



How to preserve? – Sara van Bussel

The National Library of The Netherlands



## How to preserve?

---

- ❑ Types of preservation action
- ❑ Preservation action tools in Planets
- ❑ Planets Core Registry



# Types of preservation action



## Planets digital preservation

---

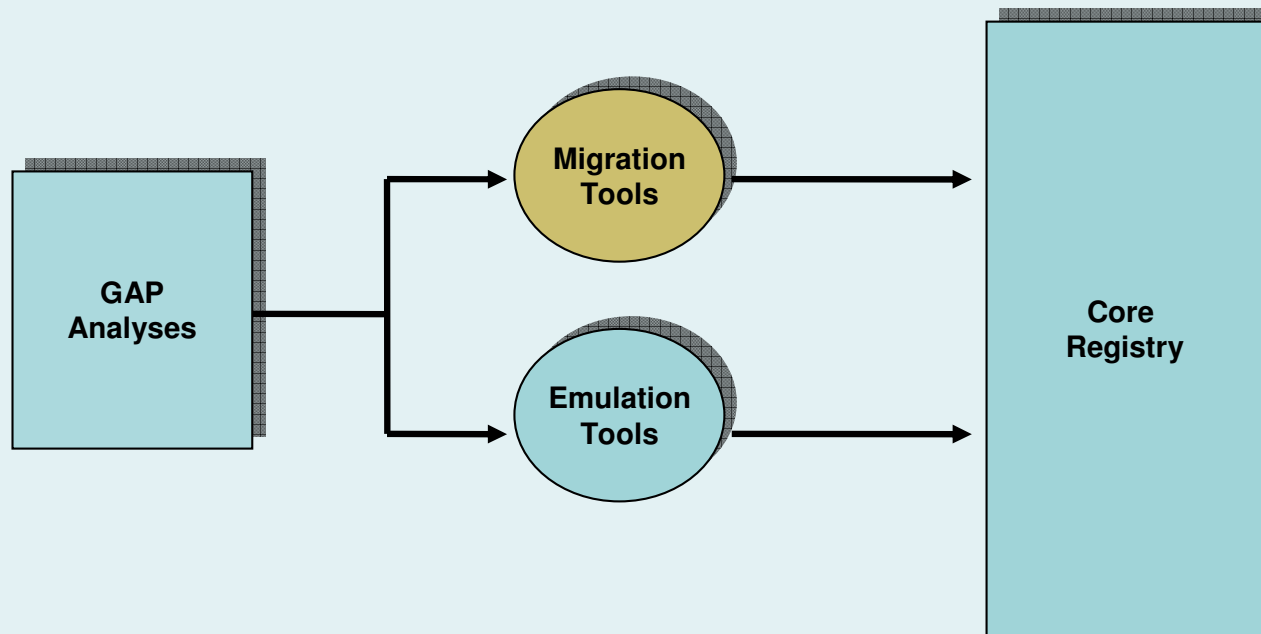
- ❑ Logical preservation addresses the problem of accessing bit streams, whose interpretation may depend on obsolete operating systems, applications, or formats.
- ❑ Preservation Actions
  - Migration
  - Emulation



# Migration



## Relationships between PA activities



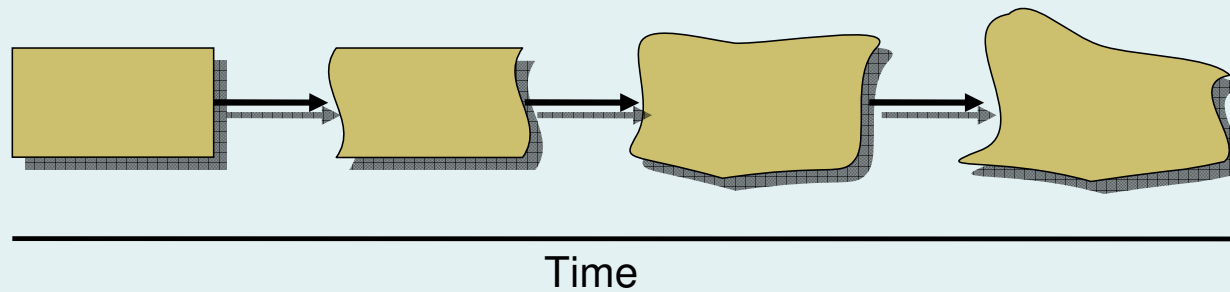
11/17/09



## Migration

**Migration** = changing the object

- The object is available in the current environment
- Re-using information is simple

