



How to preserve? – Sara van Bussel

The National Library of The Netherlands



# How to preserve?

---

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

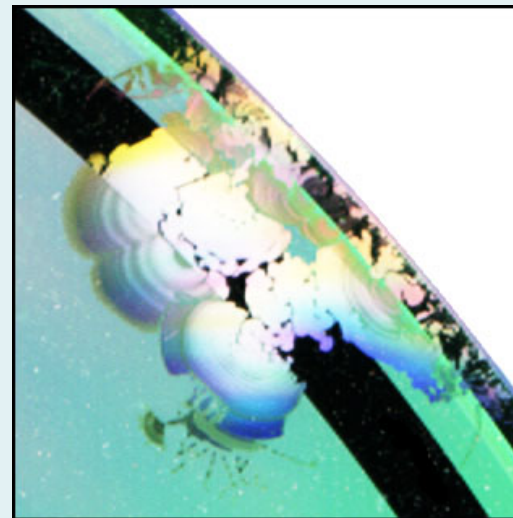


# The digital preservation problem

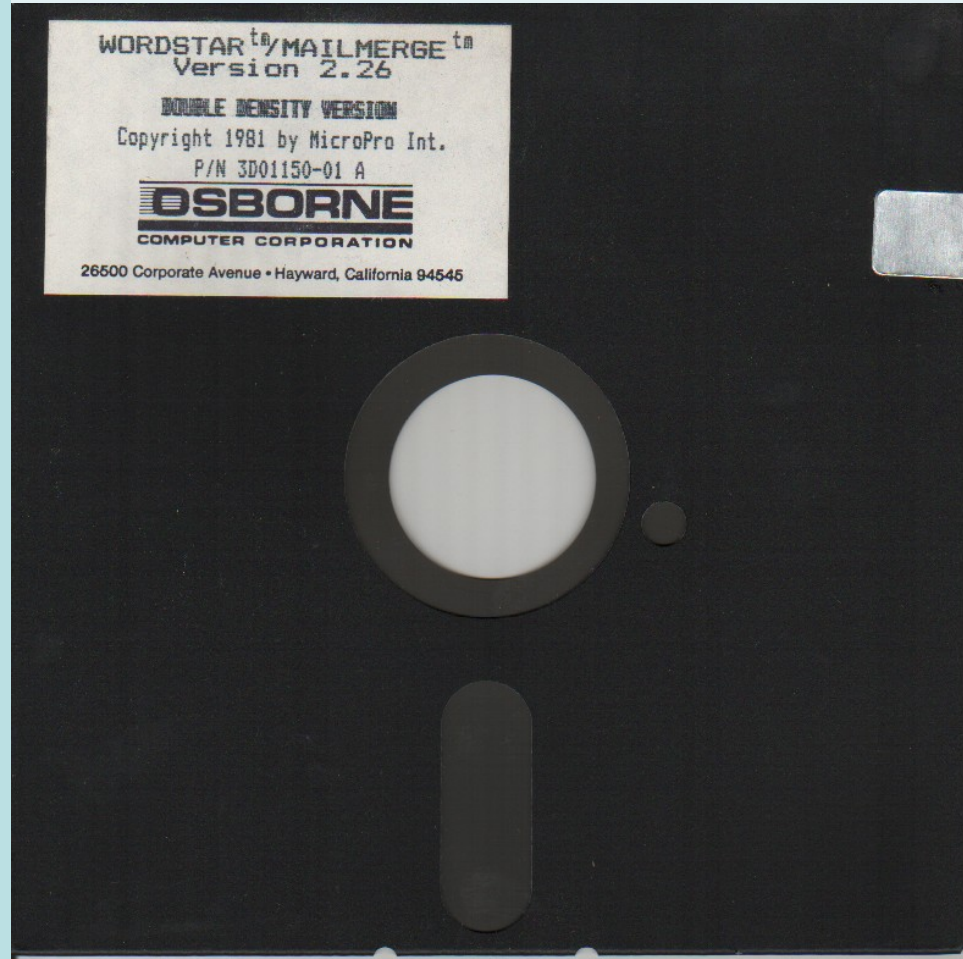


# Detoriation of storage medium

---



# Obsolescence of storage medium





# Obsolescence of hardware



# Obsolescence of formats

A,AAM,AAS,ABF,ABK,ACE,ACL,ACM,ACP,ACR,ACT,ACV,AD,ADA,ADB(2),ADD,ADF,ADI,ADM,ADP(2),ADR,ADS,AFM,AF2,AF3,AI,AIF,AIFF,AIFC,AIFF,AIM,AIS(2),AKW,ALAW,ALB,ALL,AMS(2),ANC,ANI,ANS,ANT,API,APR,APS,ARC,ARI,ARJ,ART(5),ASA,ASC(2),ASD,ASE,ASF,ASM,ASO,ASP(3),AST(2),ASV,ASX(3),ATT,AU(2),AVB,AVI,AVR,AVS,AWD,AWR,AXX,A3M,A4M,A4P,A3W,A4W,A5W,BAK,BAS,BAT,BDF,BFC,BG,BGL,BI,BIF,BIFF,BIN,BK;BK\$,BKS,BMK,BMP,BM1,BOOK,BOX,BPL,BQY,BRX,BSC,BSP,BS1,BS\_,BTM,BUD,BUN,BW,BYU,B4,C,C01,CAB,CAD,CAL(2),CAM,CAP,CAS,CAT,CB,CBI,CCA,CCB,CCF,CCH,CCM,CCO,CCT,CDA,CDF,CDI,CDR(2),CDT,CDX(2),CEL,CER,CFB,CFG,CFM,CGI,CGM,CH,CHK,CHM,CHR,CHP,CHT(2),CIL,CIM,CIN,CK1,CK2,CK3,CK4,CK5,CK6,CLASS,CLL,CLP,CLS,CMD(3),CMF,CMP(2),CMV,CMX,CNF,CNM,CNQ,CNT,COB,COD,COM,CPD(2),CPE,CPI,CPL(2),CPO,CPP,CPR,CPT,CPX,CRD,CRP,CRT,CSC,CSP,CSS,CST,CSV,CT(2),CTL,CUE,CUR,CUT,CV(2),CWK,CWS,CXT,CXX,DAT(3),DB,DBC,DBF,DBX(2),DCM,DCR,DCS,DCT,DCU,DCX(3),DC5,DDF,DDIF,DEF(2),DEFI,DEM,DER,DEWF,DGN,DIB,DIC,DIF,DIG(2),DIR,DIZ,DLG,DLS,DLL,DMF,DOC(5),DOT,DPL,DPR,DRAW,DRV,DRW,DSF,DSG,DSM,DSP,DSQ,DST,DSW,DTD,DTED,DTM,DTF,DUN,DV,DWD,DWG(2),DXF(2),DXR,EDA,EDE,EDD,EDK,EDQ,EDS,EDV,EFA,EFE,EFK,EFQ,EF3,EFV,EMD,EMF,EML,ENC,ENFF,EPHTML,EPS,ER1,ERR,ERX,ESPF,ESPS,EUI,EVY,EWL,EXC,EXE,F,F2R,F3R,F77,F90,FAR,FAV,FAX,FBK,FCD,FDB,FDF,FEM,FFA,FFF,FFL,FFO,FFT,FFX,FH3,FIF,FIG,FITS,FITS,FLA,FLC,FLF(3),FLI,FLT(3),FM,FMB,FML,FMT(2),FMX,FND,FNG,FNK,FOG,FON,FOR,FOT,FP,FP1,FP3,FPT(2),FPX,FRM(6),FRX(2),FRT,FSF,FSL(2),FSM,FT,FTG,FTS,FW2,FW3,FW4,FXP,FZB,FZF,FZV,G721,G723,GAL,GCD,