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.
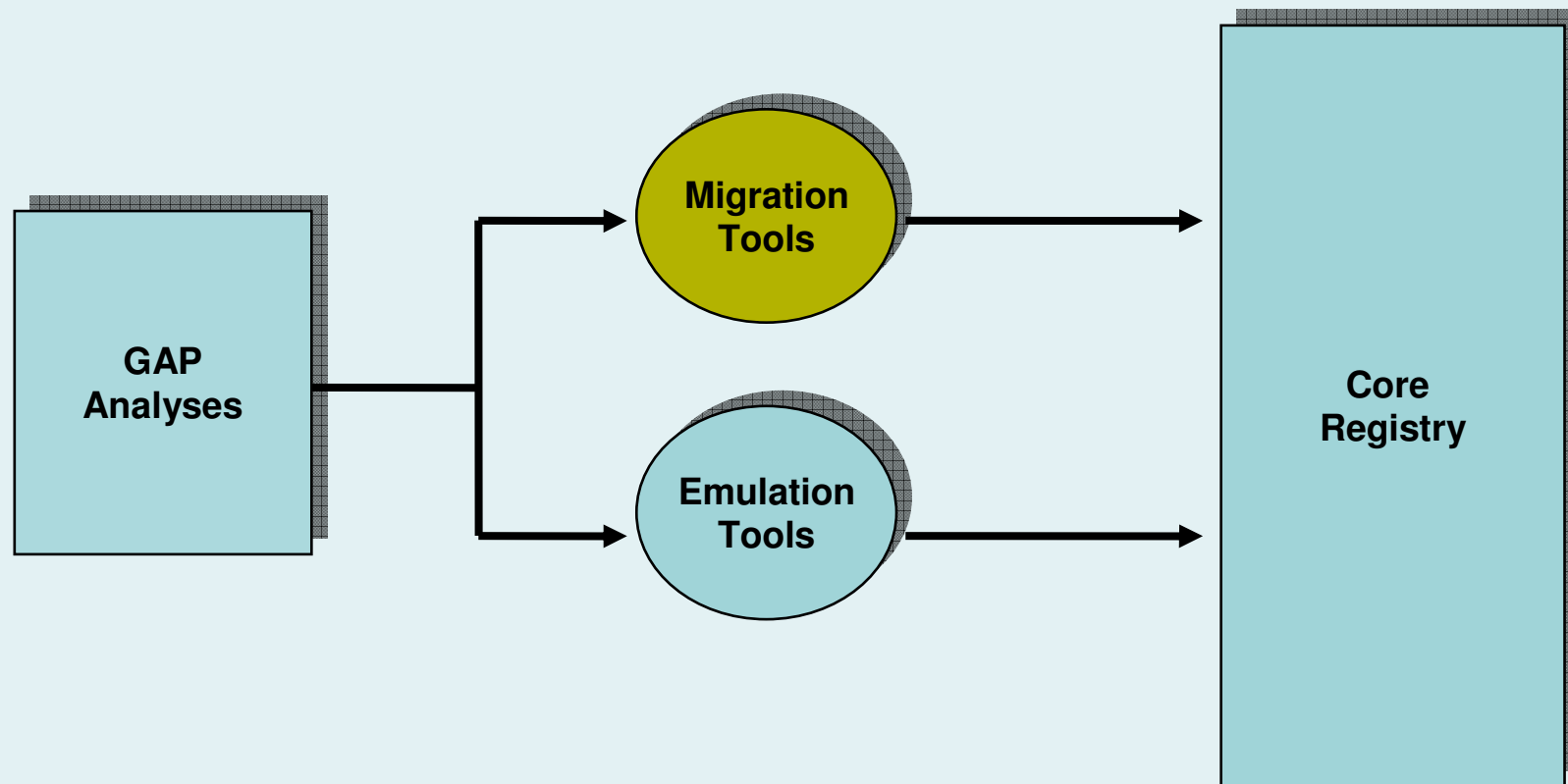
❑ Preservation Actions
  ▪ Migration
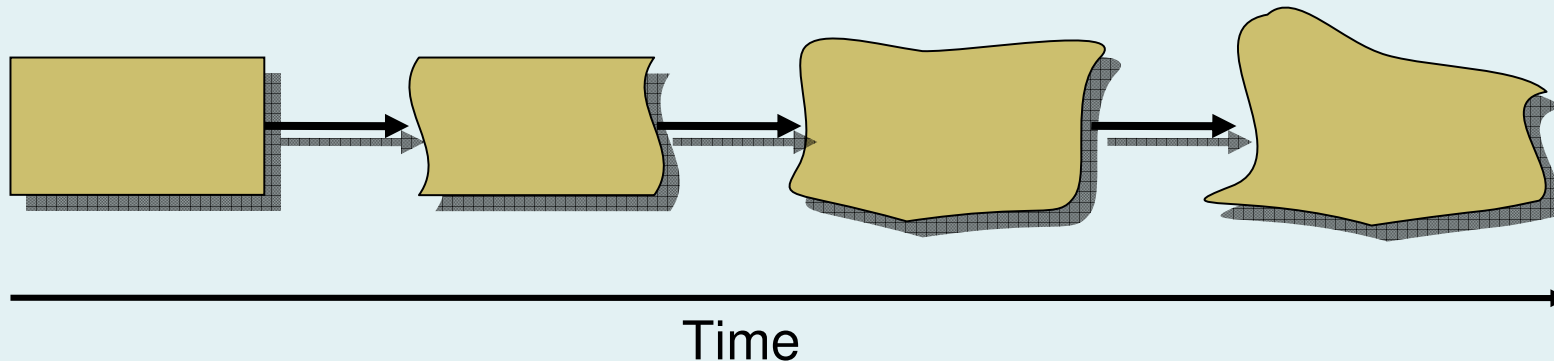 ▪ Emulation