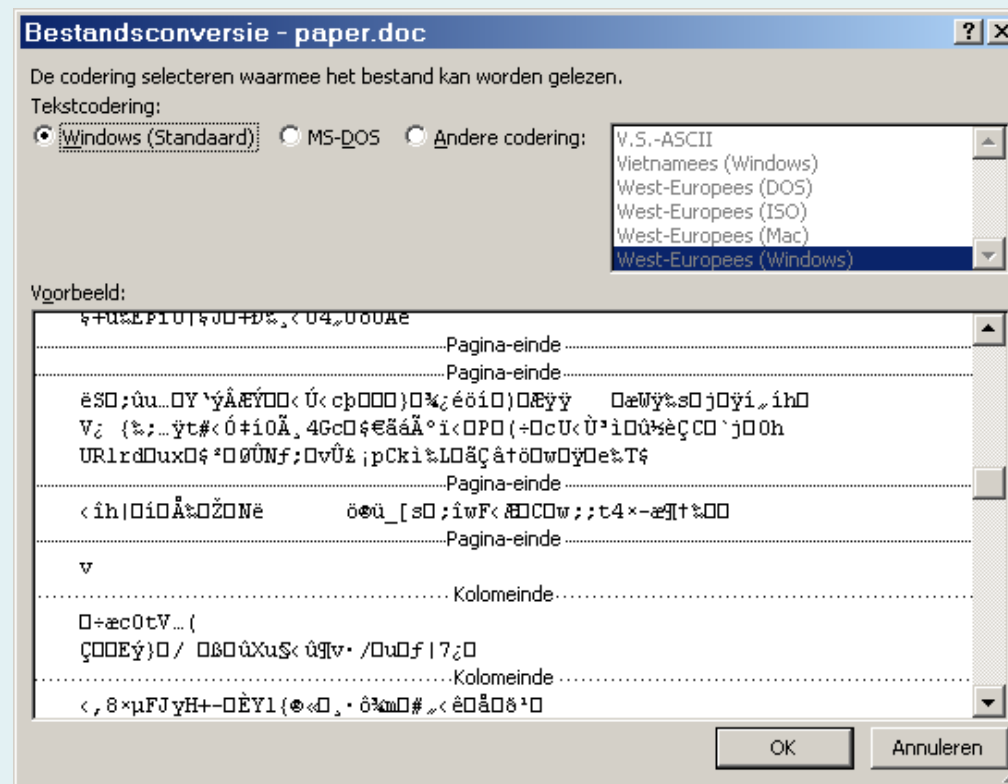


Risks:

- Inconsistencies can occur
- Functionality can be lost
- Quality Assessment is difficult

