

CeBIT 2008

4-9 March 2008, Hannover, Germany



Planets: Integrated Digital Preservation Services

Planets overview

- ❑ 4-year project co-funded by the European Union
- ❑ Under the same programme as CASPAR and DPE
- ❑ Started in June 2006 with €15m budget
- ❑ Coordinated by the British Library
- ❑ 16 partners including national libraries and archives, technology companies and research universities
- ❑ Focuses on the needs of libraries and archives



Aims and objectives

- ❑ Help ensure long-term access to cultural and scientific heritage
 - Improve decision-making
 - Ensure long-term access
 - Control costs
 - Ensure wide adoption across user community
 - Establish market place for preservation services and tools
- ❑ Build practical solutions
 - Integrate existing expertise, designs and tools
 - Deliver tools and services that can be used in an operational environment

Planets partners



KB

Koninklijke Bibliotheek



STATSBIBLIOTEKET

Österreichische

Nationalbibliothek



DET KONGELIGE BIBLIOTEK

NATIONALBIBLIOTEK OG KØBENHAVNS UNIVERSITETSBIBLIOTEK



the national archives



Schweizerische Eidgenossenschaft

Confédération suisse

Confederazione Svizzera

Confederaziun svizra

Swiss Confederation

nationaal archief

- ❑ The British Library
- ❑ National Library, Netherlands
- ❑ Austrian National Library
- ❑ State and University Library, Denmark
- ❑ Royal Library, Denmark

- ❑ National Archives, UK
- ❑ Swiss Federal Archives
- ❑ National Archives, Netherlands



Planets partners



- ❑ Tessella Plc
- ❑ IBM Netherlands
- ❑ Microsoft Research
- ❑ Austrian Research Centers GmbH



- ❑ Hatii at University of Glasgow
- ❑ University of Freiburg
- ❑ Technical University of Vienna
- ❑ University of Cologne

