



# **The Digital Preservation Scenario**

**Vittore Casarosa  
HATII at the University of Glasgow**

**Digital Preservation – The Planets Way  
Copenhagen, 22 – 24 June 2009**

# Outline of presentation

---

- ❑ Overview of Day 2 and Day 3 program
- ❑ The “real” sample collection
- ❑ The real “user needs”



# Program of Day 2 and Day 3

---

- ❑ Preservation planning with Planets
  - Exercise: define and discuss objective tree
- ❑ Characterisation of digital documents
  - Demonstration
- ❑ Preservation actions (emulation)
  - Demonstration
- ❑ The Testbed environment
  - Demonstration
- ❑ Finalising a Preservation Plan
  - Completion of first exercise
- ❑ Experiencing the Testbed environment
  - Exercise: experiment with different scenarios (migration)



# Preservation strategies

---

## ❑ Digital preservation approaches

- techno-centric: keeping and maintaining the object's original hardware and software (technology preservation)
- data-centric: maintaining objects in current data formats and ensuring that the HW and SW needed to render the object can be “emulated” in the future technologies (technology emulation)
- process-centric: migrating the digital information so that it can be (faithfully) rendered by the new processes that will evolve from the present ones (information migration)
- post-hoc: digital archaeology or forensics (rescue)



