.eu/xcl/s</td><td>chemas/x</td><td>cl"><td></td></xcdl>														
xsi:schemaLocatio <object 1d="</td" id="o</td><td></td><td>.planets-p</td><td>roject.</td><td>eu/xcl</td><td>/schem</td><td>as/xc.</td><td>L XCDL</td><td>Core.xs</td><td>d"><td>"U" ></td><td></td><td></td><td></td><td></td></object>	"U" >													
	id="nd1" >ff	ff ff ff	00 00 f	7 bd f	f ff O	0 00 4	ef 7b	ff ff O	0 00 e	7 39 ff ff	00 00			
de f7 ff ff 00 00														
bd ef ff ff 00 00														
9c e7 ff ff 00 00														
7b de ff ff 00 00 5a d6 ff ff 00 00														
39 ce ff ff 00 00														
18 c6 ff ff 00 00	10 84 ff ff	00 00 08 4	2 ff ff	00 00	00 00	ff f	E 00 0	0						
ff ff f7 bd 00 00														
de f7 f7 bd 00 00 bd ef f7 bd 00 00														
9c e7 f7 bd 00 00														
7b de f7 bd 00 00														
5a d6 f7 bd 00 00														
39 ce f7 bd 00 00 18 c6 f7 bd 00 00														
ff ff ef 7b 00 00														
de f7 ef 7b 00 00														
bd ef ef 7b 00 00														
9c e7 ef 7b 00 00 7b de ef 7b 00 00														
5a d6 ef 7b 00 00														
39 ce ef 7b 00 00														
18 c6 ef 7b 00 00														
ff ff e7 39 00 00 de f7 e7 39 00 00														
bd ef e7 39 00 00														
9c e7 e7 39 00 00														
7b de e7 39 00 00														20
5a d6 e7 39 NN NN <	52 94 e7 39	IIII IIII 4A 5	2 e7 39		42 11	e/ .1	ə IIII II	11						>
	1			<u> </u>										
1 basi2c16.png						×	TEL							
2 ArcTriomphe-cHRM-red-green-sw	ap.png					C:,	Dokumente	und Einstellung	jen/j/Desktoj	p/PlanetsReview2007	7/extractor/xcel/xc	el_docs/xcel_pn	ig.×ml	····
3 ArcTriomphe-colortest-good.png														
4 basi0g02.png														
5 basi0g04.png														GO
6 basi0g08.png														
			Ex	Microsoft P			Reader	_		Adobe Photo		esktop » phd		» 📢 🗊 12:38





Third-party tools

	🖗 DROID (Digital Record Object Identification)	
👙 Jhove	Eile Edit Identify Tools Help	
File Edit Help		110 ASCI
	PUID Format Version Status Warning fmt/44 JPEG File Interchange Format 1.02 Positive (Specific Format) Version Version	
**** * ***		plane

	You are not logged in
Active Preservation Management Console	Username login
	Password

Command	CHR3	XMLOutputFile /home/carj/tmp/F/chr-output.xml
Machine	potent.tessella.co.uk	XMLInputFile /home/carj/droid-input.xml
Date Received	2006-Deo-15 04:50:12	
Working Area	/home/carj/tmp/F/	
Date Completed		
Client Web Service	http://localhost:8080/jobQueueManager/services/JobCompleteService	

Name	Tool Id	Date Started	Date Completed	Priority	% Completed	Status	Acti	оп	Message
Characterise File Set	100	2006-Deo-15 04:51:24		Normal Priority 💌	0	Running		×	Characterise File Set started
DROID	100	2006-Deo-15 04:51:27	2006-Deo 15 04:51:52	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Jpeg2000	100	2006-Deo-15 04:51:53	2006-Deo-15 04:51:53	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Gif	100	2006-Deo 15 04:51:53	2006-Deo-15 04:51:53	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Html	100	2006-Deo-15 04:51:53	2006-Deo-15 04:51:54	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Tiff	100	2006-Deo 15 04:51:54	2006-Deo 15 04:51:54	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Wave	100	2006-Deo 15 04:51:54	2006-Deo 15 04:51:54	Normal Priority	100	Completed			Tool Completed Normally
XML validatior	100	2006-Deo 15 04:51:54	2006-Deo 15 04:51:55	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Aiff	100	2006-Deo 15 04:51:54	2006-Deo 15 04:51:55	Normal Priority	100	Completed			Tool Completed Normally
Jhove-Pdf	100	2006-Deo-15 04:51:55		Normal Priority 💌	50	Running		X	ZA544606.800-UK-A.pdf (8/16)
Jhove-Jpeg	100	2006-Deo-15 04:51:57	2006-Dec-15 04:51:57	Normal Priority	100	Completed			

Questions?



