



Preservation Actions – Sara van Bussel

The National Library of The Netherlands



# Preservation Actions

---

## ❑ Types of preservation action

- Migration
  - SIARD
- Emulation
  - Dioscuri
  - GRATE

## ❑ Planets Core Registry



# Types of preservation action

Migration



# Migration Tools

---

- ❑ Migration tools are available through the Planets Framework
- ❑ Can be invoked from:
  - Core Registry
  - Plato
  - Testbed
  - IF Framework



# Types of preservation action

Migration - SIARD



# SIARD - Software-Independent Archiving of Relational Databases

---

- ❑ Archiving databases today means mainly archiving relational databases
- ❑ Relationships must be preserved
- ❑ Field values must be converted which means field should have:
  - No codes
  - No encryption
  - Data type must be suitable for archiving



# SIARD - The SIARD Format

---

- ❑ The SIARD format stores database content
- ❑ A SIARD file is a ZIP file (ZIP64) containing XML files
- ❑ One XML file documents the metadata (based on SQL:1999)
- ❑ The other XML files contain the table data
- ❑ The SIARD file format is based on open standards:
  - SQL:1999, XML, XML Schema, UNICODE



# SIARD -Specifications

---

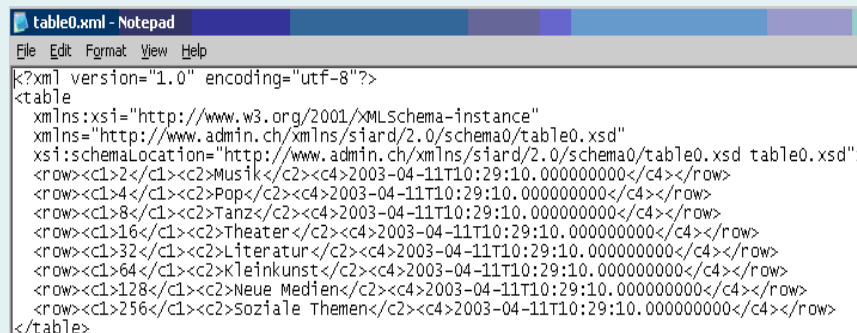
- ❑ SIARD is platform independent
- ❑ It operates in a JAVA 1.5 environment
- ❑ SIARD can run on a single computer with a common GUI
- ❑ Installation
  - Use SIARD directly from a USB stick or simply click & install SIARD
- ❑ Databases
  - Oracle, MS SQL, MS Access





# SIARD - Primary data & Metadata

- ❑ **Primary data** is stored in the folder content
- ❑ This data will be stored in a XML file format
- ❑ For every table in the database SIARD automatically generates a separate XML file
- ❑ **Metadata** is stored one single file: **metadata.xml**
- ❑ An XML storage file can make it possible to search within the tables. Specific database information can be extracted according to your needs



```
<?xml version="1.0" encoding="utf-8"?>
<table
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="http://www.admin.ch/xmlns/siard/2.0/schema0/table0.xsd"
  xsi:schemaLocation="http://www.admin.ch/xmlns/siard/2.0/schema0/table0.xsd table0.xsd">
  <row><c1>2</c1><c2>Musik</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>4</c1><c2>Pop</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>8</c1><c2>Tanz</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>16</c1><c2>Theater</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>32</c1><c2>Literatur</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>64</c1><c2>Kleinkunst</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>128</c1><c2>Neue Medien</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
  <row><c1>256</c1><c2>soziale Themen</c2><c4>2003-04-11T10:29:10.000000000</c4></row>
</table>
```