# OAIS conceptual model

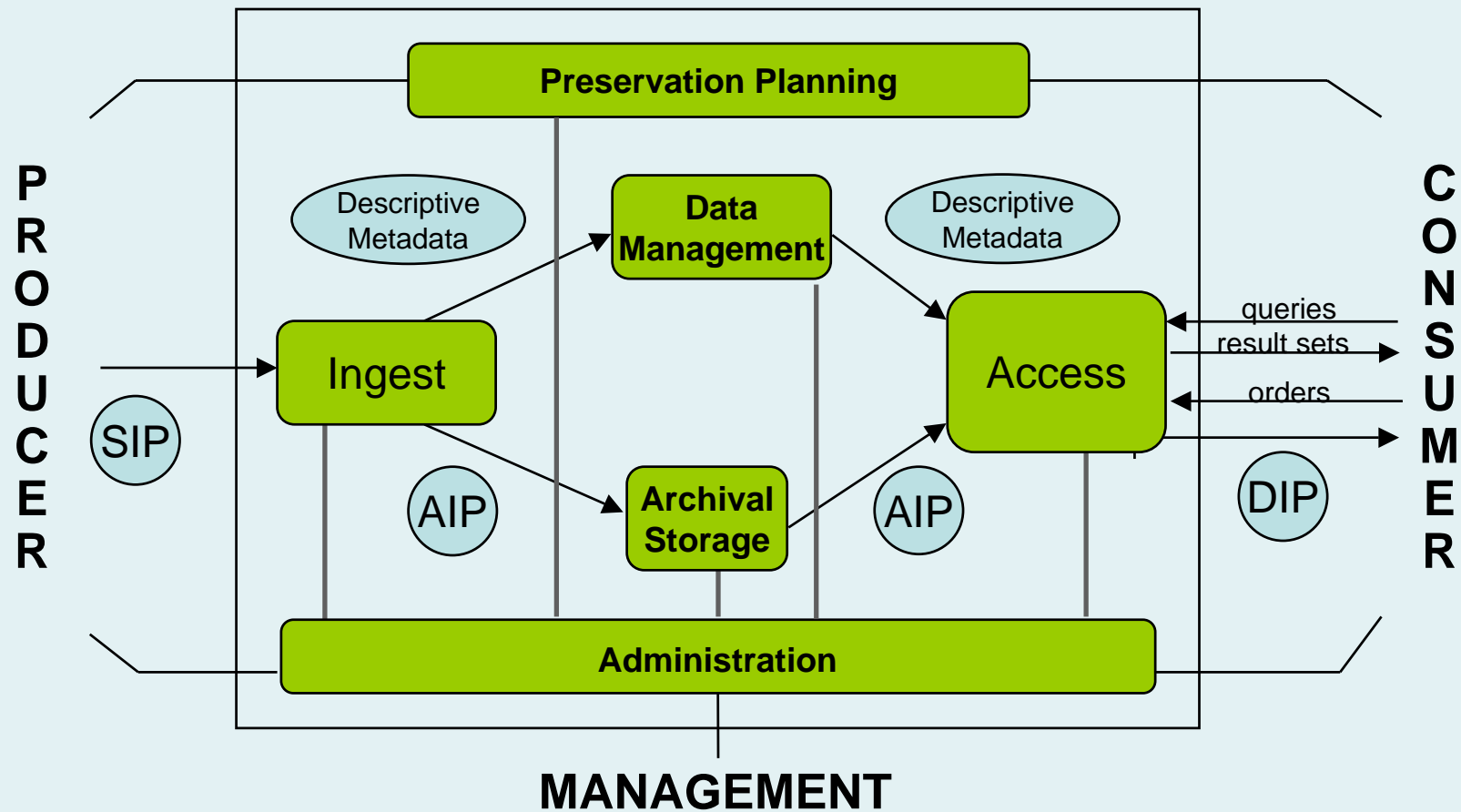
---

## Open Archival Information System

- ❑ The term 'open' means that the document was developed in an open way, and does not imply that access to any OAIS should be unrestricted
- ❑ An archive is defined as an "organization that intends to preserve information for access and use by a ***“designated community”***"
- ❑ While an OAIS itself need not be permanent, the information being maintained has been deemed to need "Long Term Preservation"
  - Long term = long enough for there to be a concern about the impact of changing technologies



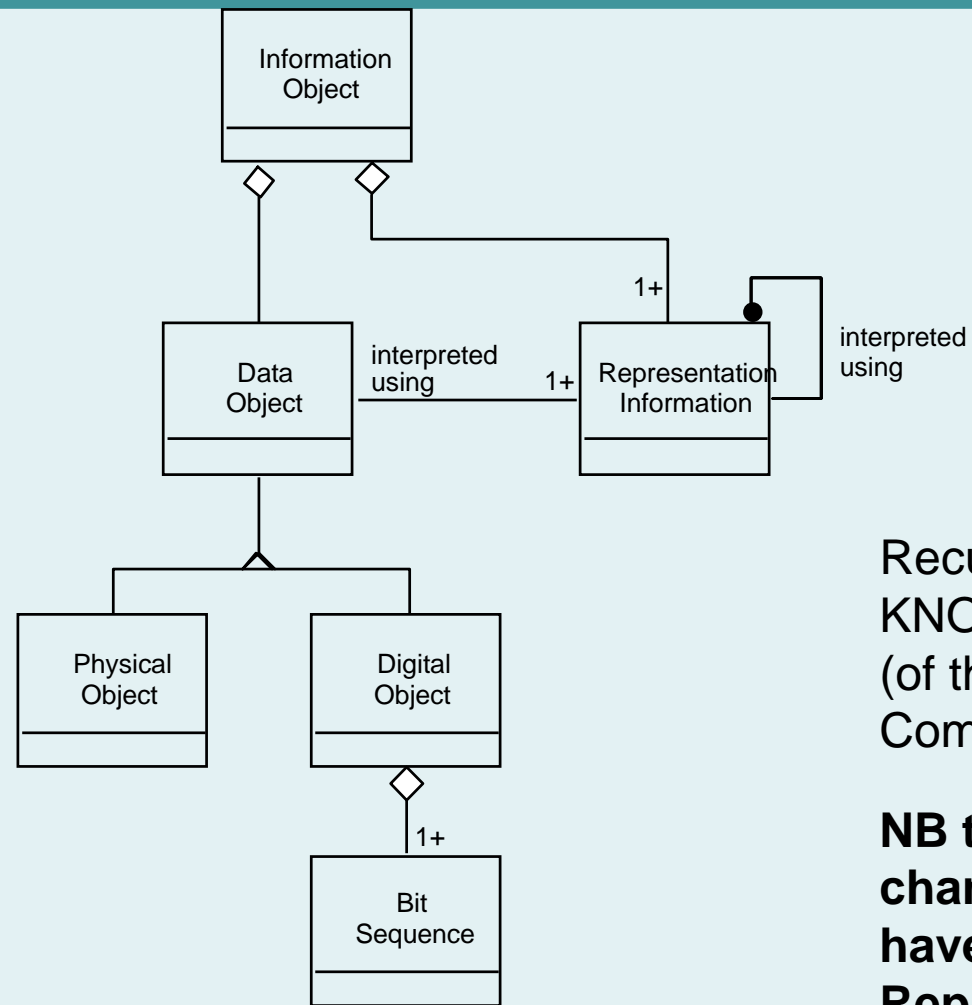
# OAIS Repository Model



SIP = Submission Information Package  
AIP = Archival Information Package  
DIP = Dissemination Information Package



# OAIS data model



Recursion ends at  
KNOWLEDGE BASE  
(of the Designated  
Community)