Migration

Relationships between activities

Migration

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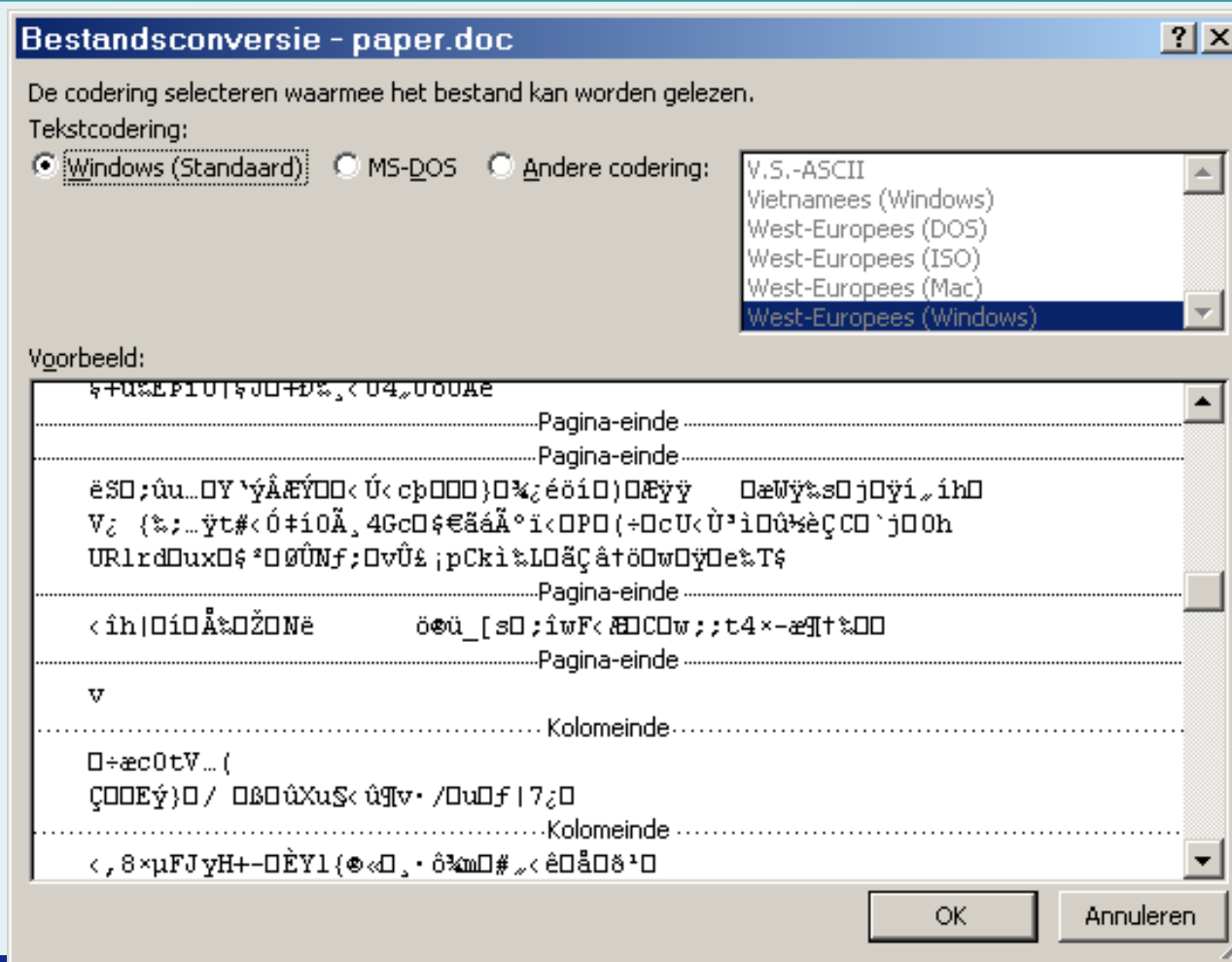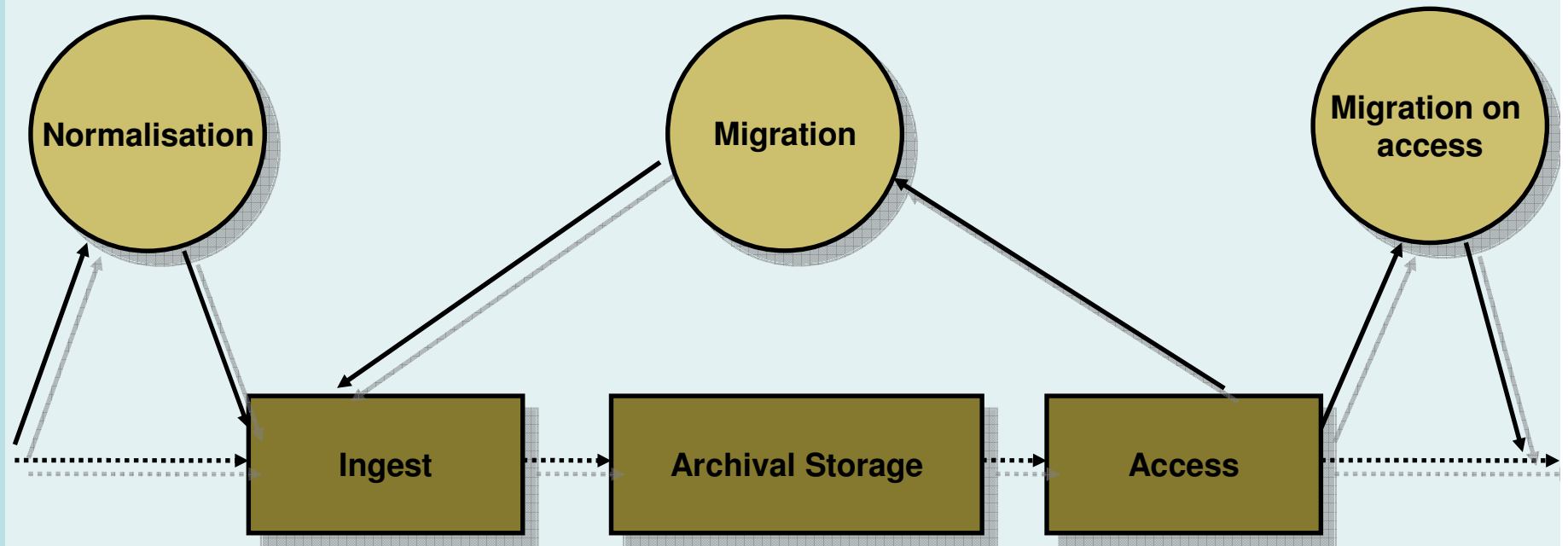- Re-using information is simple