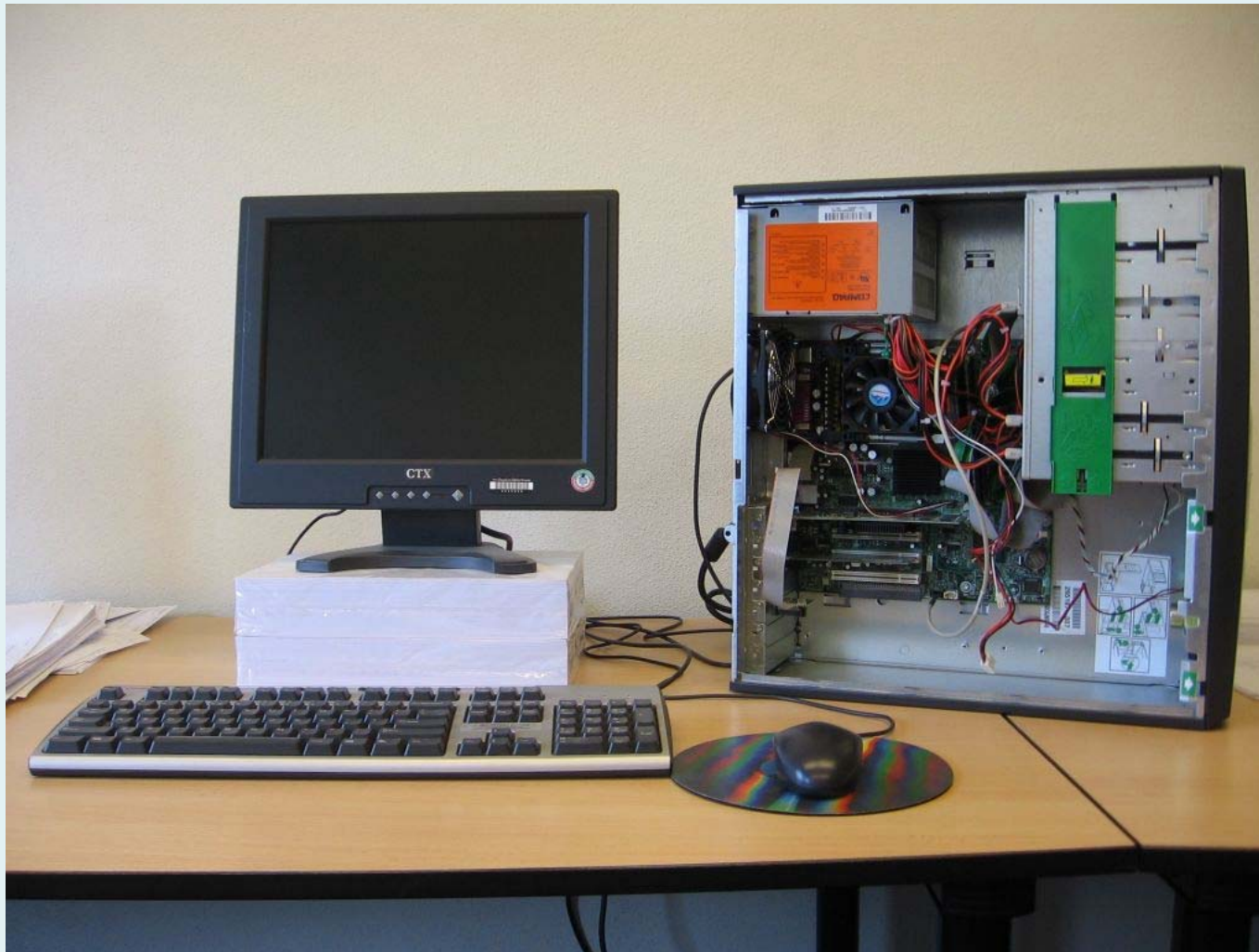


# Types of preservation action

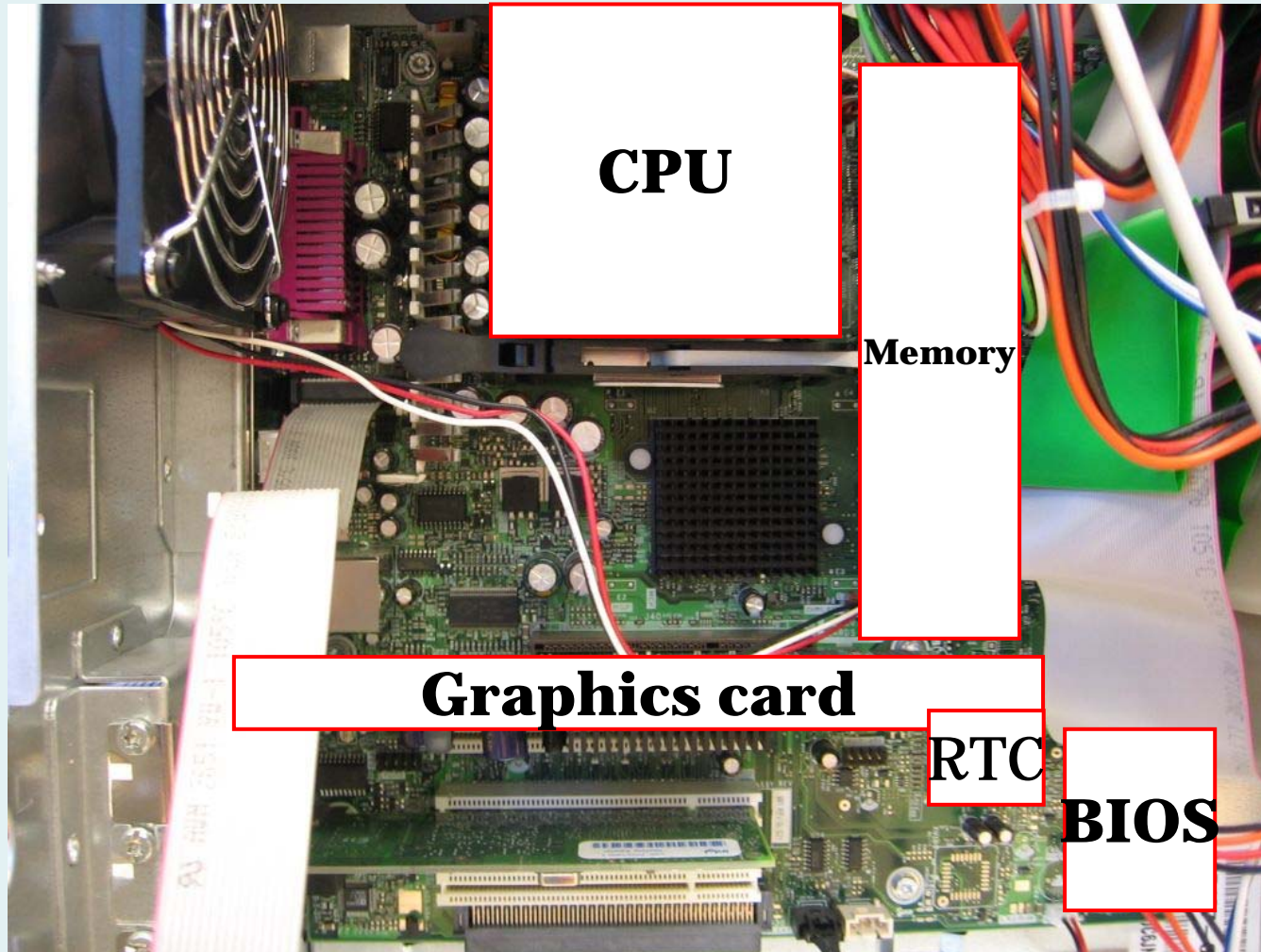
Emulation - Dioscuri



# Dioscuri - Original source PC

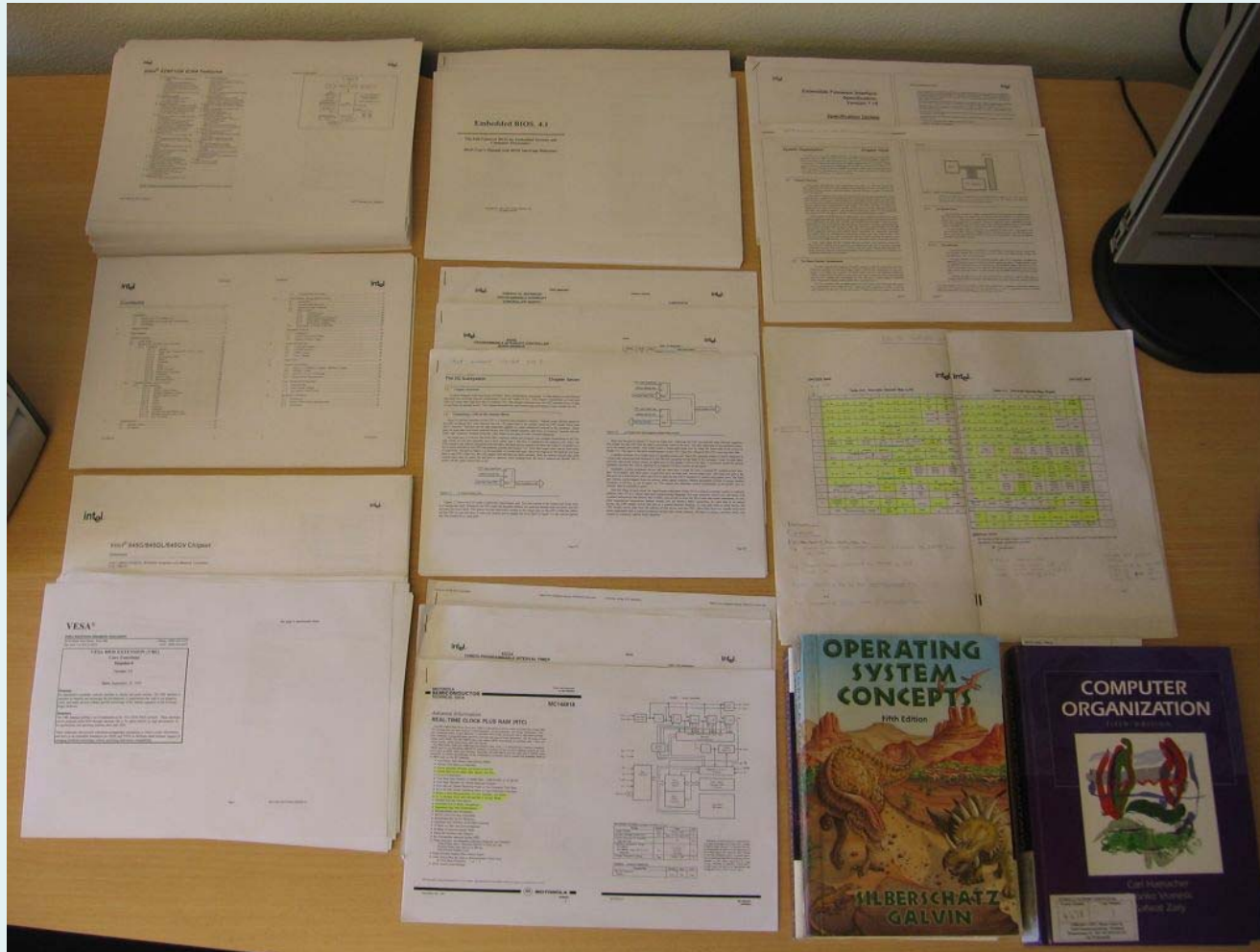


# Dioscuri - Original source PC





# Dioscuri - Documentation



# Dioscuri - Results

- ❑ Modular emulator for digital preservation
- ❑ Programmed in Java using JVM
- ❑ Current version: 0.4.0 (Beta)
- ❑ Easy to use interface (GUI) offering process control, media management and configuration editor



# Dioscuri - Specifications

---

## ❑ Virtual machine hardware

- 16-bit Intel 8086-based CPU, DMA-support, IRQ-handling
- 1 MB RAM
- Storage devices: floppy, HDD
- Input devices: keyboard
- Output devices: VGA, screen
- System BIOS: Bochs BIOS
- VGA BIOS: Plex86/Bochs VGA BIOS