GCP,GDB,GDM,GED(2),GEM,GEN,GetRight,GFC,GFI,GFX,GID,GIF,GIM,GIX,GKH,GKS,GL,GNA,GNT,GNX,GRA,GRD,GRF,GRP,GSM(4),GTK,GT2,GWX,GWZ,GZ,H,HCOM,HDF,HED,HEL,HEX,HGL,HH,HLP(2),HOG,HPJ,HPP,HQX,HST,HT,HTM,HTML,HTT,HTX,HXM,ICA,ICB,ICC,ICL,ICM,ICO,IDB,IDD,IDF,IDQ,IDX(3),IFF,IGES,IGF,IIF,ILBM,IMA,IMG,INC,INF,INI(3),INP,INRS,INS(4),INT,IOF,IQY,ISO,ISP,IST,ISU,IT,ITI,ITS(2),IV,IW,J62,JAR,JAVA,JBF,JFF,JIF,JFIF,JMP,JN1,JPE,JPEG,JPG,JS,JSP,JTF,K25,KAR,KDC,KEY,KFX,KIZ,KKW,KMP,KQP,KR1,KRZ,KSF,KYE,LBM,LBT,LBX,LDB,LDL,LEG,LES,LFT,LHA,LIB,LIN,LIS,LLX,LNK,LOG,LPD,LRC,LSL,LSP,LST,LU,LVL,LWLO,LWOB,LWP,LWSC,LYR,LZH,LZS,M1V,M3D,M3U,MAT,MAC,MAD,MAF,MAG,MAGIC,MAK,MAM,MAN,MAP(2),MAQ,MAR,MAS,MAT,MAUD,MAX(3),MAZ(2),MB1,MBOX,MBX,MCC(2),MCR,MCW,MDA(2),MDB,MDE,MDL(2),MDN,MDW,MDZ,MED,MER,MET,MGF,MHTM,MHTML,MI,MIC,MID,MIF,MIFF,MIM,MIME,MLI,MME,MMF(2),MMM,MMP,MN2,MND,MNI,MNG,MNT,MNX,MOD(3),MOV,MP2,MP3,MPA,MPE,MPEG,MPG,MPP(2),MPR,MP2,MP3,MRI,MSA,MSDL,MSG,MSN(2),MSP,MTM,MUL,MUS,MUS10,MVB,MWP,NAN,NAP,netCDF,NCB,NCD,NCF(2),NDO,NFF,NIL,NIST,NLB,NLM,NLU,NSF,NS2,NST,NTF,NWC,NWS,O01,OBD(2),OBJ,OBZ,OCX,ODS,OFF,OFN,OFT,OKT,OLB,OLE,OOGL,OPL,OPO,OPT,OPX,ORC,ORG,OR2,OR3,ORA,OSS,OST,OTL,P10,P65,P7C,PAB,PAC,PAK,PAL,PAS,PAT(3),PBD,PBF,PBK,PBL,PBM,PBR,PCD(2),PCE,PCL,PCM(2),PCP,PCS,PCT,PCX,PDF(2),PDB,PDQ,PF,PFA,PFB,PFC,PFM,PGL,PGM,PGP,PH,PHTML,PIC(3),PICT,PIF(2),PIG,PIN(2),PIX,PJ,PJX,PJT,PKG,PKR,PL,PLG,PLI,PLM,PLS(2),PLT(3),PM5,PM6,PNG(2),PNT,PNTG,POG,POT,POV,PP4,PPA,PPF,PPM,PPP(2),PPS(2),PPT,PRC,PRE,PRF(2),PRG(2),PRJ,PRN(2),PRS,PRT,PRV,PRZ,PS,PSB,PSD,PSI,PSM(2),PST,PTM,PUB(2),PWD,PWL,PWP,PWZ,PXL,PY,QAD,QBW,QDT,QD3D,QFL,QIC,QIF(2),QLB,QM,QRY,QST,QT,QTI,QTIF,QTQ,QTP,QTS(2),QTX,QW,QXD,R,RA,RAM,RAR,RAS,RAW(3),RBH,RDF,RDL,REC(2),REG,RES,RGB;SGI,RFT,RLE,RL2,RM,RMD,RMF,RMI,ROV,RPM,RPT,RRS,RSL,RTF,RTM,RTK,RTS(2),RUL,RVP,RXX,S,S3I,S3M,SAM(2),SAV,SB,SBK,SBL,SC2(2),SCC,SCD(2),SCF,SCH,SCI,SCN,SCP,SCR(2),SCT(3),SCT01,SCV,SCX,SD,SD2(2),.....