Time

Risks:
- Inconsistencies can occur
- Functionality can be lost
- Quality Assessment is difficult
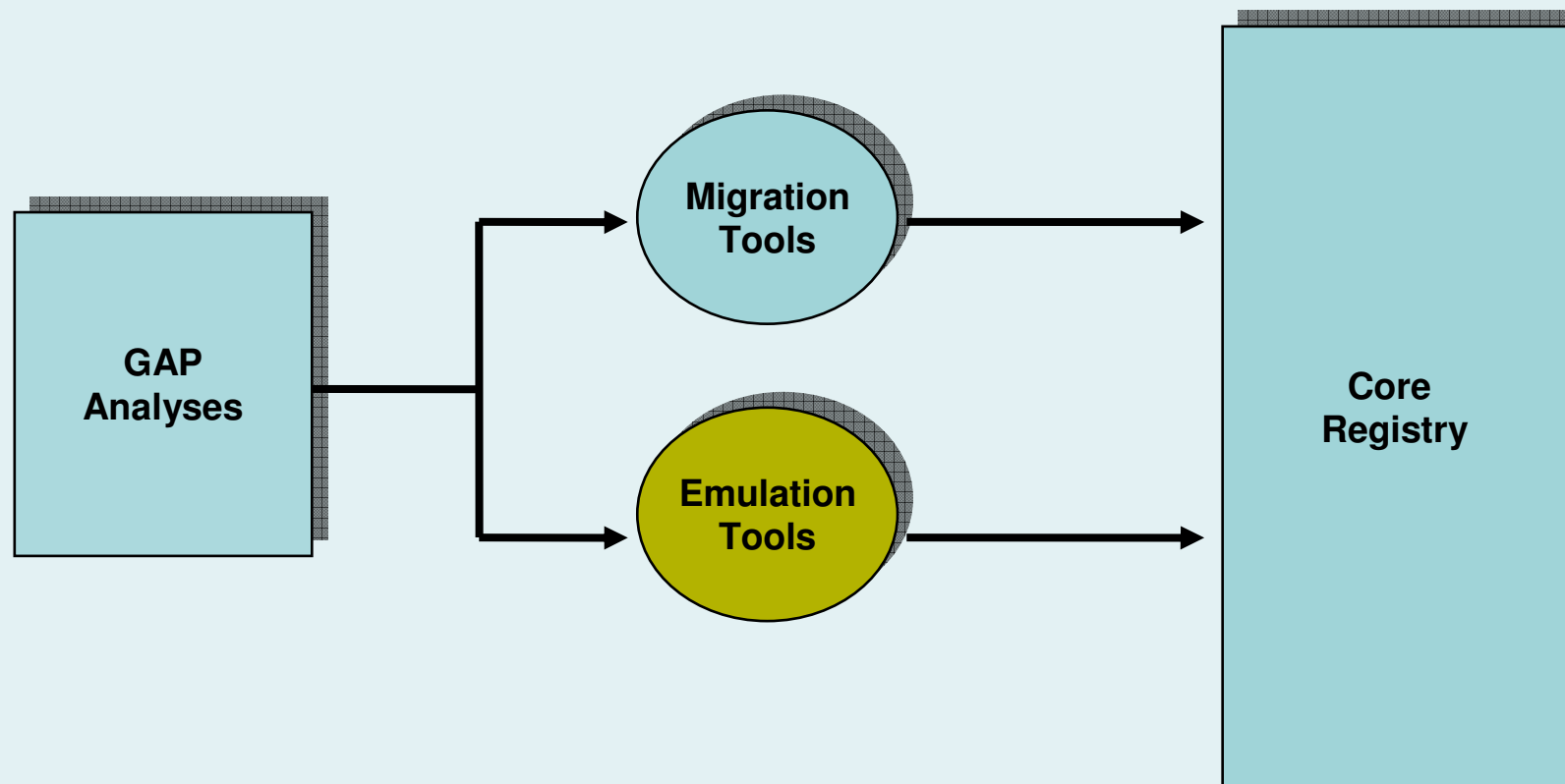
WP5.1 -> Word 95 -> Word 97 -> Word XP -> …