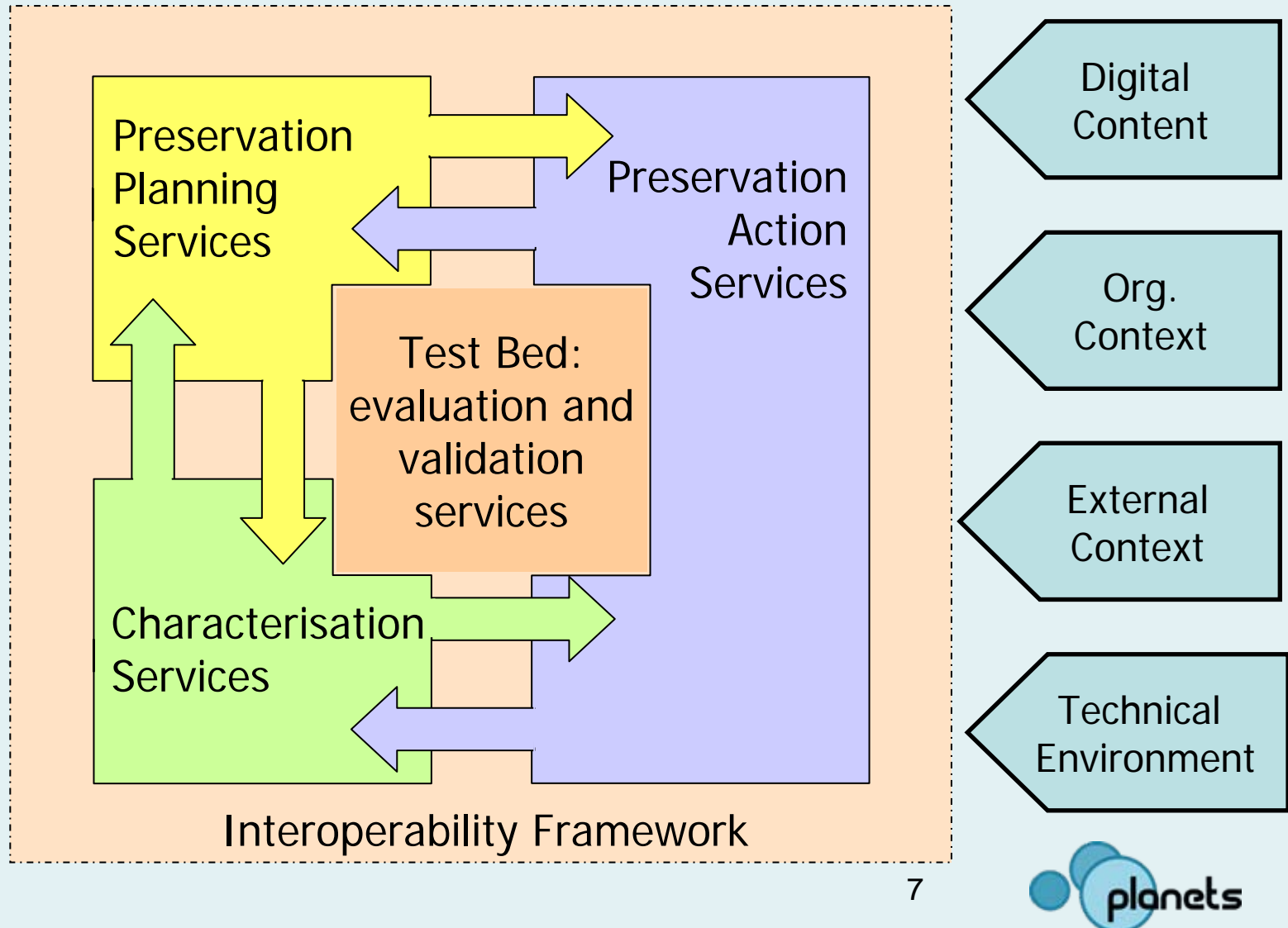


The Planets team



All Staff Meeting, Feb 2007

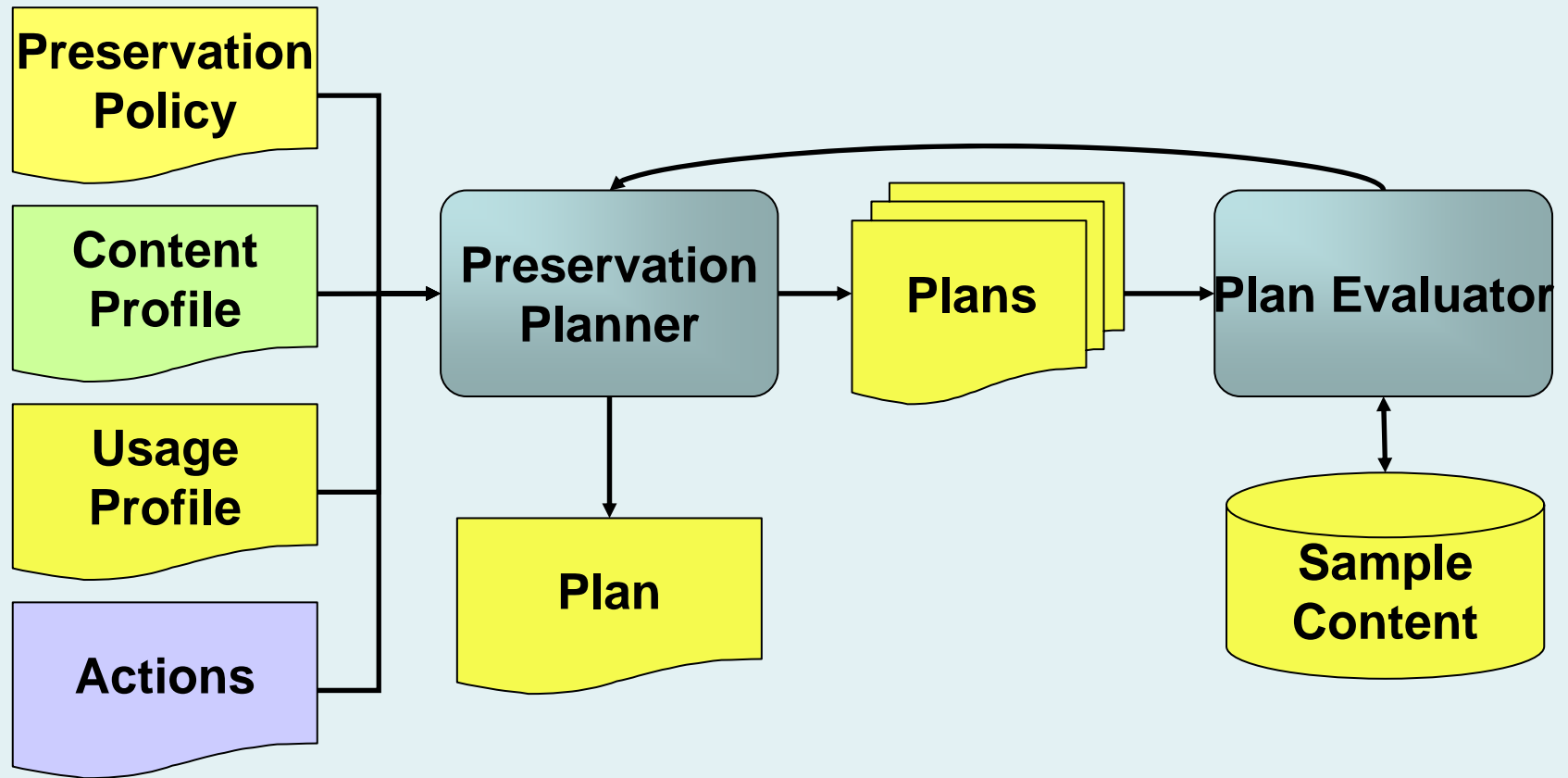
Planets architecture



Planets architecture: key components

- ❑ **Preservation planning**
 - tools and services for formulating and selecting preservation plans
- ❑ **Preservation characterisation**
 - tools and services for automatic analysis of digital objects' technical and intellectual characteristics
 - supporting registry of characterisation information
- ❑ **Preservation action**
 - methodology for describing preservation action tools
 - supporting registry
 - migration and emulation tools
- ❑ **Testbed**
 - hardware and software environment for comparing digital preservation tools and assessing their effectiveness
- ❑ **Interoperability framework**
 - service-oriented architecture
 - provides shared functions
 - integrates the Planets tools and services