# Obsolescence of software

```
H:INTRO PAGE 1 LINE 9 COL 11 INSERT ON  
      < < < MAIN MENU > > >  
--Cursor Movement-- | -Delete- | -Miscellaneous- | -Other Menus-  
^S char left ^D char right | ^G char | ^I Tab ^B Reform | (from Main only)  
^A word left ^F word right | DEL chr lf | ^V INSERT ON/OFF | ^J Help ^K Block  
^E line up ^X line down | ^T word rt | ^L Find/Replce again | ^Q Quick ^P Print  
      --Scrolling-- | ^Y line | RETURN End paragraph | ^O Onscreen  
^Z line down ^W line up | | ^N Insert a RETURN |  
^C screen up ^R screen down | | ^U Stop a command |  
!----!----!----!----!----!----!----!----!-----R
```

### 1. Introducing WordStar

WordStar is highly flexible and very visible. Watch the screens as you give commands, and information in various parts of the screen will guide you. You won't see all the information all the time, but it will be there when you need it.

#### WHERE YOU ARE

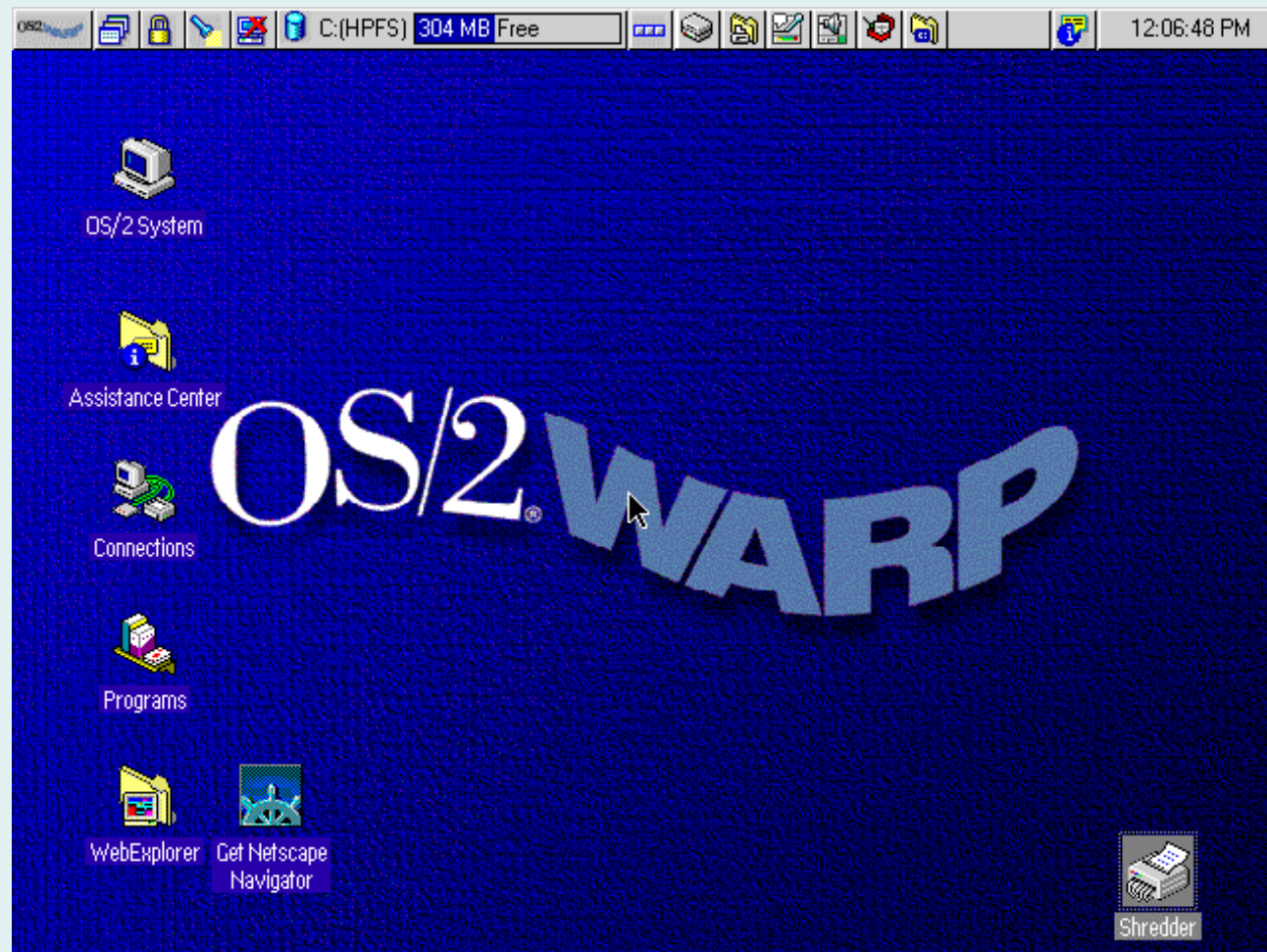
The seven WordStar menus are your greatest aids. They are like signposts at the top of your screen, showing you where you are.

```
1HELP 2INDENT 3SET LM 4SET RM 5UNDLIN 6BLDFCE 7BEGBLK 8ENDBLK 9BEGFIL 10ENDEFIL
```