Migration and digital archiving

Emulation
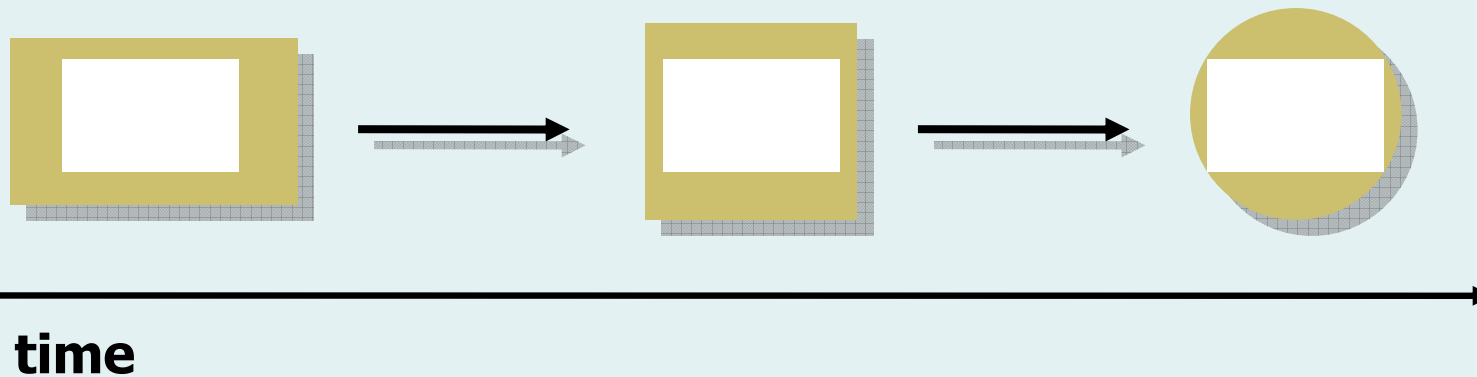
Relationships between activities

Emulation – Windows 2000 in Windows XP environment

Emulation

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

time

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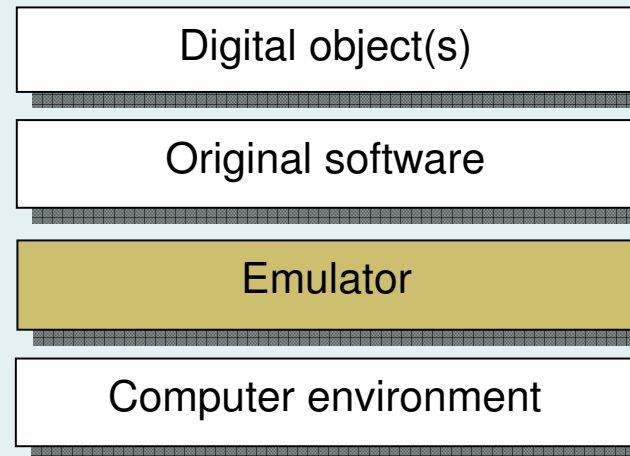- User has to have knowledge about original environment

Emulation

❑ An emulator emulates a hardware configuration

❑ An original operating system and original software are still needed to access files

Digital object(s)

Original software

Emulator

Computer environment

❑ Emulation for digital preservation has specific requirements

Panel

Left	Files	Commands	Options	Right

Name

√ Brief
Full
Info
Tree
On/Off   Ctrl-F1
─────────────────────
√ Name
eXtension
tiMe
Size
Unsorted
─────────────────────
Re-read
Drive...   Alt-F1

Name			Name	
123view	exe	qbasic	exe
Io	sys	qbasic	hlp
Msdos	sys	read	me
attrib	exe	sys	com
autoexec	bat	undelete	exe
chkdsk	exe	unformat	com
command	com	xcopy	exe
config	sys		
dbview	exe		
edit	com		
edit	hlp		
fdisk	exe		
format	com		

m   exe
           exe
small   exe

.sys   33430 11-11-91 5:00a

\>

lp   2User   3View   4Edit   5Copy   6R

D:\EMULAT~1\WORKSP~3\emu\images\floppy\nc\NC

Left	Files	Commands	Options

		Date	Time
Brief		-11-06	18:33
¹ Full		-10-06	9:45
Info		-11-06	15:29
Tree		-02-07	13:04
On/Off Ctrl-F1		-03-06	10:53
─────────		-12-05	15:14
¹ Name		-12-05	15:14
eXtension		-08-06	17:11
tiMe		-12-05	11:59
Size		-12-05	20:20
Unsorted		-07-06	15:21