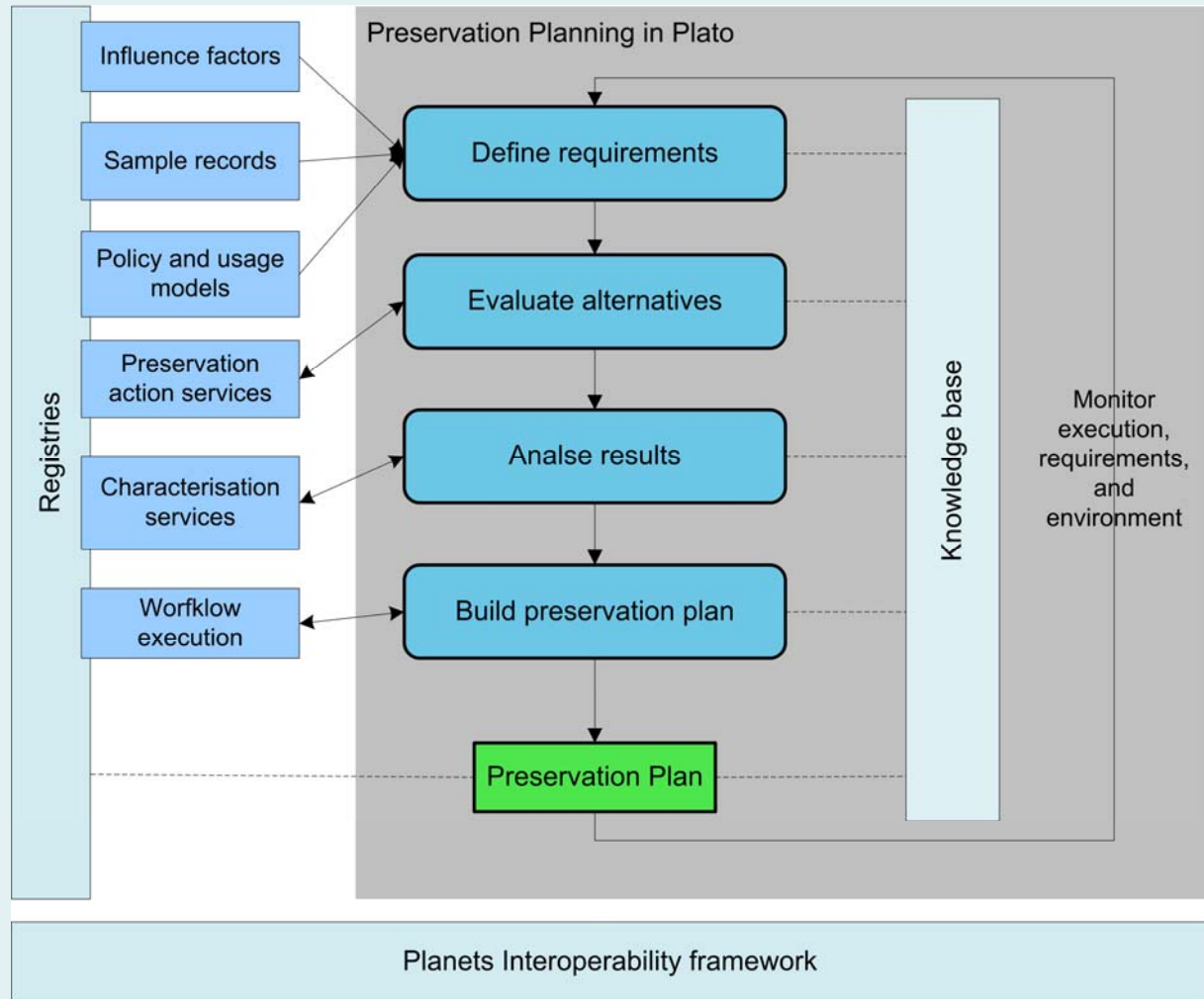
Preservation planning



Planning for a desired future

- ❑ Parameters for successful preservation strategy
 - Organisational policy
 - Usage pattern of digital collection
- ❑ Methodology and workflow for identifying preservation requirements, defining potential solutions and selecting the most appropriate
 - Plato, workflow-based decision support tool
- ❑ Validation framework for measuring and comparing characteristics (in a typical before and after scenario)
- ❑ Proactive preservation planning including collection profiling, technology watch and advice services
- ❑ Implementing the OAIS Preservation Planning function

Preservation Planning using Plato



Planets Preservation Planning Tool (Plato)



PLANETS Preservation Planning Tool (*Plato*)

Institute of Software Technology and Interactive Systems



Project | Define Requirements | Evaluate Requirements | Consider Results | Loaded project: PP4 workshop - The National Archive

Identify Requirements

[Expand All](#) | [Collapse All](#)

[Website](#)

Focus	Node	+	+	-	Single	Scale	Restriction	Unit
	Website	+	+					
X	Record characteristics	+	+					
X	Appearance	+	+					
X	Content	+	+					
X	Structure	+	+					
X	Behaviour	+	+					
X	deactivate	+	+					
X	mailto:		+		<input type="checkbox"/>	Ordinal	Yes/No	
X	preserve	+	+					
X	menus		+		<input type="checkbox"/>	Ordinal	complete/navigable/missing	
X	pop-ups		+		<input type="checkbox"/>	Ordinal	Yes/No	
X	freeze	+	+					
X	current date/time		+		<input type="checkbox"/>	Ordinal	frozen/missing/current	
X	visitor counter		+		<input type="checkbox"/>	Ordinal	frozen/missing/current	
X	Newsfeeds		+		<input type="checkbox"/>	Ordinal	frozen/missing/current	
X	Context	+	+					
X	Technical characteristics	+	+					
X	Ubiquity		+		<input type="checkbox"/>	Ordinal	Ubiquitous/Widespread/Specialised/Obs	
X	Tool Support		+		<input type="checkbox"/>	Positive Number		Number of tools
X	Documentation	+	+					
X	Quality		+		<input type="checkbox"/>	Ordinal	Primary/Secondary	
X	Disclosure		+		<input type="checkbox"/>	Ordinal	Full/Partial/None	
X	Openness		+		<input type="checkbox"/>	Ordinal	Standard/Open/Proprietary	
						Ordinal	Public/Limited/None	

© 2007 Institute of Software Technology and Interactive Systems: «office bears»

Quick Access:














PLANETS Preservation Planning Tool (*Plato*)[\[logout\]](#) [\[Export to XML\]](#) [\[help\]](#)