**NB the end point  
changes – and so we  
have to add more  
ReplInfo over time**



# Caveats

## ❑ Many different issues under “Preservation”

### ▪ Organizational

- Financial
- Political
- Educational

Planets can provide “arguments”

### ▪ Technical

- Planning
- Validating
- Executing
  - Migration
  - Emulation
- Revising

Planets can provide “tools”





# The “real collection”

---

- Click to add text



## Part of sample collection ...

---

- ☐ 00000005.gif
- ☐ 00000006.gif
- ☐ 00000007.gif
- ☐ 00000008.gif
- ☐ 00000009.gif
- ☐ 00000010.gif
- ☐ 00000011.gif
- ☐ 00000012.gif
- ☐ 00000013.gif
- ☐ 00000014.gif
- ☐ 00000015.gif
- ☐ 0001.GIF
- ☐ 0002.GIF
- ☐ 0003.GIF
- ☐ 0004.GIF
- ☐ 0005.GIF
- ☐ 0006.GIF
- ☐ 0007.GIF
- ☐ 0008.GIF
- ☐ 0009.GIF
- ☐ 0010.GIF
- ☐ 0011.GIF
- ☐ 0012.GIF
- ☐ bear.gif
- ☐ Lada\_\_1200\_E.gif
- ☐ Lamborghini\_\_Countach\_LP\_500.gif
- ☐ Matra\_Simca\_\_Rancho.gif
- ☐ Trabant\_\_P\_601\_L.gif
- ☐ Triumph\_\_Spitfire\_1500.gif
- ☐ Volvo\_\_264\_GL.gif
- ☐ VW\_\_Golf.gif
- ☐ Zastava\_\_750.gif



# Part of sample collection

---

- ❑ basi0g01.docx
- ❑ basi0g02.docx
- ❑ fourth.docx
- ❑ PNG\_capabilities.docx
- ❑ third.docx
- ❑ basi0g01.pdf
- ❑ basi0g02.pdf
- ❑ fourth.pdf
- ❑ PNG\_capabilities.pdf
- ❑ third.pdf
- ❑ oi9n0g16.png
- ❑ oi9n2c16.png
- ❑ oxford.png
- ❑ pngsuite\_logo.png
- ❑ pp0n2c16.png
- ❑ pp0n6a08.png
- ❑ ps1n0g08.png
- ❑ ps1n2c16.png
- ❑ oi9n0g16.png.svn-base
- ❑ oi9n2c16.png.svn-base
- ❑ oxford.png.svn-base
- ❑ pngsuite\_logo.png.svn-base
- ❑ pp0n2c16.png.svn-base
- ❑ pp0n6a08.png.svn-base
- ❑ ps1n0g08.png.svn-base
- ❑ ps1n2c16.png.svn-base