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
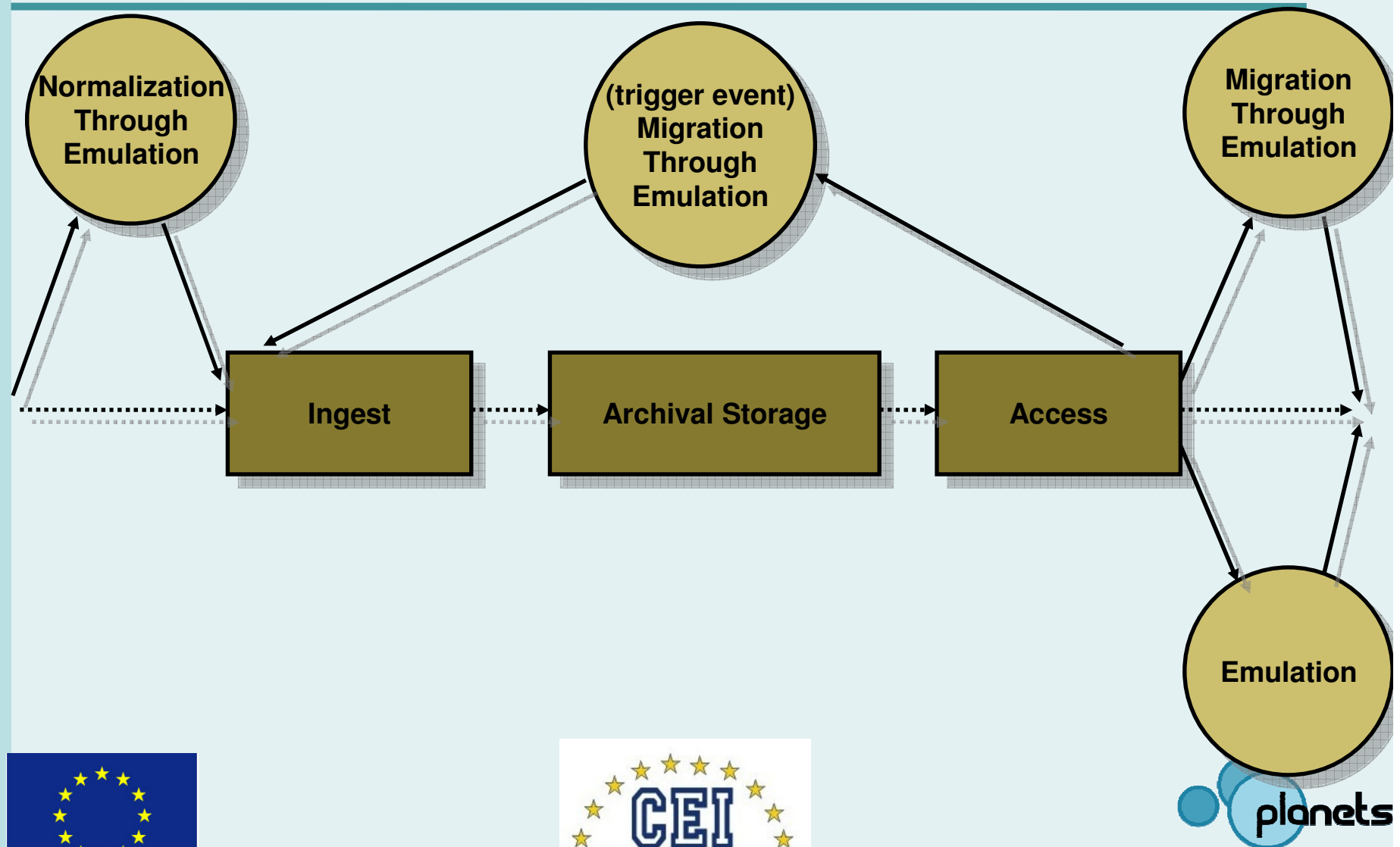
Migration through Emulation