## ❑ Host Platforms

- Windows XP, Linux Fedora Core 4, Sun Sparc Solaris, all running JRE v1.5.x



# Dioscuri - Specifications

---

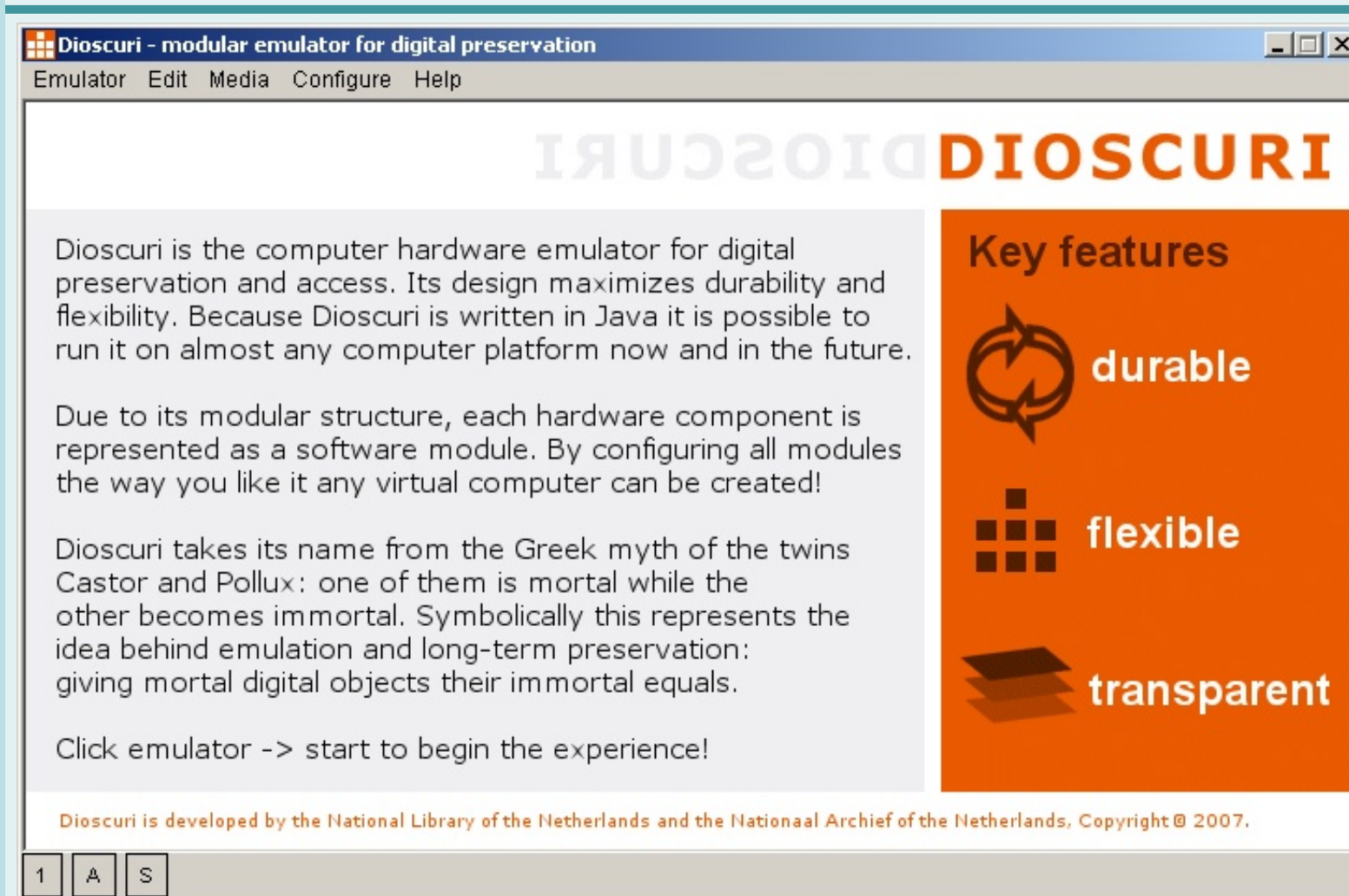
## ❑ Capable of:

- Running MS-DOS, FreeDOS, Linux 16-bit (ELKS), MS Windows 3.0
- Norton Commander 3.0, WordPerfect 5.1, DrawPerfect 1.1, many games like PC-versions of PacMan, Tetris, Chess and many more. Also DOS-based webbrowser Arachne
- XML-based module configuration
- Text extraction from emulated environment into the clipboard of host computer.