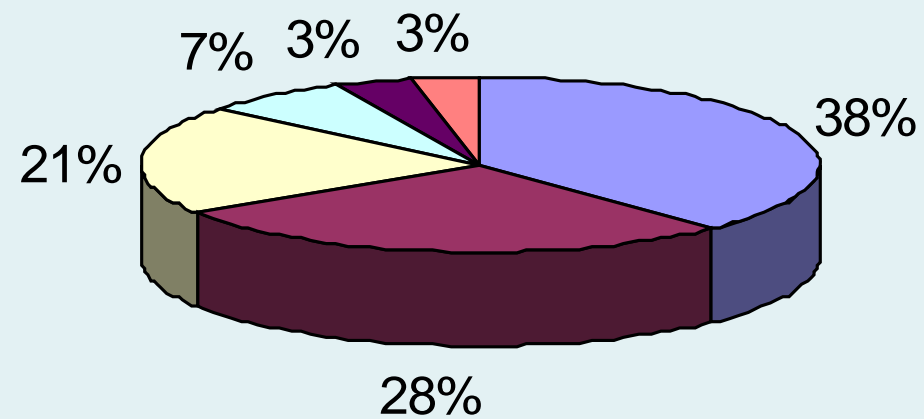


# The Testbed Corpora

- ☐ directory amh
- ☐ directory AVI
- ☐ directory BMP
- ☐ directory btest
- ☐ directory DOC
- ☐ directory EPS
- ☐ directory FLV
- ☐ directory GIF
- ☐ directory HTML
- ☐ directory JPEG
- ☐ directory JPEG2000
- ☐ directory lost+found
- ☐ directory MOV
- ☐ directory MPEG
- ☐ directory MS\_XLS
- ☐ directory MSG\_Outlook
- ☐ directory mvd
- ☐ directory Naneth Databases
- ☐ directory Naneth XML
- ☐ directory OGG
- ☐ directory OOXML
- ☐ directory OpenDocument
- ☐ directory PBM
- ☐ directory PCD
- ☐ directory PCX
- ☐ directory PDF
- ☐ directory PDFA-1b
- ☐ directory PICT
- ☐ directory PNG
- ☐ directory PPM
- ☐ directory RTF
- ☐ directory SVG Testsuite W3C
- ☐ directory test
- ☐ directory test\_ba
- ☐ directory testacct
- ☐ directory Testfiles from KB-DK
- ☐ directory TIFF
- ☐ directory tlr
- ☐ directory training
- ☐ directory TXT
- ☐ directory WAV
- ☐ directory WMV
- ☐ directory WordPerfect
- ☐ directory XHTML



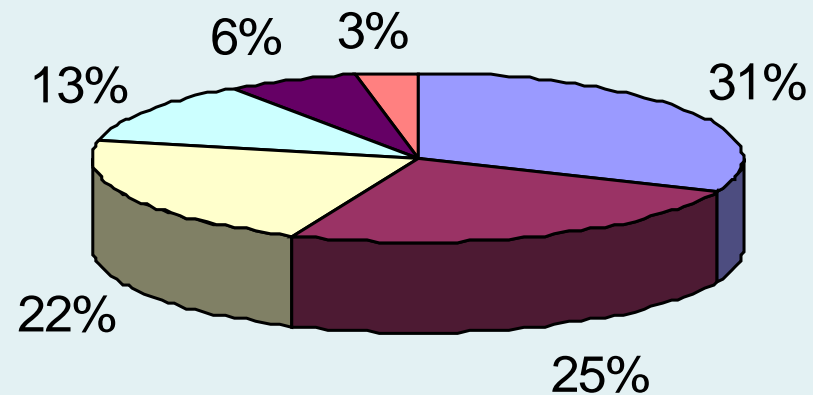
# The real users (where they come from)



# The real users (what they are doing)

---

- Digital archivist
- Software developer
- Digital librarian
- Preservation advisor
- Researcher
- Museum curator



# Preservation within the organisation

---

☐ Does your organisation have a digital preservation policy?

- yes 17
- no 9

☐ If yes, is the policy reviewed

- Annually or more frequently 3
- Every 2-5 years 6
- Has not been reviewed yet 9

☐ Does your organisation have a budget for digital preservation?

- yes 18
- no 8



# Plans for preservation

---

- ❑ Is digital preservation present in your:
  - Operational planning 21
  - Financial planning 18
  - Business-continuity planning 19
- ❑ Plans for the long-term management of digital information:
  - Long-term solution in development 10
  - Long-term solution in place 8
  - No plans 5
  - Assessing needs 2
  - Tendering for a solution 1





# Source of digital information

---

	Now	in 10 years
<input type="checkbox"/> File system	20	9
<input type="checkbox"/> Scanner	15	4
<input type="checkbox"/> EDMS	13	8
<input type="checkbox"/> Internet	12	6
<input type="checkbox"/> Media store	10	2
<input type="checkbox"/> E-mail	9	9
<input type="checkbox"/> ERMS database	6	6
<input type="checkbox"/> CAD	3	7
<input type="checkbox"/> Lab systems	2	4
<input type="checkbox"/> Educational Research and testing data		
<input type="checkbox"/> Institutional repositories		
<input type="checkbox"/> Project funded outputs   Legacy AHDS collections		
<input type="checkbox"/> Rich media digitisation programmes		



# Type of digital information

---

❑ Documents	26
❑ Images	22
❑ Video	19
❑ Audio	19
❑ Websites	17
❑ Databases	15
❑ e-Journals	12
❑ e-Mails	12
❑ e-Books	10

❑ Software	8
❑ GIS	6
❑ Scientific data	6
❑ Disc images	5
❑ Computer games	
❑ Complex objects	



# Quantity of digital information

---

## □ How much digital information do you store

	now	in 10 years
▪ Less than 20TB	12	1
▪ 20-100TB	8	6
▪ 100TB-1PB	3	7
▪ More than 1PB	3	12



# Conclusions

---

- ❑ All in all, many (most ?) organizations still need to be converted to “preservation” (and Planets can help in preaching)
- ❑ For preservation planning and execution, there is no “one size fits all” solution; each organisation is a particular case
- ❑ At the end of these two days you should have (hopefully) a better understanding of the (technical) issues underlying preservation, and some knowledge of the possible solutions offered by Planets