❑ A migration tool can become obsolete

❑ An emulator can run the hardware/operating system/software combination needed to keep using the migration tool
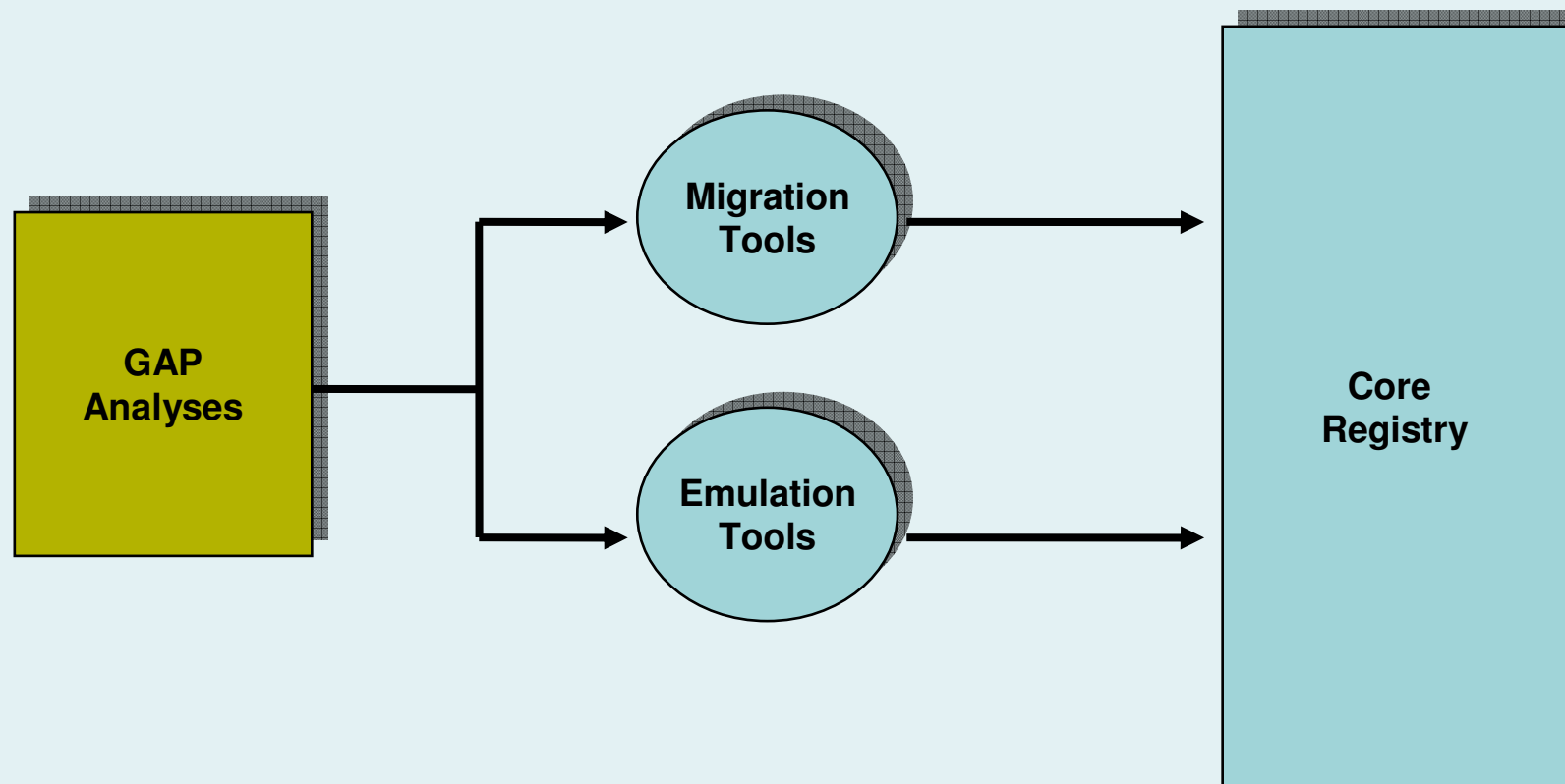
Emulation and digital archiving

Preservation action tools in Planets
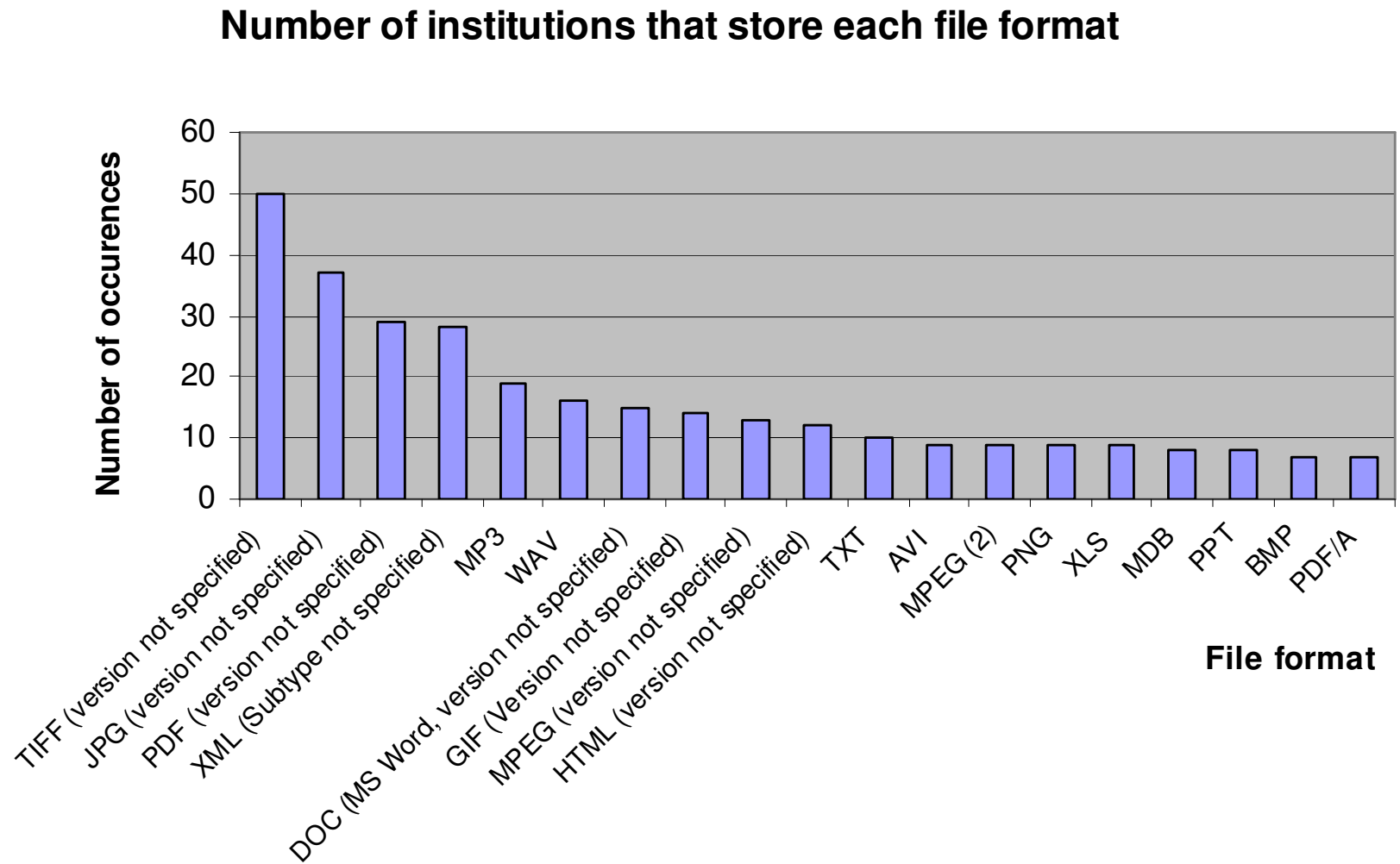
Gap analysis