# Obsolescence of operating system



OS/2 Warp 4.0 (1996)



# 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
  - Emerging technologies
  - Migration
  - Emulation



# Emerging technologies



# Emerging Technologies

---

- ❑ Context-aware Digital Entities
- ❑ Binary Code Translation
- ❑ Database Preservation Technologies
- ❑ Personal Archiving





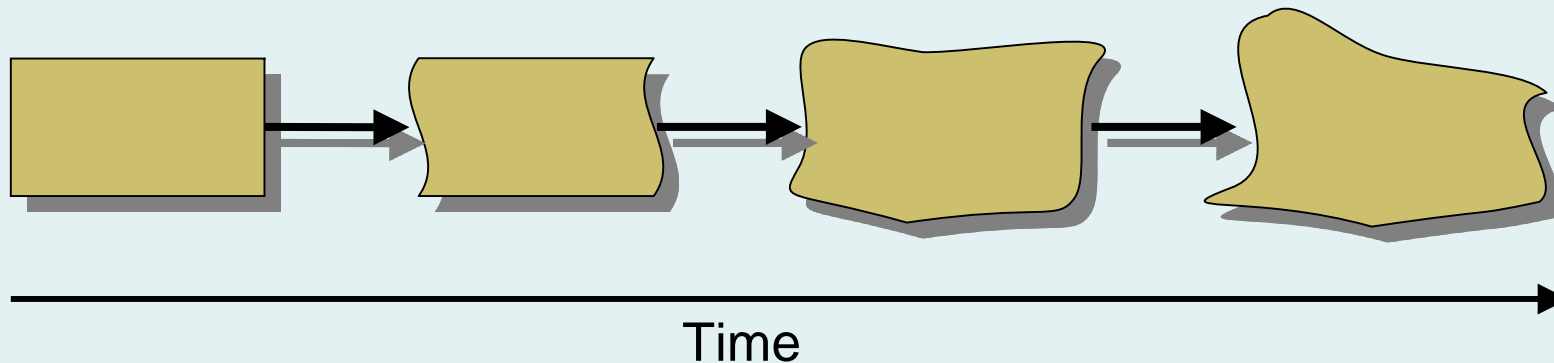
# Migration



# Migration

**Migration** = changing the object

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

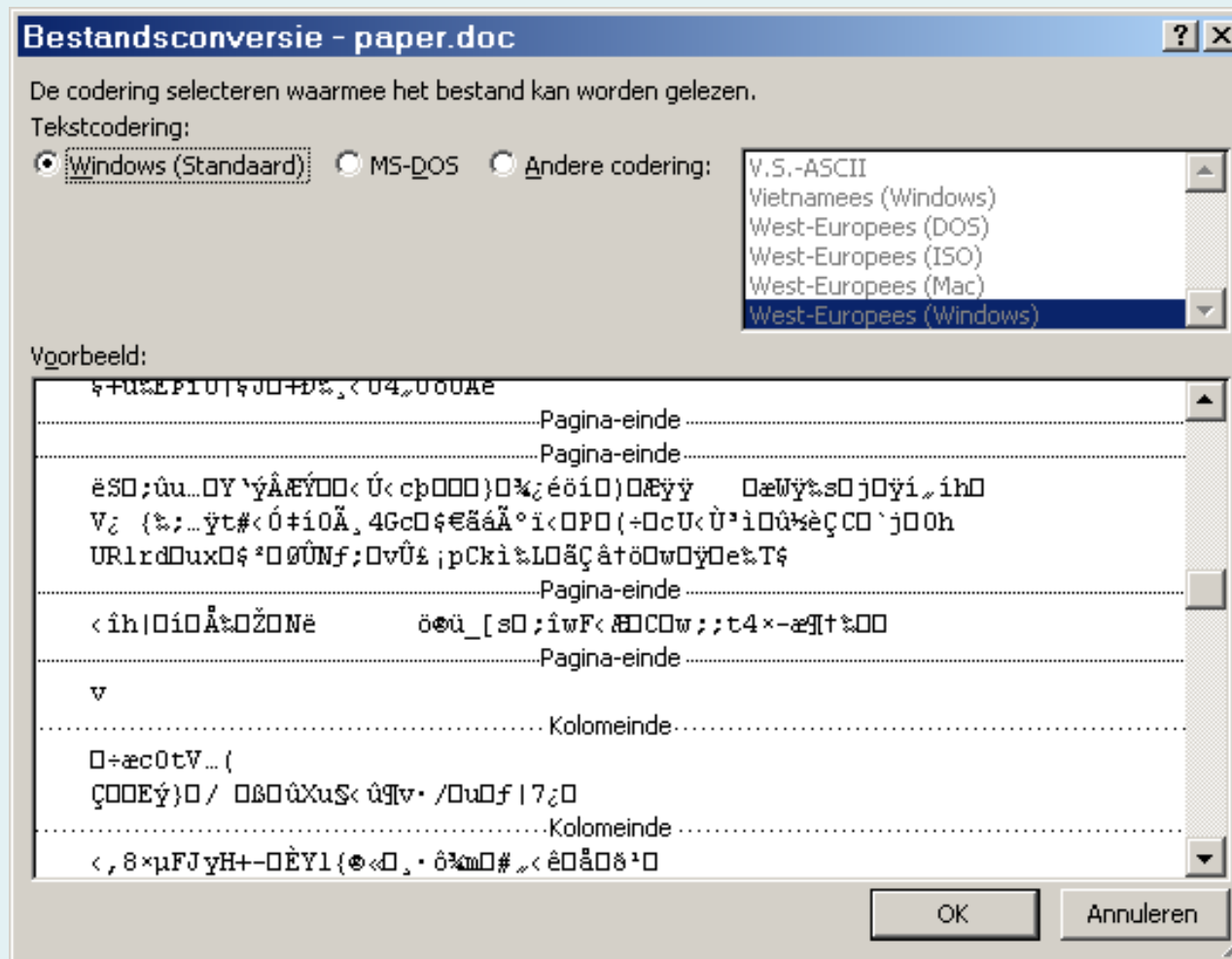


Risks:

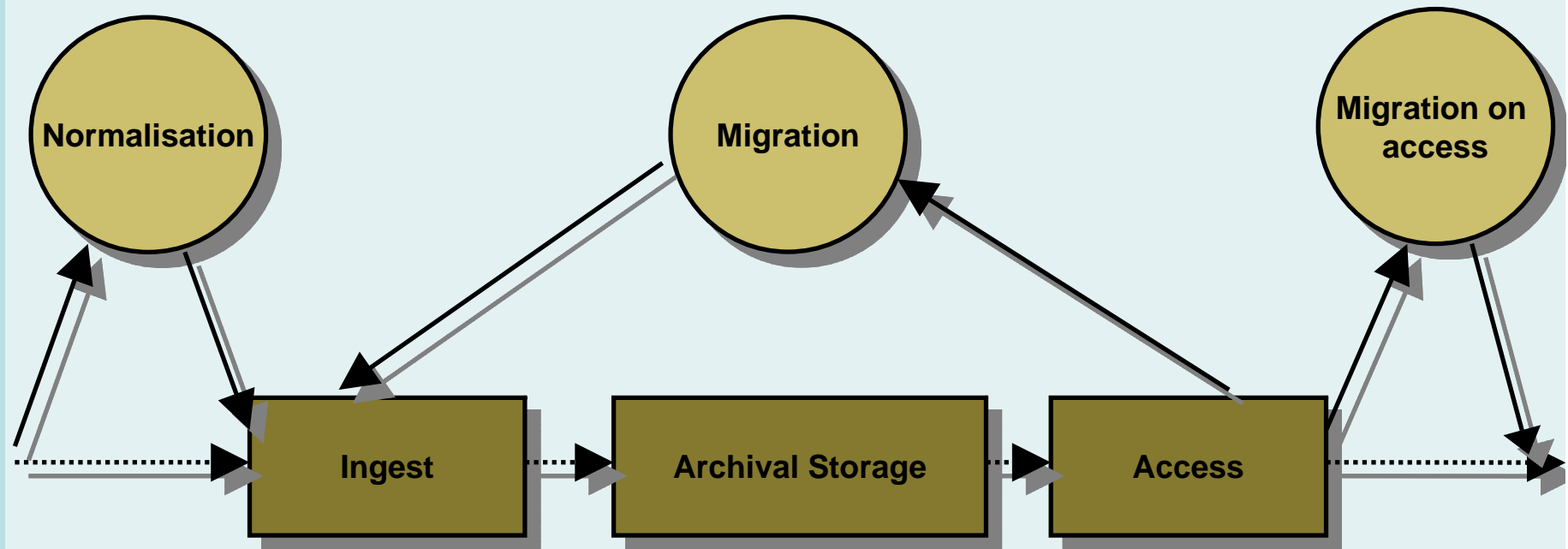
- Inconsistencies can occur
- Functionality can be lost
- Quality Assessment is difficult