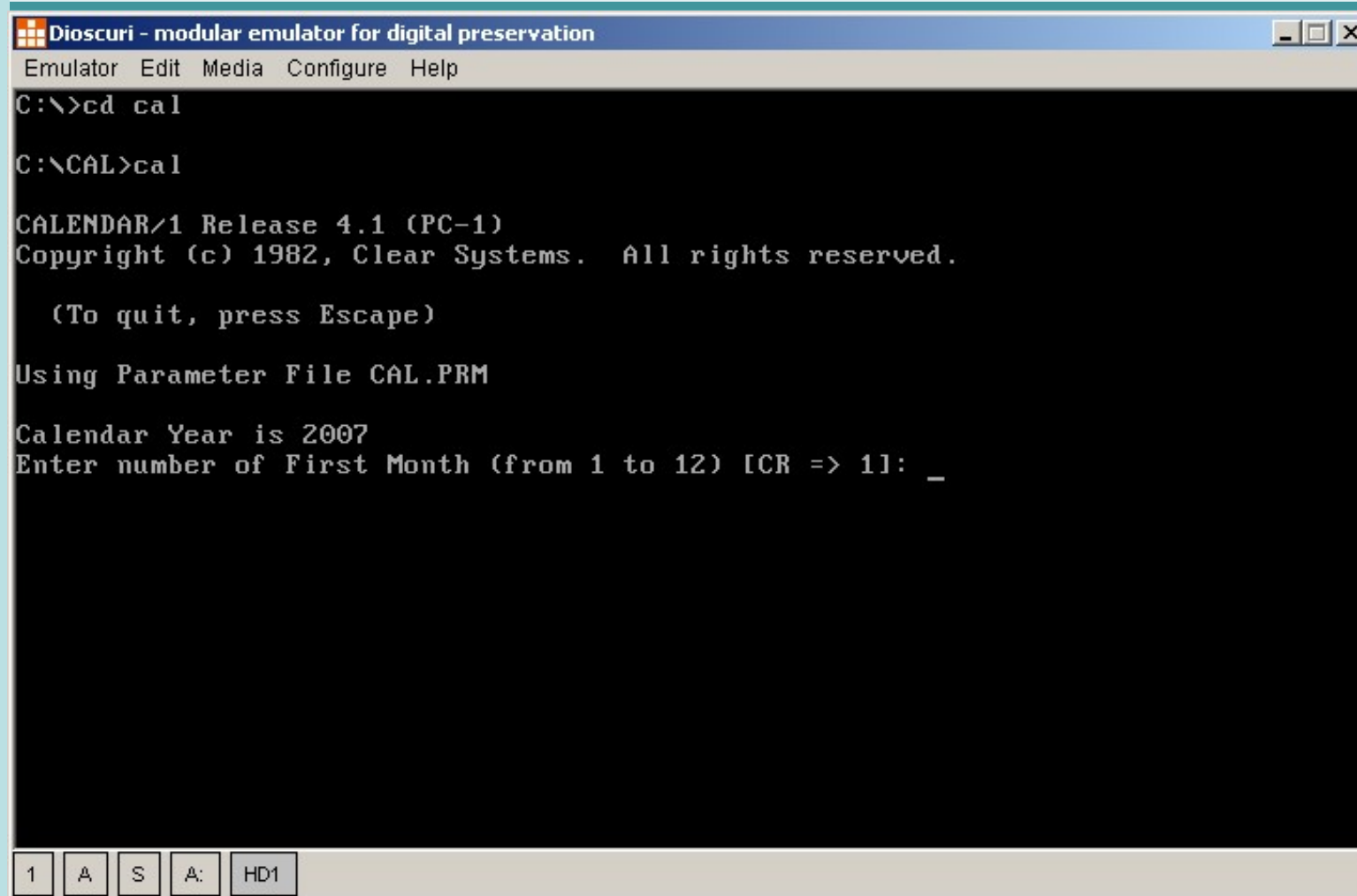




# Dioscuri – Start up screen



# Dioscuri – MSDos running Calendar



The screenshot shows a window titled "Dioscuri - modular emulator for digital preservation". The menu bar includes "Emulator", "Edit", "Media", "Configure", and "Help". The command prompt shows the following sequence of commands and output:

```
C:\>cd cal  
C:\CAL>cal  
CALENDAR/1 Release 4.1 (PC-1)  
Copyright (c) 1982, Clear Systems. All rights reserved.  
  
  (To quit, press Escape)  
  
Using Parameter File CAL.PRM  
Calendar Year is 2007  
Enter number of First Month (from 1 to 12) [CR => 1]: _
```

At the bottom of the emulator window, there is a status bar with five buttons labeled "1", "A", "S", "A:", and "HD1".



# Dioscuri – Running Ironman



# <http://dioscuri.sourceforge.net> for more information and download

## Dioscuri - the modular emulator

Dioscuri is an x86 computer hardware emulator written in Java. It is designed by the digital preservation community to ensure documents and programs from the past can still be accessed in the future.

The Dioscuri emulator has two key features: it is durable and flexible. Because it is implemented in Java, it can be ported to any computer platform which supports the Java Virtual Machine (JVM), without any extra effort. This reduces the risk that emulation will fail to work on a single architecture in the future, as it will continue to work on another architecture.

Dioscuri is flexible because it is completely component-based. Each hardware component is emulated by a software surrogate called a module. Combining several modules allows the user to configure any computer system, as long as these modules are compatible. New or upgraded modules can be added to the software library, giving the emulator the capability to run these.

*Dioscuri is the best choice to retain access to your old documents, games and other applications!*

## Latest news

11 December 2008

A new release of Dioscuri has been made available by the Dioscuri project team. The new version (0.4.0) offers the following improvements:

- Added backwards compatibility with JRE 1.5
- Added command-line interface
- Improved 32-bit CPU
- Fixed minor bugs in modules CPU
- Updated GUI

For an overview of the full change history, please check the changelog.

 **Download: Dioscuri version 0.4.0**



## Dioscuri

- [Idea and key features](#)
- [Digital Preservation](#)
- [Screenshots](#)
- [Latest news!](#)

## Downloads

- [Latest version](#)
- [All versions / sourcecode](#)
- [Disk images](#)

## Documentation

- [User manual](#)
- [Reference docs](#)
- [Javadoc](#)
- [Changelog](#)

## Support

- [FAQ](#)
- [Forum](#)

## Development

- [Buglist \(tracker\)](#)
- [Feature requests \(tracker\)](#)
- [Roadmap](#)

## Contact

- [About project team](#)
- [Join development!](#)
- [Mailinglist](#)



# Types of preservation action

Emulation - GRATE



# GRATE - Emulation as a Service

---

- ❑ Emulation requires some effort until object of interest is actually accessible
  - Average archive user is often not trained computer professional
  - Range of problems to setup emulation environment on average machines (with unknown software environment)
  - Many software components needed are proprietary
- ❑ GRATE provides Global Remote Access To Emulation
  - PRONOM detection of object type
  - Recommended view path is provided
  - By clicking on an URL the emulator is opened with the object



# GRATE – Global Remote Access To Emulation

---

- ❑ Global remote access to emulation (services)
  - Client side: Java application executable in average browsers with JRE 1.5
  - Server side: Standard Linux environment to host the several emulators
  - Running different emulators like Dioscuri, QEMU, MESS, ...
  - Extensible to more emulators, environments
  - Up- and download of objects
  - Object transport via virtual floppies





# GRATE - Data Exchange for Emulation Environments

---

- ❑ Major issue: Object transport into emulation environment (and out of it)
- ❑ Means of object transport, depending on emulator used
  - Virtual optical (ISO) or floppy disks as images
  - Network connections like FTP, SMB/CIFS
  - “Shared Folders”
  - Copy & Paste