Relationships between activities

Gap analysis in tool provision

❑ Inventory of file formats in existing collections, archives and repositories

❑ Insight into the demand for preservation action tools

❑ 107 different file formats, 65 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

Number of institutions that store each file format

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
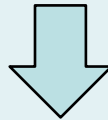   - Digital preservation issues are solved by knowledgeable users

CEI

CENTRAL EUROPEAN INITIATIVE

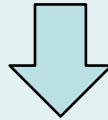planets

Preservation action tools in Planets
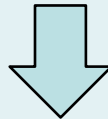
Services

Preservation Action Tools and Planets

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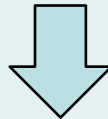
- The preservation action tool is tested

- 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

Tool	Input Format	Output Format
Microsoft Conversion Tools	Microsoft binary Office Formats; OOXML; ODF; legacy formats (WinWord2, WordPerfect, ...)	OOXML Formats; ODF; HTML
Dvips	Dvi	Ps (PostScript)
PS2PDF	PDF	PDF/A
PDF2PDFA	PDF	PDF/A
BullZip PDF Printer	Print to a PDF document from any Microsoft Windows application	PDF
PDF/A Converter	PDF	PDF/A
Ghostscript	PS (tool can be used for conversions)	PDF
GIMP	GIF, EPS, JPG, PNG, PS, TIFF, BMP, XCT	GIF, EPS, JPG, PNG, PS, TIFF,BMP
JMagik	GIF, TIFF, JPG	GIF, TIFF, JPG
Sanselan	GIF	TIFF
MsgText	MSG	TXT
SIARD	Relational Database	XML
Xena	Various	XENA
Sox	Various audio formats	Various audio formats