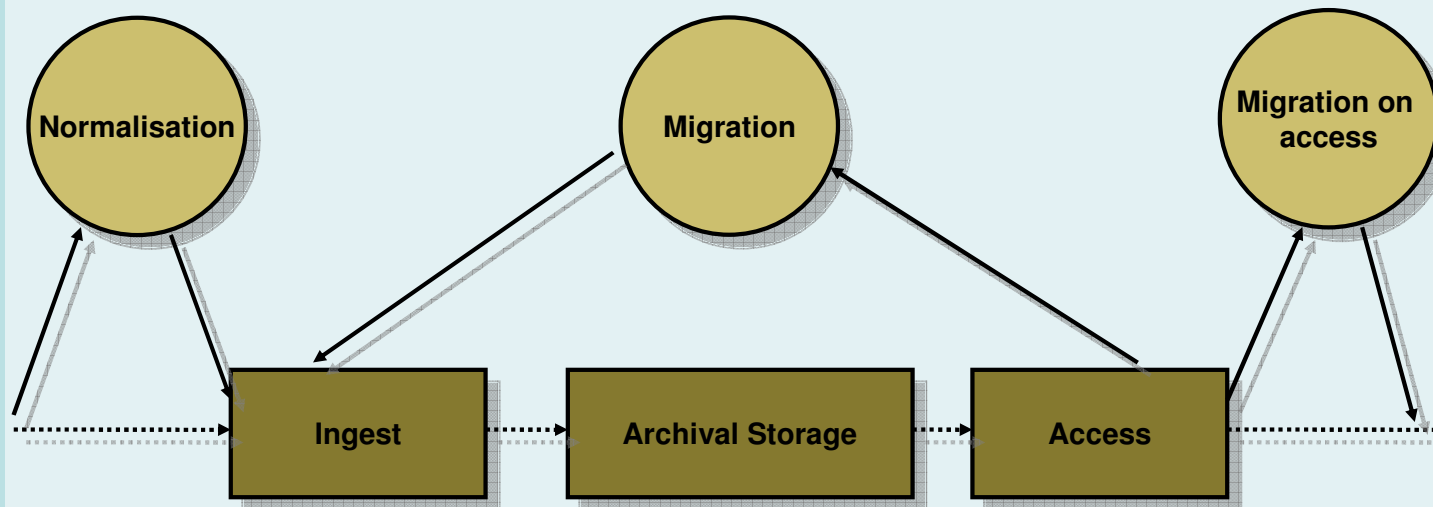


## WP5.1 -> Word 95 -> Word 97 -> Word XP -> ...





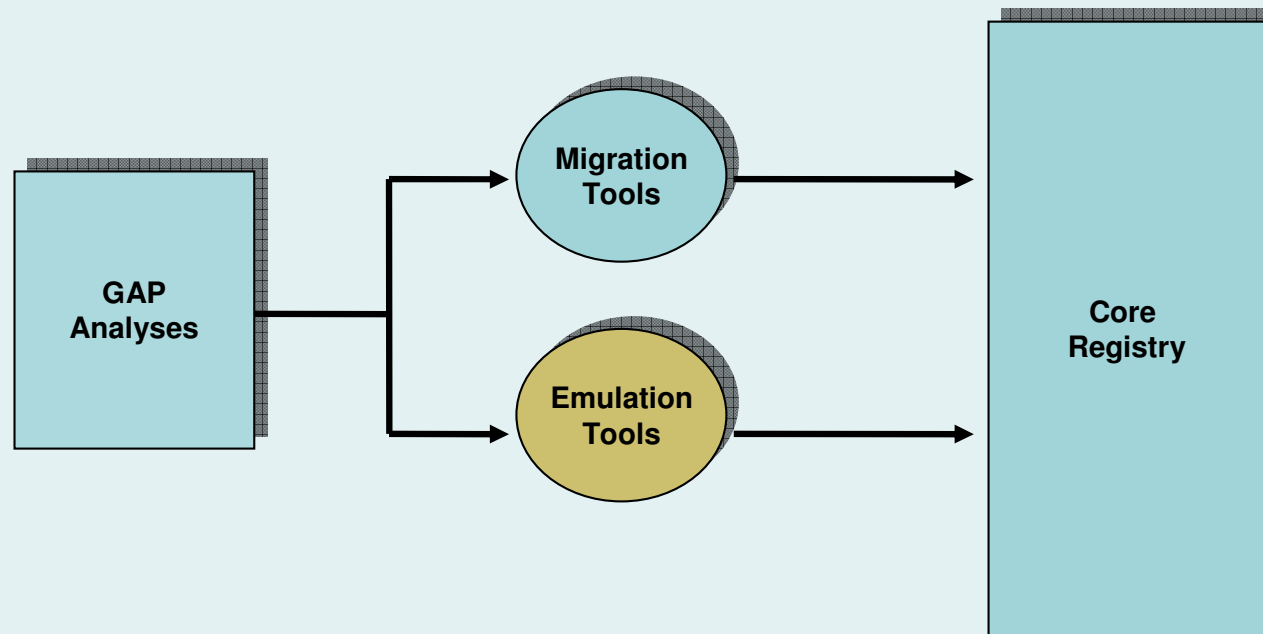
## Migration and digital archiving



# Emulation



## Relationships between PA activities



11/17/09



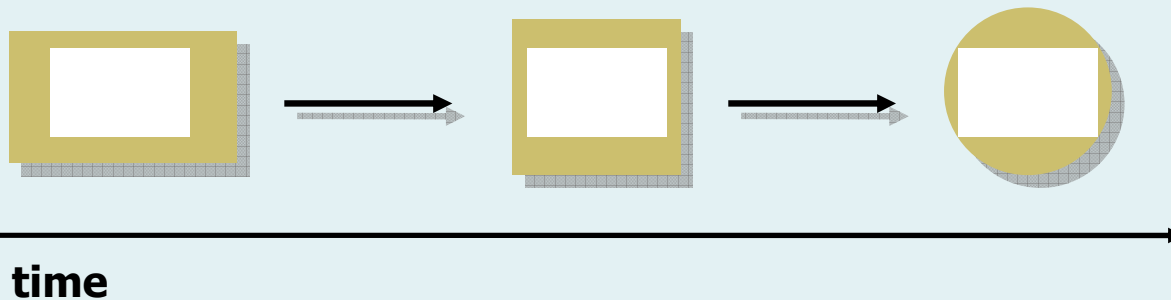
## Emulation – Windows 2000 in Windows XP environment



## Emulation

**Emulation** = changing the environment

- Authentic environment (including functionality)
- No changes to the object are needed.



Risks:

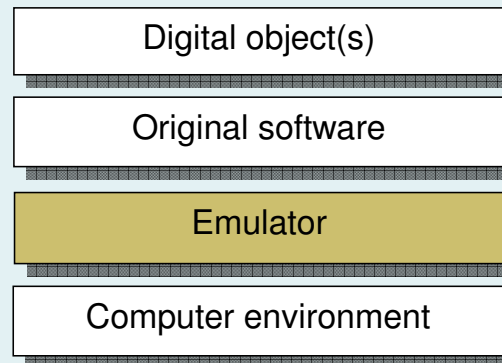
- Technically challenging
- User has to have knowledge about original environment



## Emulation

---

- ❑ An emulator emulates a hardware configuration
- ❑ An original operating system and original software are still needed to access files



- ❑ Emulation for digital preservation has specific requirements





## Emulation – Use cases

---

- ❑ Rendering old websites
  - Use emulation to view websites in a web archive
- ❑ Opening old files
  - WordPerfect files in an archive
  - Files containing geographical data in a library
  - Databases
- ❑ Executing old programs
  - Games, which are more and more part of collections
  - Scientific applications
  - Multimedia applications
  - Digital art