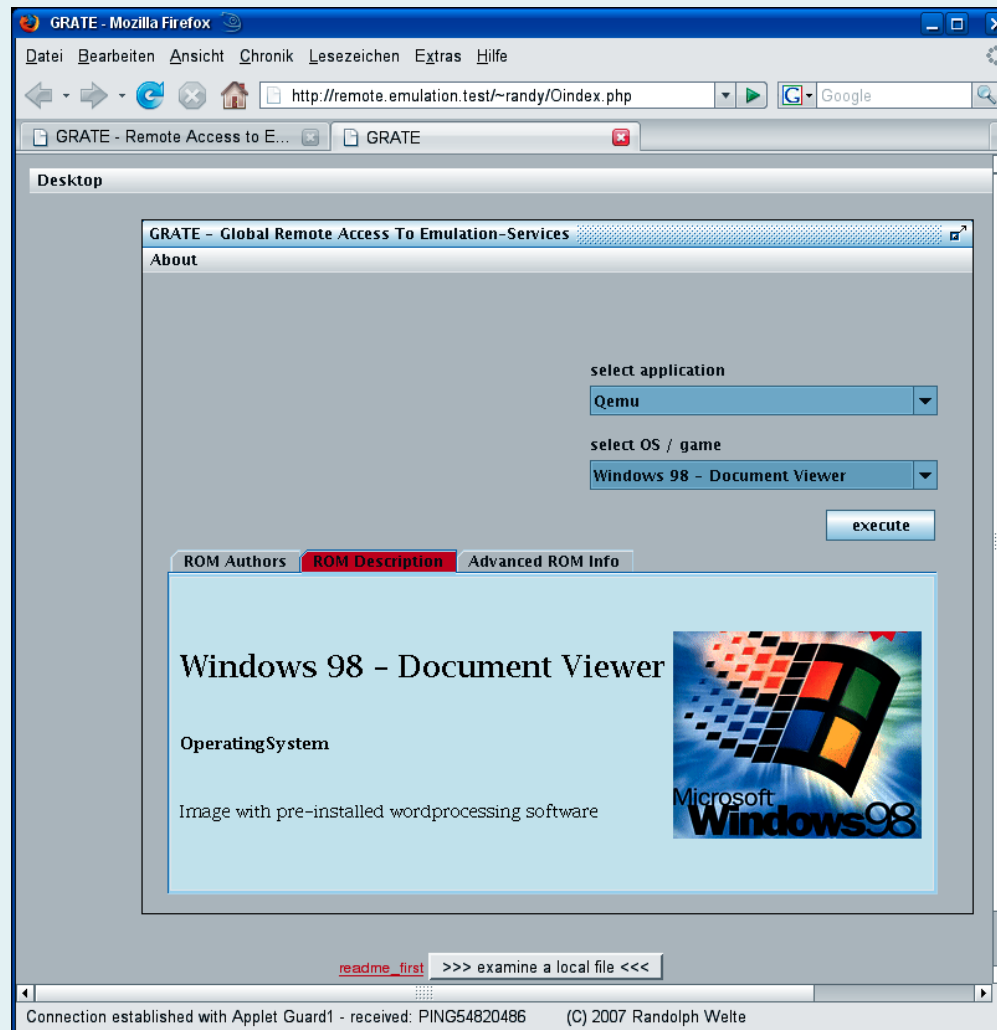
# Emulation Examples in GRATE

---

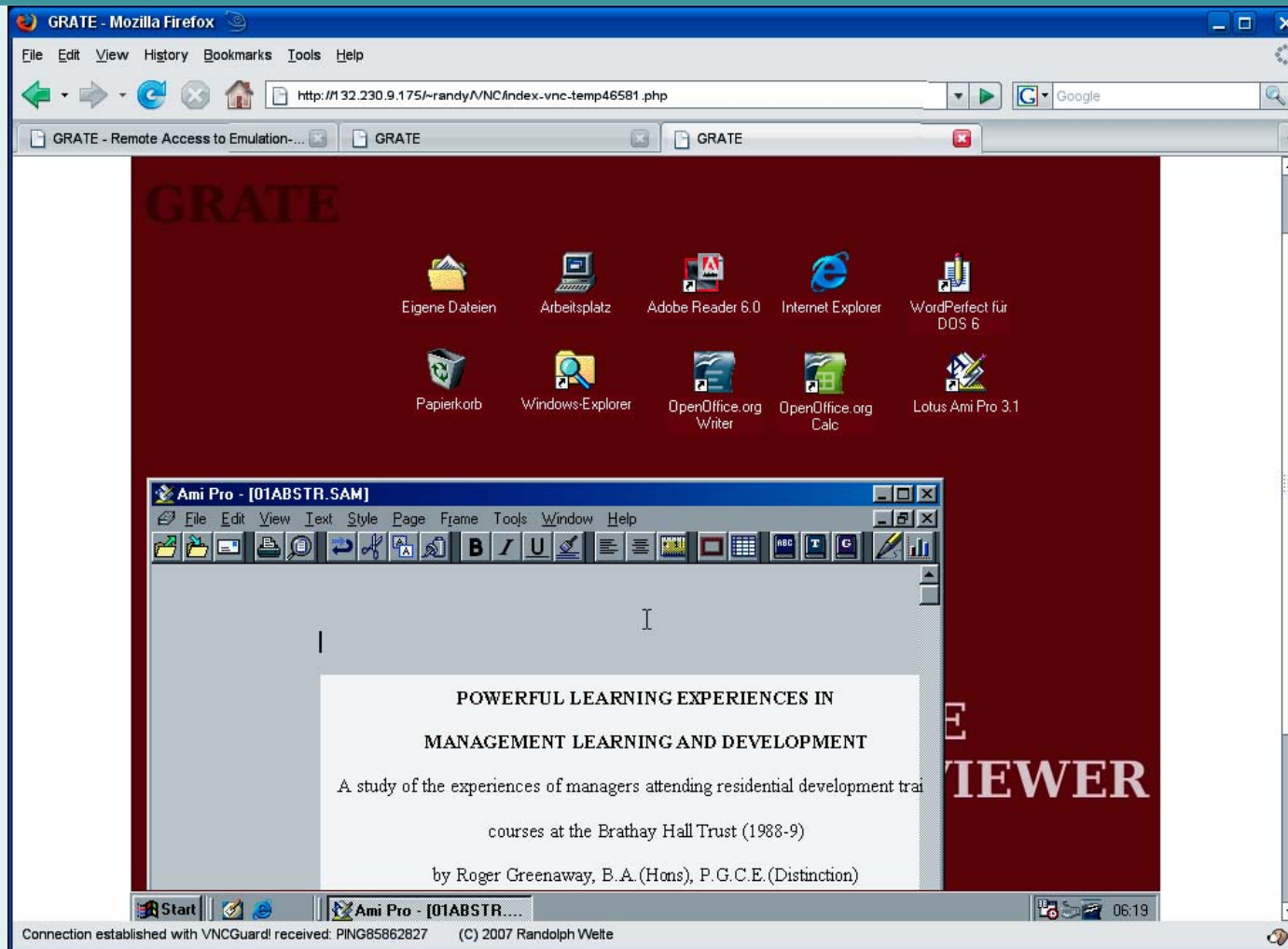
- ❑ Dioscuri X86 emulator recreating an 286, 386 PC of the early 1990s
  - Java programming language, modular approach – components like disk, floppy, VGA, CPU, RAM put together form the machine
  - Running DOS and Windows 3.0
- ❑ QEMU – popular C programming language multi architecture emulator for X86, PPC, Sparc, ...
  - Used for Windows 3.1 and 95, 98 environments in GRATE
  - MESS, DOSBOX, Hatari, ...