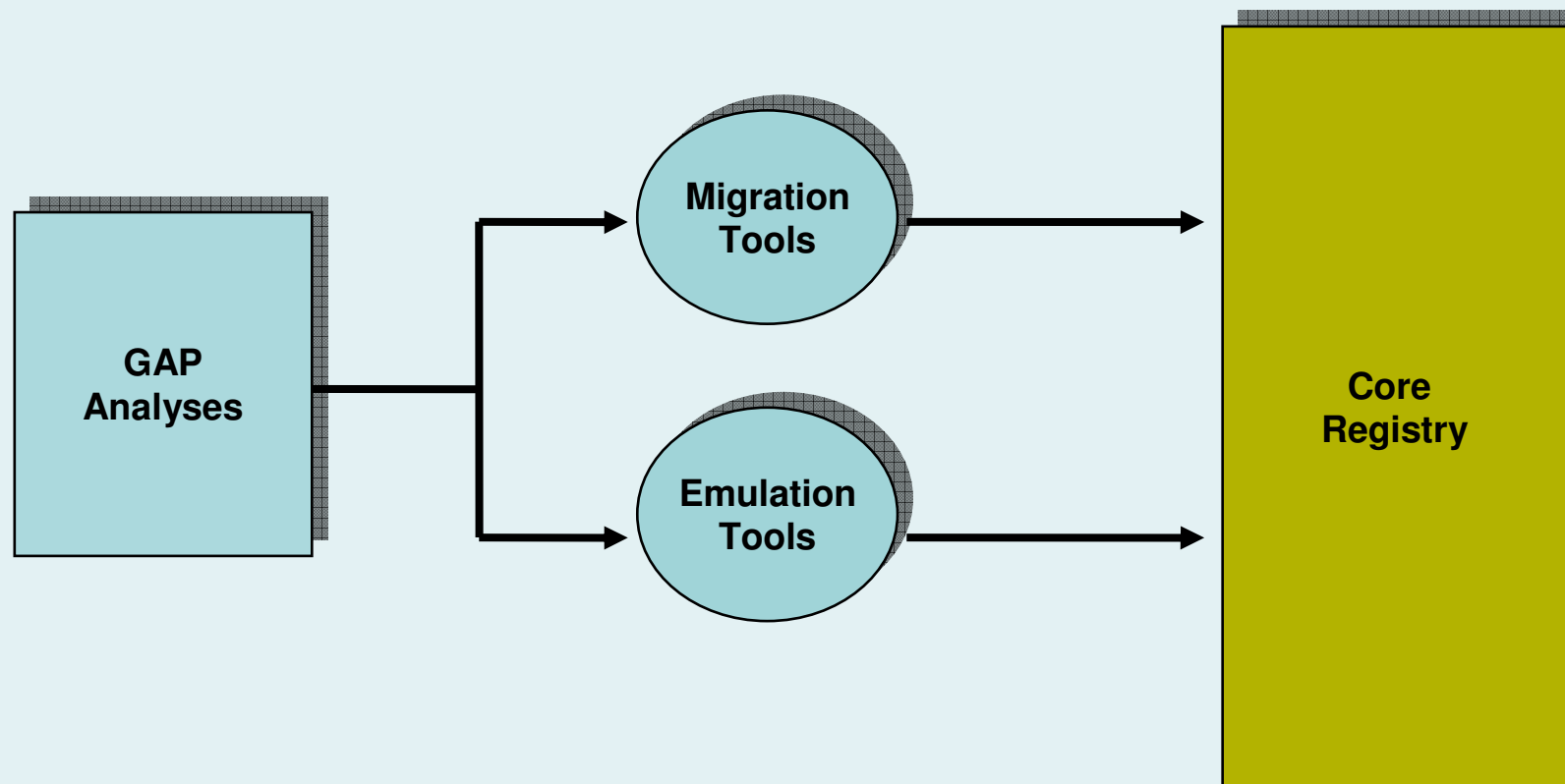# Emulation – Making available in the Planets Framework

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

Planets Core Registry

Relationships between activities
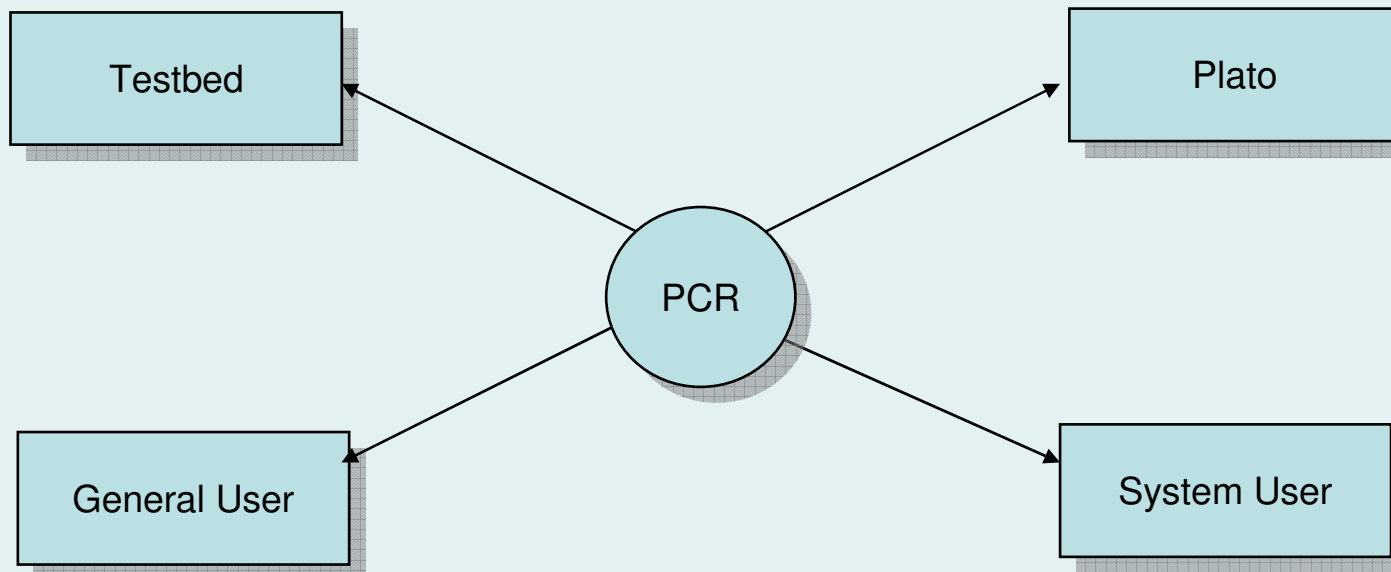
Planets Core Registry

- ❑ 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
are stored in the registry.

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

Testbed

Plato

PCR

General User

System User

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

Thank you for listening

Sara.vanbussel@kb.nl