## Games accessible through emulation

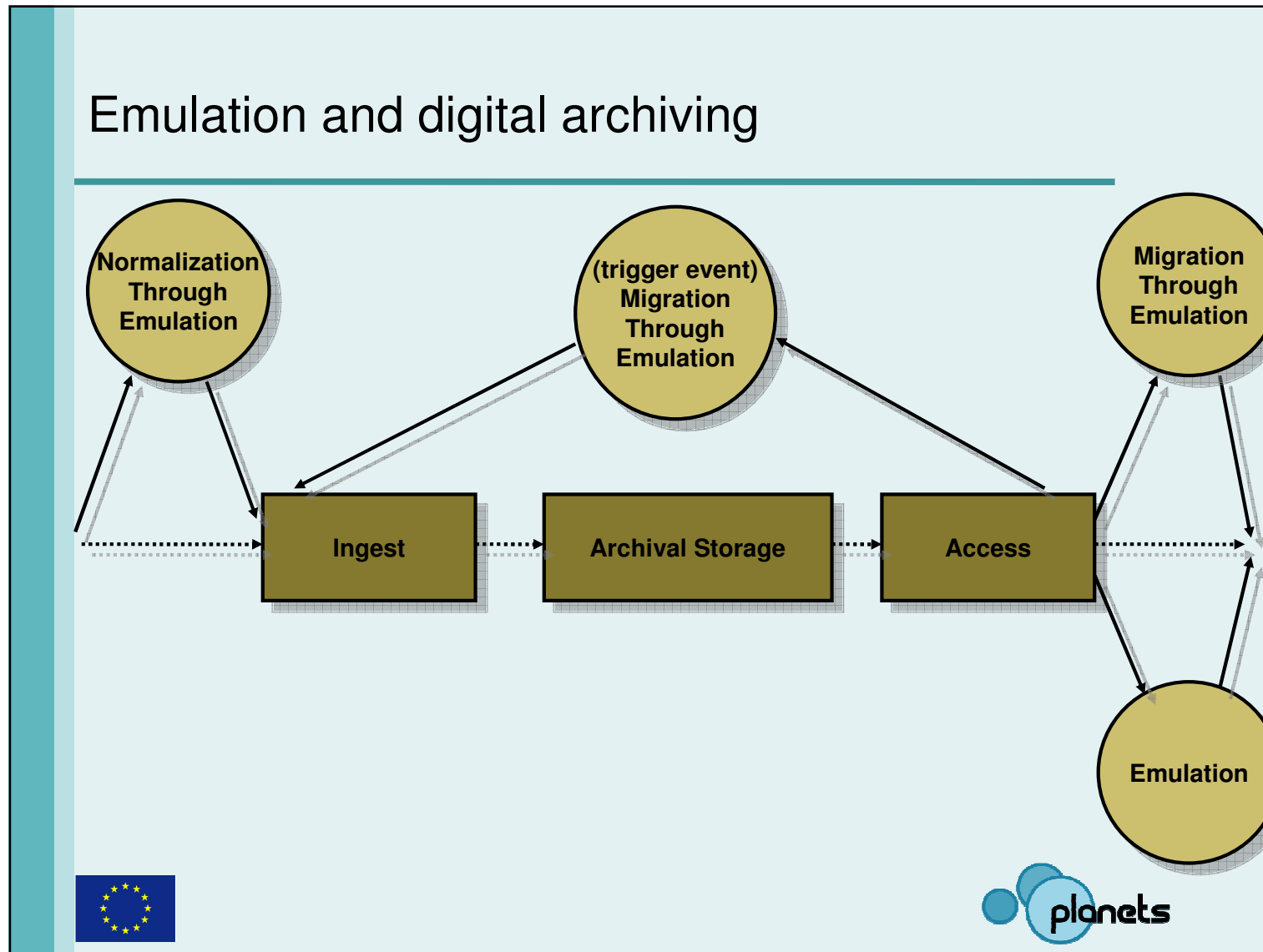


## Migration through Emulation

---

- ❑ A migration tool can become obsolete
- ❑ An emulator can run the hardware/operating system/software combination needed to keep using the migration tool