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



# Migration and digital archiving



# Emulation





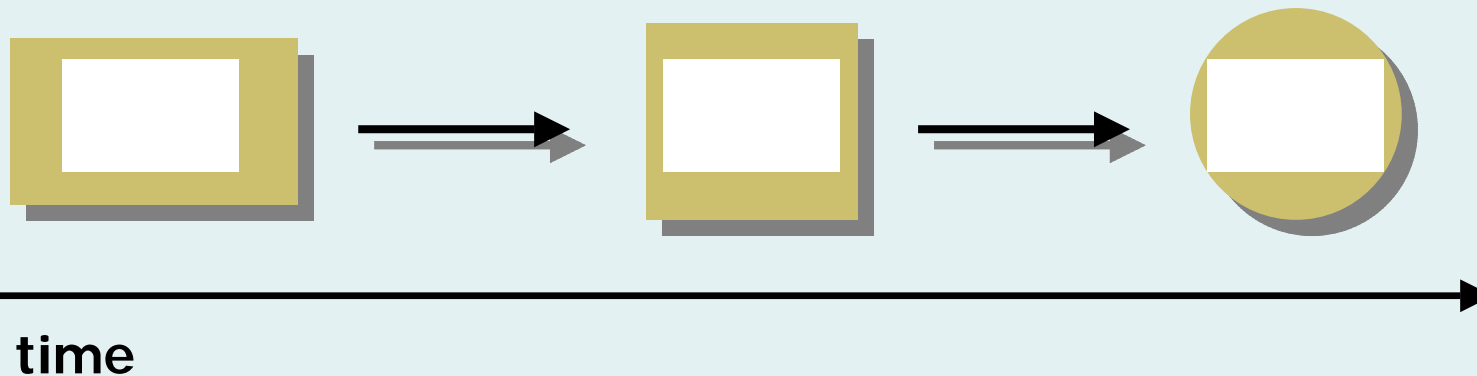
# Emulation – Windows 2000 in Windows XP environment



# Emulation

**Emulation** = changing the environment

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



Risks:

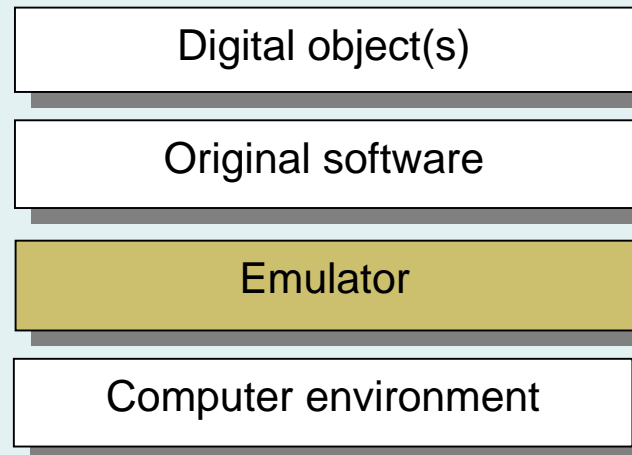
- Technically challenging
- User has to have knowledge about original environment



# 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



The screenshot displays the DOS 6.22 boot screen. The command menu is visible, with the 'Brief' option highlighted by a red circle. The file list shows various files and their attributes. An inset window in the bottom right corner shows a file list with columns for Date and Time.

Left	Files	Commands	Options
Brief			
Full			
Info			
Tree			
On/Off		Ctrl-F1	
Name			
eXtension			
tiMe			
Size			
Unsorted			
Re-read			
Drive...		Alt-F1	

Date	Time
11-06	18:33
10-06	9:45
11-06	15:29
02-07	13:04
03-06	10:53
12-05	15:14
12-05	15:14
08-06	17:11
12-05	11:59
12-05	20:20
07-06	15:21

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

Left	Files	Commands	Date	Time
	rief		-11-06	18:33
	ll		-10-06	9:45
	ifo		-11-06	15:29
D	Free		-02-07	13:04
E	On/Off	Ctrl-F1	-03-06	10:53
E			-12-05	15:14
I	1 Name		-12-05	15:14
P	extension		-08-06	17:11
R	time		-12-05	11:59
S	Size		-12-05	20:20
n	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



Start 1 2 3 4 x xmess (SDL) version 0.56...