Project | Define Requirements | Evaluate Requirements | Consider Results | Project 'Minimalist test project covering all features' is in state WEIGHTS_SET

Analyse Results

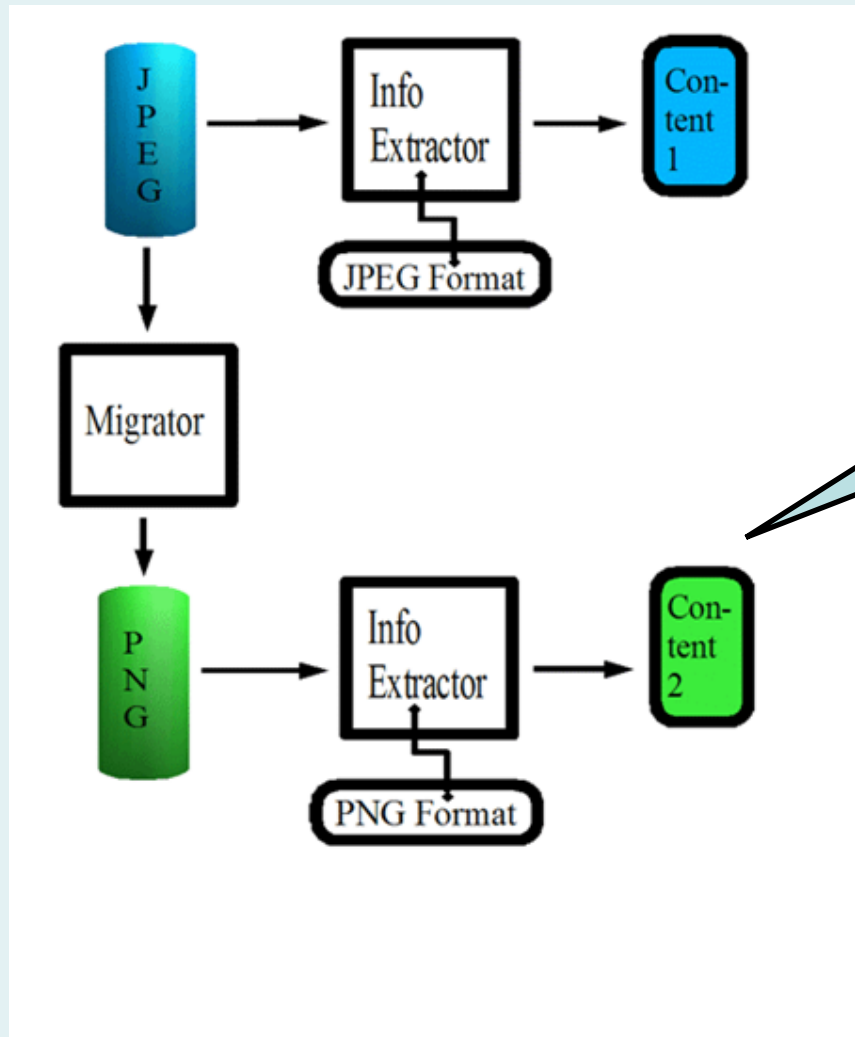
Sum ☒ PDF/A (Tool A)☒ PDF/A (Tool B)[Show](#)[Expand All](#) | [Collapse All](#)**Minimalist root node**

Focus	Name	Result
	▼ Minimalist root node	PDF/A (Tool A): 2,98  PDF/A (Tool B): 3,19 
X	► Image properties	PDF/A (Tool A): 0,70  PDF/A (Tool B): 0,80 
X	▼ Karma	PDF/A (Tool A): 0,40  PDF/A (Tool B): 0,00
X	▼ Filesize (in Relation to Original)	PDF/A (Tool A): 0,78  PDF/A (Tool B): 0,99 
X	▼ A Single-Leaf	PDF/A (Tool A): 0,40  PDF/A (Tool B): 0,80 
X	▼ IntRange 0-10	PDF/A (Tool A): 0,70  PDF/A (Tool B): 0,60 