# GRATE – Selection of emulator and service



# GRATE – Running Windows 98 in browser



# Planets Core Registry



# Planets Core Registry

---

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



# Planets Core Registry

---

## ☐ Core entities

[File formats](#) [Software packages](#) [Hardware](#) [Storage media](#) [Character encodings](#) [Compression techniques](#)

## ☐ Subsidiary entities

[Agents](#) [Documents](#) [IPRs](#) [Technical environments](#) [Pathways](#) [Properties](#)

## ☐ Reference objects

[Reference files](#) [Ext. Identifiers](#) [Images](#) [Xcdl files](#) [Xcel files](#) [CMTs](#) [External sigs](#) [Internal sigs](#)

## ☐ Search

[By PUID](#) [By text](#) [By pathway](#)



# Planets Core Registry – File Formats

PUID:	fmt/40						
Name:	Microsoft Word for Windows Document						
Description:	With the release of Word 97, Microsoft revised the native binary word processing format, which is based on its generic OLE2 Compound Document Format. The format is proprietary and Microsoft does not make details of its structure public. The information here is derived primarily from OpenOffice.org's reverse-engineered documentation of the format and should not therefore be regarded as definitive. A Word document is stored as a 'WordDocument' stream within a Compound Document Format file. The format remained unchanged with the releases of Word 2000, 2002 and 2003.						
Version:	97-2003						
Release date:							
Entity note:							
Provenance source:	Digital Preservation Department						
Provenance note:							
Provenance provided date:	11-mrt-2005						
Provenance last updated:	2-aug-2005						
Byte order:	Little-endian						
Disclosure level:	None						
Orientation:	Binary						
Risk score:	0						
Assessments:	<table><tr><th>Name</th><th>Description</th><th>URL</th></tr><tr><td></td><td></td><td></td></tr></table>	Name	Description	URL			
Name	Description	URL					
Component manifestation types:	<table><tr><th>Name</th></tr><tr><td>Microsoft Word Document</td><td>A Microsoft Word document</td></tr></table>	Name	Microsoft Word Document	A Microsoft Word document			
Name							
Microsoft Word Document	A Microsoft Word document						
External signatures:	<table><tr><th>Name</th></tr><tr><td>File extension: doc</td></tr></table>	Name	File extension: doc				
Name							
File extension: doc							



# Planets Core Registry – Software Packages

PUID: sfw/279 [x-sfw/1]  
Name: Word  
Description: No description available  
Version: 97 (8.0) for Windows  
Release date: 1-jan-1997  
Entity note:  
Provenance source: Digital Preservation Department  
Provenance note:  
Provenance provided date: 23-sep-2003  
Provenance last updated: 2-aug-2005  
Local source availability: Unknown  
Local package availability: Unknown  
Service pack level:  
Associated tools:

Tool name	Tool type
-----------	-----------

Processes:

Type	Object	
Render	Binary Interchange File Format (BIFF) Workbook 4W	<a href="#">View</a>
Render	Plain Text File	<a href="#">View</a>
Create	MS-DOS Text File	<a href="#">View</a>
Render	Microsoft Works for Windows 4.0	<a href="#">View</a>
Render	Rich Text Format 1.3	<a href="#">View</a>
Render	Rich Text Format 1.2	<a href="#">View</a>
Render	Microsoft Outlook Address Book	<a href="#">View</a>





# Planets Core Registry – Relationships between entities

## ❑ Relationships of Microsoft Word 97 for Windows

Relationships (source):	<b>Relationship type</b>	<b>Target</b>
	sfw/279 Available On	med/2 CD ROM
	sfw/279 Is previous version of	sfw/280 Word 2000 (9.0) for Windows
Relationships (target):	<b>Target</b>	<b>Relationship type</b>
	fmt/40 Microsoft Word for Windows Document 97-2003	Is default for sfw/279

## ❑ Relationships of Microsoft Word for Windows document

Relationships (source):	<b>Relationship type</b>	<b>Target</b>
	fmt/40 Is default for	sfw/545 Word 2003
	fmt/40 Is default for	sfw/279 Word 97 (8.0) for Windows
	fmt/40 Is default for	sfw/280 Word 2000 (9.0) for Windows
	fmt/40 Has priority over	fmt/111 OLE2 Compound Document Format
	fmt/40 Is subsequent version of	fmt/39 Microsoft Word for Windows Document 6.0/95
Relationships (target):	<b>Target</b>	<b>Relationship type</b>
	fmt/111 OLE2 Compound Document Format	Is supertype of fmt/40
	fmt/39 Microsoft Word for Windows Document 6.0/95	Is previous version of fmt/40



# Planets Core Registry - Pathways

- ❑ Pathways are step by step preservation actions
- ❑ Pathway roles are:
  - Preservation
  - Presentation
  - Extraction

PUID:	pth/11										
Name:	WordPerfect for MS-DOS/Windows Document 6.0 to XML 1.0										
Description:	WordPerfect for MS-DOS/Windows Document 6.0 to XML										
Pathway type:	File Format Migration										
Pathway steps:	<table><thead><tr><th>Name</th><th>Description</th><th>Order</th><th></th></tr></thead><tbody><tr><td>WordPerfect for MS-DOS/Windows Document 6.0 to XML 1.0</td><td>WordPerfect for MS-DOS/Windows Document 6.0 to XML</td><td>1</td><td><a href="#">View</a></td></tr></tbody></table>			Name	Description	Order		WordPerfect for MS-DOS/Windows Document 6.0 to XML 1.0	WordPerfect for MS-DOS/Windows Document 6.0 to XML	1	<a href="#">View</a>
Name	Description	Order									
WordPerfect for MS-DOS/Windows Document 6.0 to XML 1.0	WordPerfect for MS-DOS/Windows Document 6.0 to XML	1	<a href="#">View</a>								
Documents:	<table><thead><tr><th>Document</th></tr></thead><tbody></tbody></table>			Document							
Document											
Pathway roles:	<table><thead><tr><th>Role</th><th>Status</th></tr></thead><tbody><tr><td>Preservation</td><td>Unknown</td></tr></tbody></table>			Role	Status	Preservation	Unknown				
Role	Status										
Preservation	Unknown										