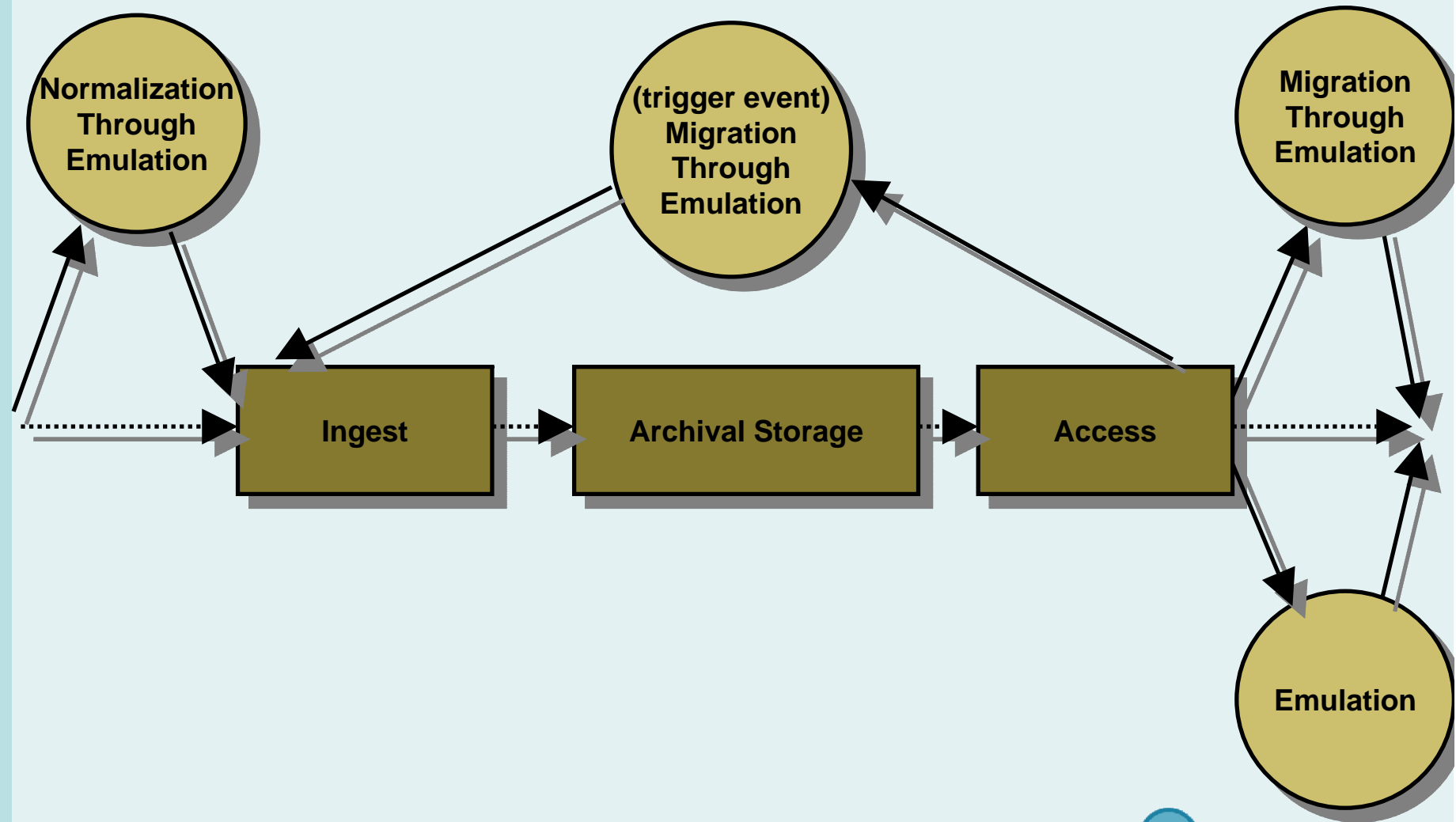
# Migration through Emulation

---

- ❑ 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



# Gap analysis in tool provision

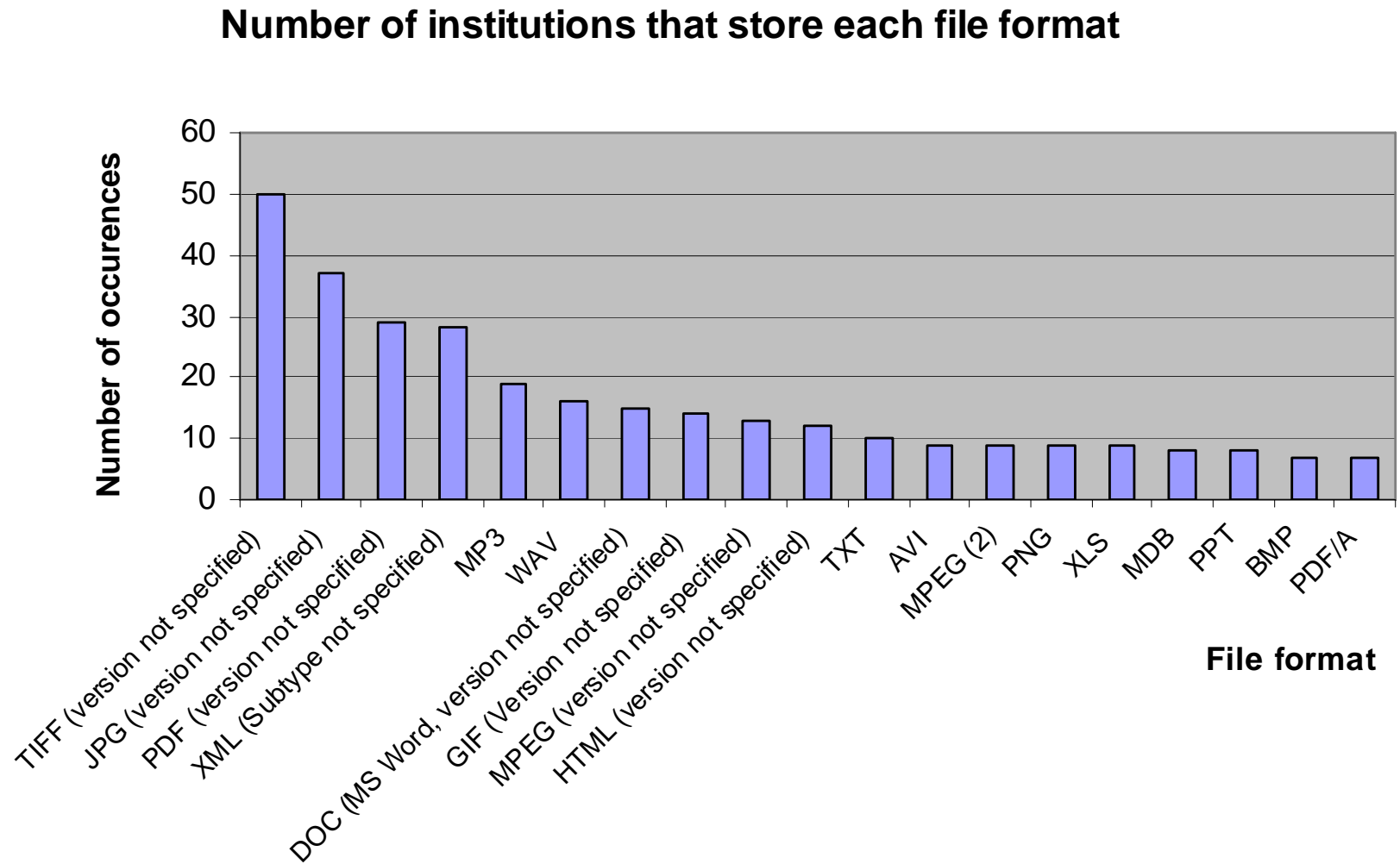
- ❑ Inventory of file formats in existing collections, archives and repositories
- ❑ Insight into the demand for preservation action tools
- ❑ 107 different file formats, 65 institutions
- ❑ Most archived file formats in three main types of institutions