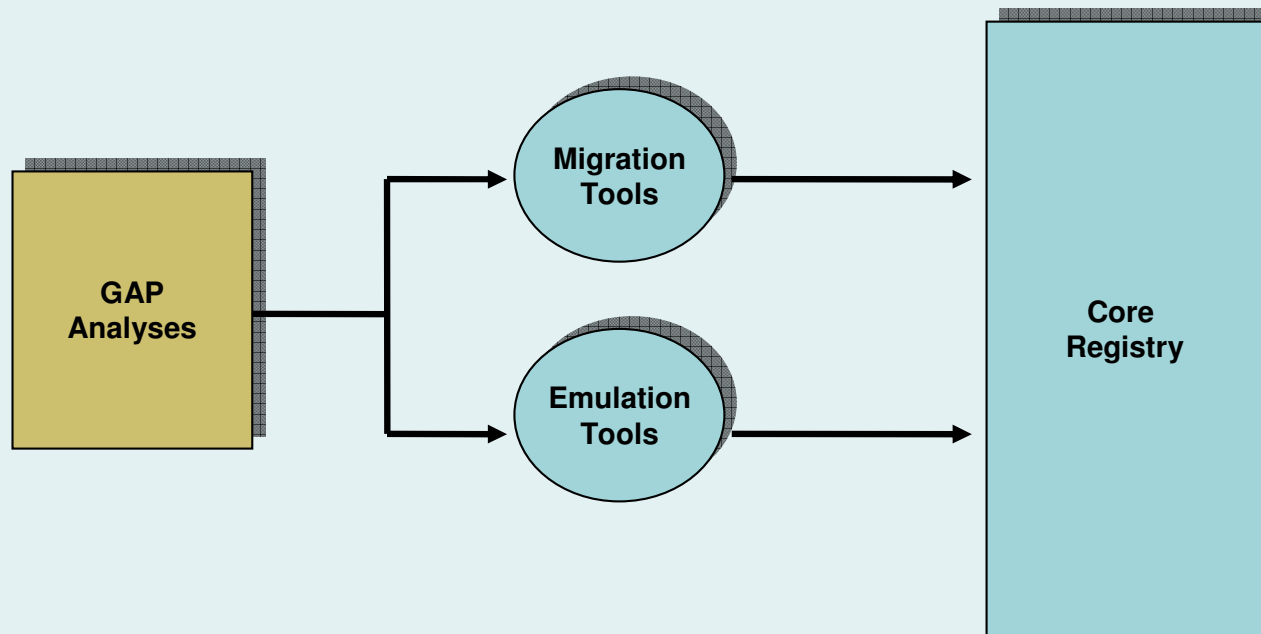


# Preservation action tools in Planets

## Gap analysis



## Relationships between PA activities



11/17/09



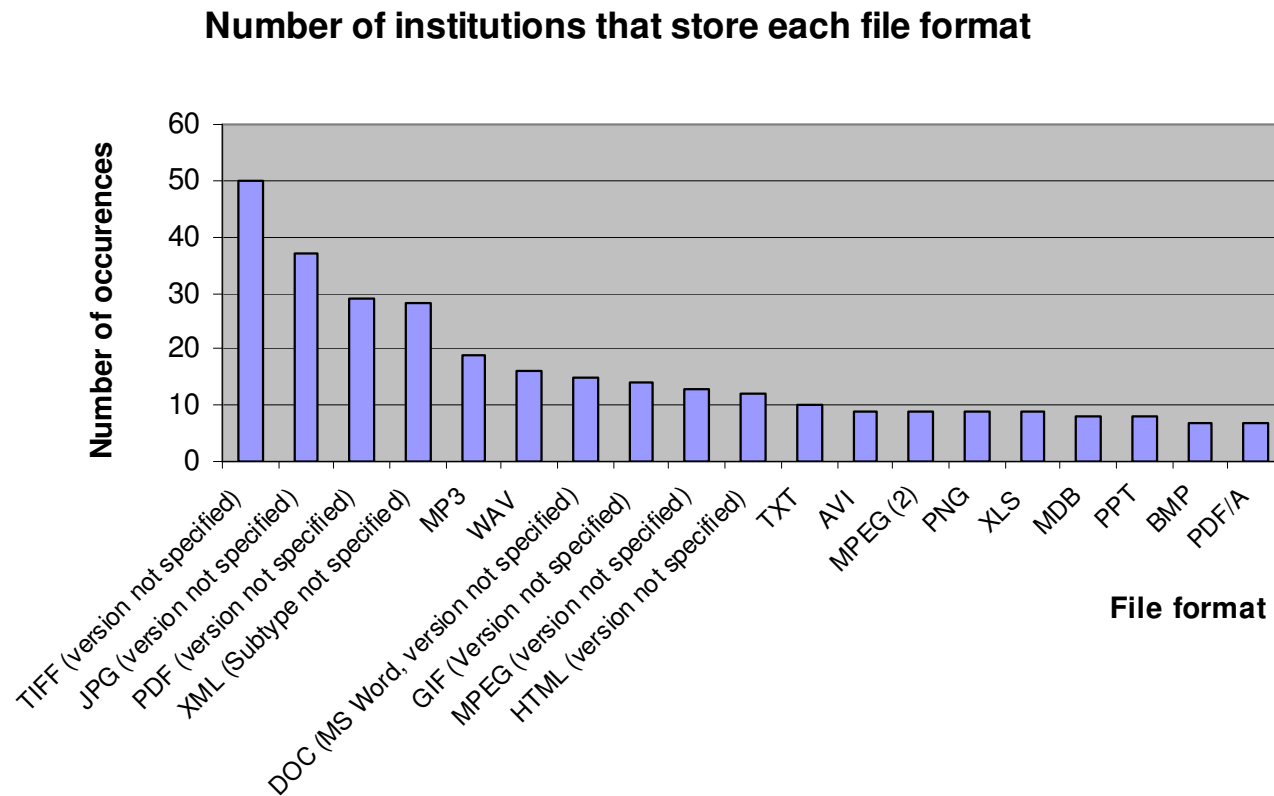
## Gap analysis in tool provision

- ❑ Inventory of file formats in existing collections, archives and repositories
- ❑ Insight into the demand for preservation action tools
- ❑ 137 different file formats, 76 institutions
- ❑ Most archived file formats in three main types of institutions

	Archives	Libraries	Museums
TIFF	29%	23%	29%
JPG	23%	26%	29%
PDF	17%	20%	17%
XML	17%	14%	
DOC	14%		
MP3		17%	13%
HTML			13%



## Gap analysis in tool provision



## What does this mean?

---

- ❑ Only 22% of all file formats found in four or more institutions
- ❑ Most preservation action tools for 'big' file formats, however:
- ❑ DAISY: Audio books for the blind
  - Archived by a few institutions but large and international user base
  - Issues arising with new versions of the format are solved by a consortium of content providers
- ❑ Sheet music: Many file formats
  - Many commercial providers of proprietary file formats
  - Due to fragmentation no main solution
- ❑ FITS: Astronomical data
  - Developed and used by scientific organizations
  - Digital preservation issues are solved by knowledgeable users





## Gap analysis – conclusions and recommendations

---