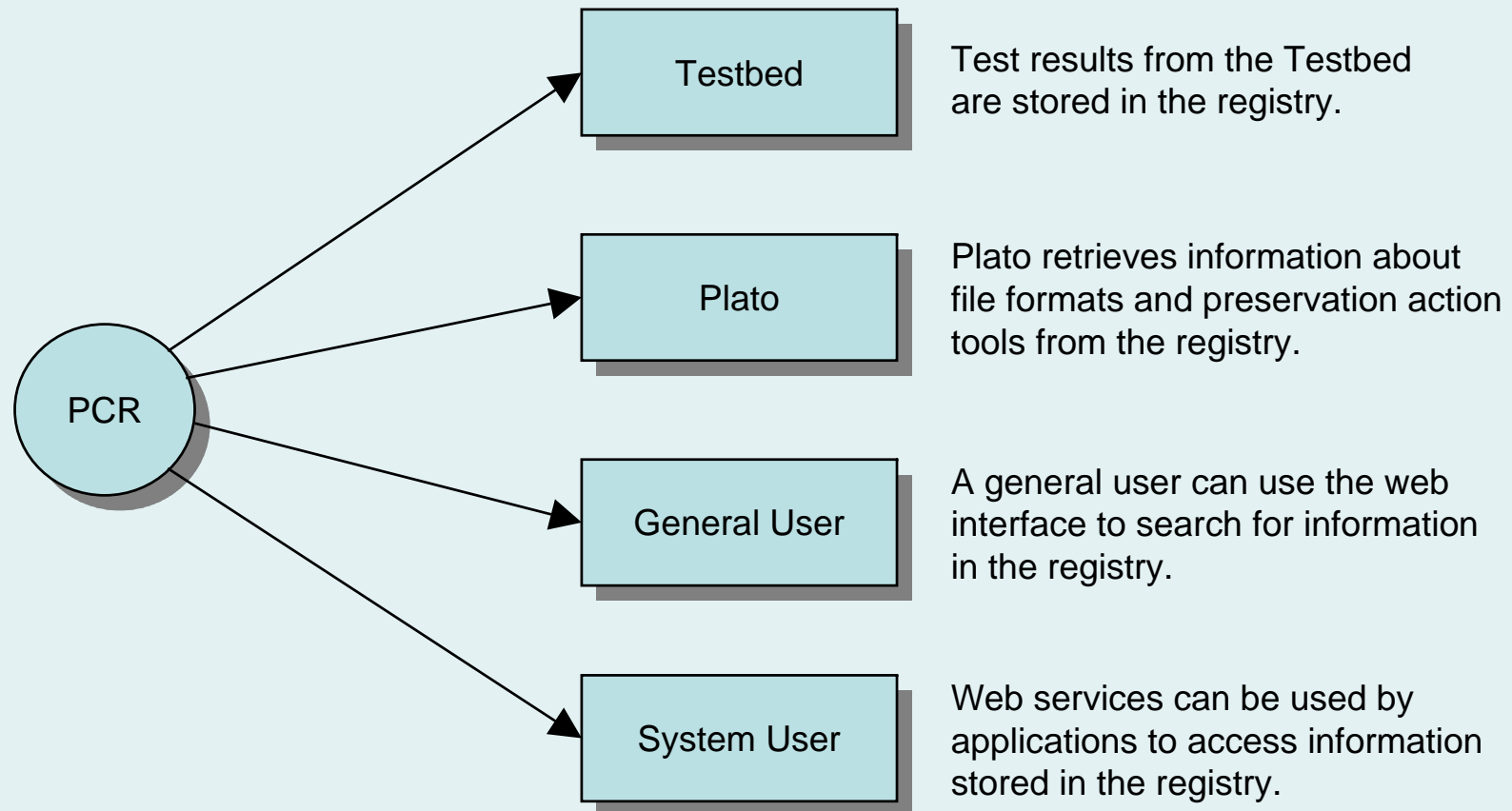
# Planets Core Registry - Pathways

- ❑ Separate actions are described in Pathway Steps
  - Related to Software Package, a specific service and invocation
  - Can contain source and target file format
  - Specific properties of the step can be described

Name:	WordPerfect for MS-DOS/Windows Document 6.0 to XML 1.0						
Description:	WordPerfect for MS-DOS/Windows Document 6.0 to XML						
Sequence position:	1						
Software package:	SoftwareToolsPackage						
Software package tool:	xmlExportWrapper						
Software package tool invocation:	File Format Migration						
Source file format:	fmt/474 WordPerfect for MS-DOS/Windows Document 6.0						
Max file size (bytes):							
Target file format:	fmt/101 Extensible Markup Language 1.0						
Properties:	<table><tr><th>Source</th><th>Type</th><th>Target</th><th>Type</th><th>Variance</th><th>Notes</th></tr></table>	Source	Type	Target	Type	Variance	Notes
Source	Type	Target	Type	Variance	Notes		



# Relationships of the Core Registry



# Testbed

---

- ❑ Test results will be integrated into the Planets Core Registry, on Pathway records
- ❑ Test results will be accessible from:
  - File formats
  - Software packages
  - Software tools
- ❑ Test results will be shared with:
  - Plato
  - System users



# Plato

---

- ❑ Plato will retrieve information from the Planets Core Registry about:
  - File formats
  - Software tools
  - Risks
  - Pathways



# General User

---

- ❑ A general (human) user can search the Planets Core Registry through a web interface
- ❑ Can be used in the same manner as Pronom, to search for background information
- ❑ Added information about
  - Software packages
  - Pathways
  - Technical environments



# System user

---

- ❑ Extracting information from the Planets Core Registry is possible through web services
- ❑ Requested information is delivered in XML





Thank you for listening  
Questions?

[Sara.vanbussel@kb.nl](mailto:Sara.vanbussel@kb.nl)