	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



# 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



# Preservation action tools in Planets

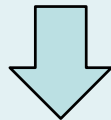
Services



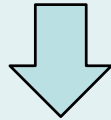
# Preservation Action Tools and Planets

---

- ❑ 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



# Planets Core Registry

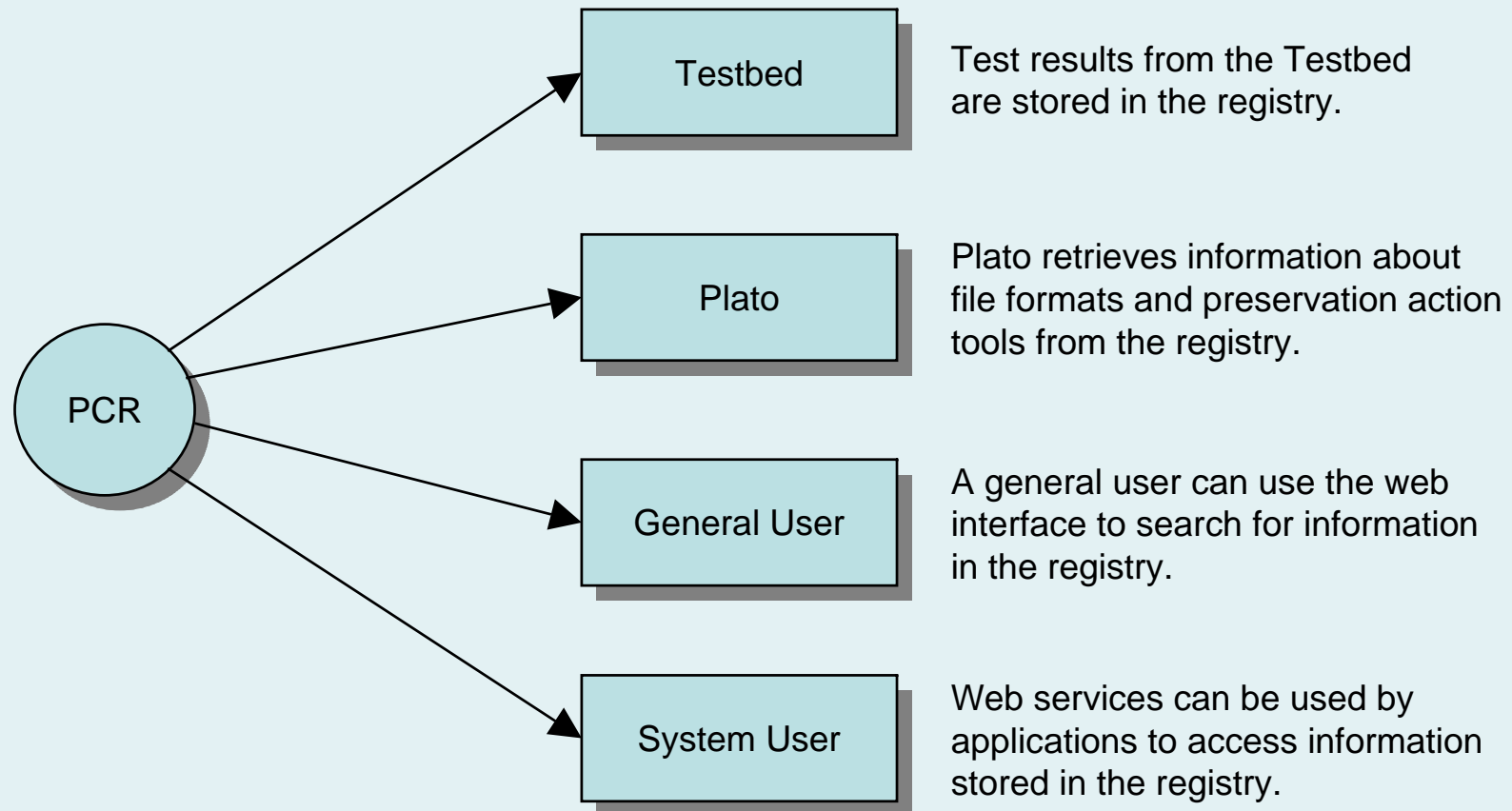
---

- ❑ 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



# 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