- ❑ 57 migration tools known and/or used by Planets Partners
- ❑ Out of the top ten file formats, nine are covered by these tools
  - XML is not covered, is often the output file format
- ❑ Are there gaps in tool provision?
  - No, because some tools are extremely flexible (wrapping, printing, emulation)
  - Yes, because each organisation has specific demands that might not be fulfilled by available tools
- ❑ Gap analysis should be redone periodically to take advantage of new information and new developments



# Preservation action tools in Planets Services



## Preservation Action Tools and Planets

---

- ❑ List of preservation action tools to be wrapped is prioritized based on the gap analysis



- ❑ The preservation action tool is tested



- ❑ The preservation action tool is wrapped as a service



- ❑ The service is tested



- ❑ The service is integrated in the Planets environment



## Migration – Tools wrapped for the Planets environment

---

- ☐ AbiWord
- ☐ AviDemux
- ☐ Dia
- ☐ Extractor (XcdlMigrate)
- ☐ ffmpeg
- ☐ FloppyImageHelper
- ☐ Ghostscript
- ☐ Gimp
- ☐ GraphicsMagickMigrate
- ☐ ImageMagick
- ☐ Im4JavaImageMagischMigrate
- ☐ Inkscape
- ☐ Jasper 19
- ☐ Java-SE
- ☐ JJ2000
- ☐ Jtidy
- ☐ Mdb2Siard
- ☐ MsgText
- ☐ NetPBM
- ☐ OpenJpeg
- ☐ Pdf2PdfAMayComputer
- ☐ Pdf2Ps
- ☐ PdfBox
- ☐ Ps2Pdf
- ☐ SanselanMigrate
- ☐ SoX
- ☐ Xena



## Emulation – Making available in the Planets Framework

---

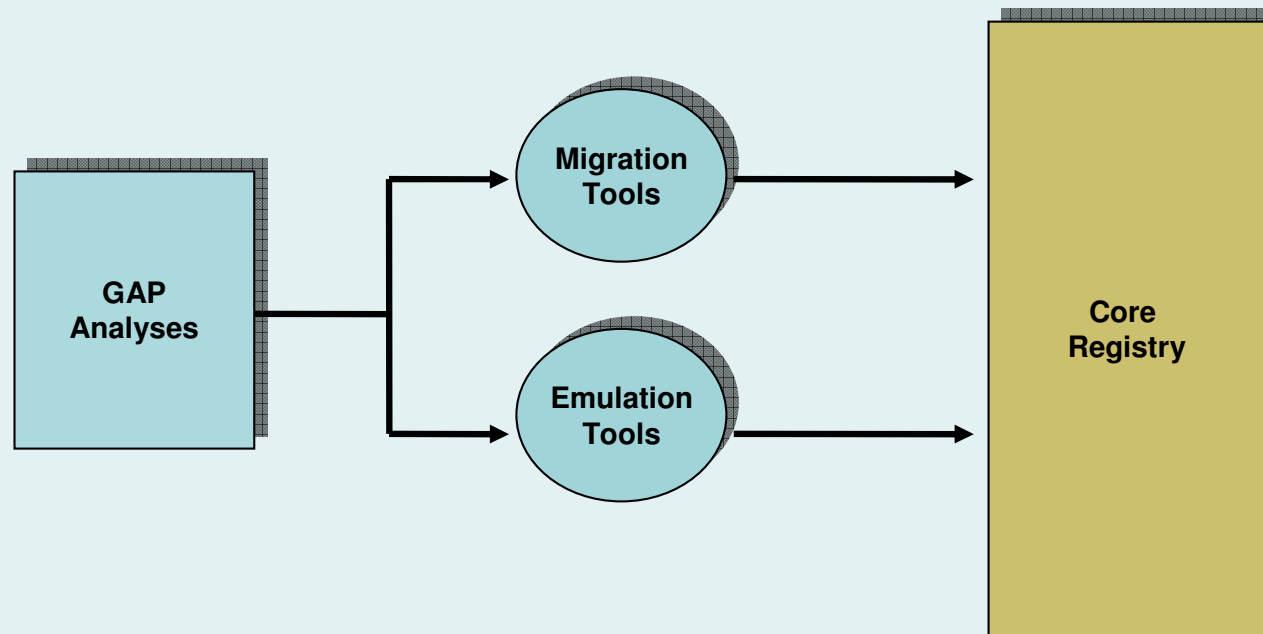
- ❑ Migration through emulation
  - ARJ (archiving format) using Dioscuri (emulator)
  - PNM to PNG using Dioscuri
  - GIF using the Universal Virtual Computer
- ❑ Emulation
  - CreateView using GRATE



# Planets Core Registry



## Relationships between PA activities



11/17/09



## Planets Core Registry

---

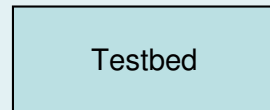
- ❑ Based upon Pronom
  - Existing file format registry developed by The National Archives
- ❑ Combined registry for preservation action tools and file formats
- ❑ Contains information about:
  - File formats
  - Software
  - Hardware
  - Media



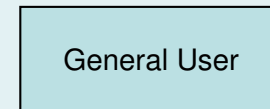


## Relationships of the Planets Core Registry

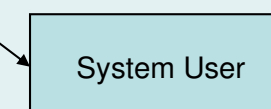
Test results from the Testbed  
can be stored in the registry.



Plato can retrieve information about  
file formats and preservation action  
tools from the registry.



A general user can use the web  
interface to search for information  
in the registry.



Web services can be used by  
applications to access information  
stored in the registry.



## Planets Core Registry

---

- ❑ The data in the Core Registry is checked and verified
- ❑ Keeping the data complete and up to date will be a collaborative community effort
- ❑ Whole community benefits