Content characterisation

- ❑ Automating the process of identifying, describing and extracting the essential characteristics of digital objects
- ❑ Supporting registry with technical information about significant properties of digital object types
- ❑ Generic XML language supporting automatic validation of document conversions and evaluation of migration quality
 - *eXtensible characterisation definition language (XCDL)* and *eXtensible characterisation extraction language (XCEL)*
 - Hierarchically decompose and represent documents in an abstract XML language
- ❑ Extractor and Comparator software

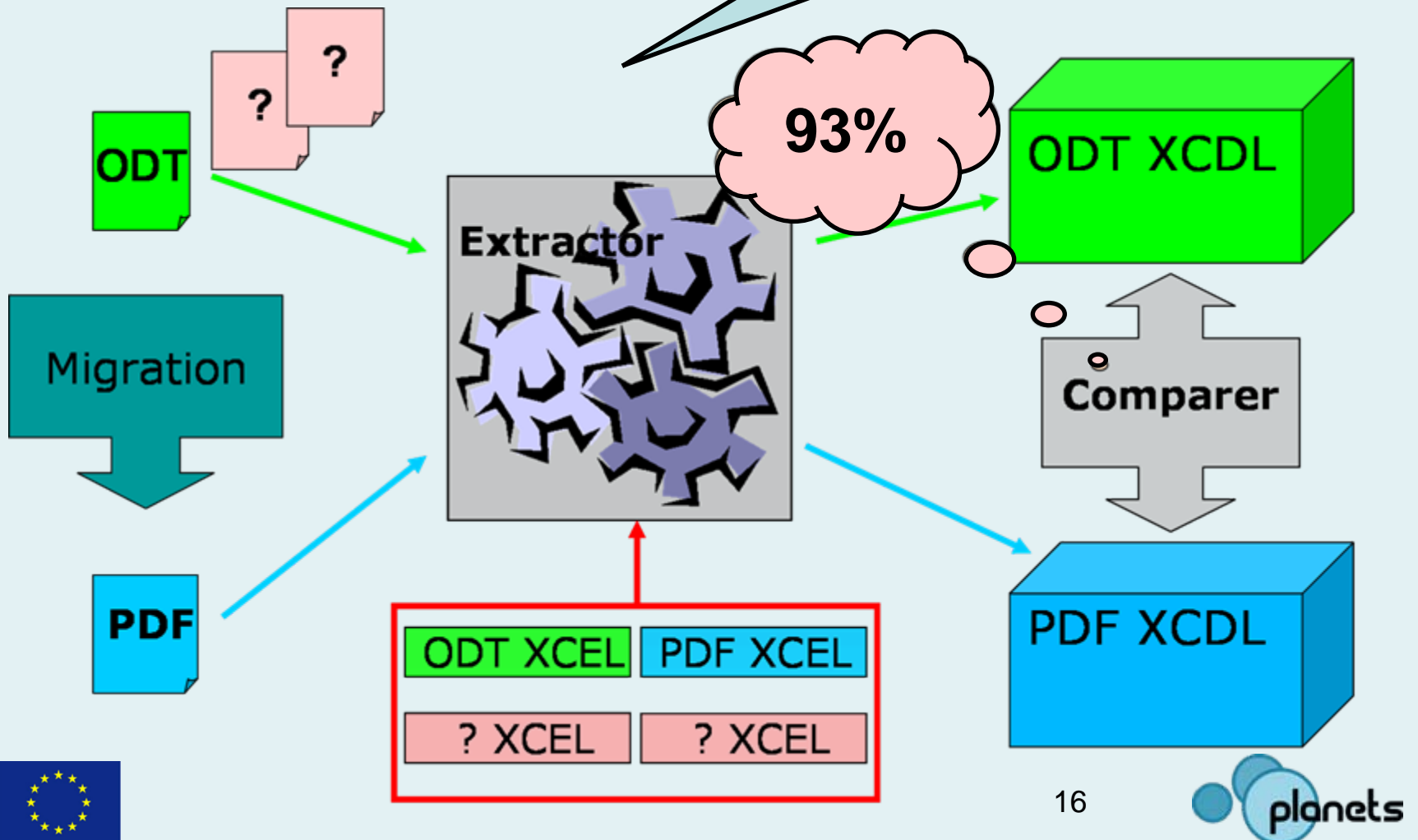
XCEL extractor



XCEL processing software which extracts content of a file and express it in XCDL

XCDL comparator

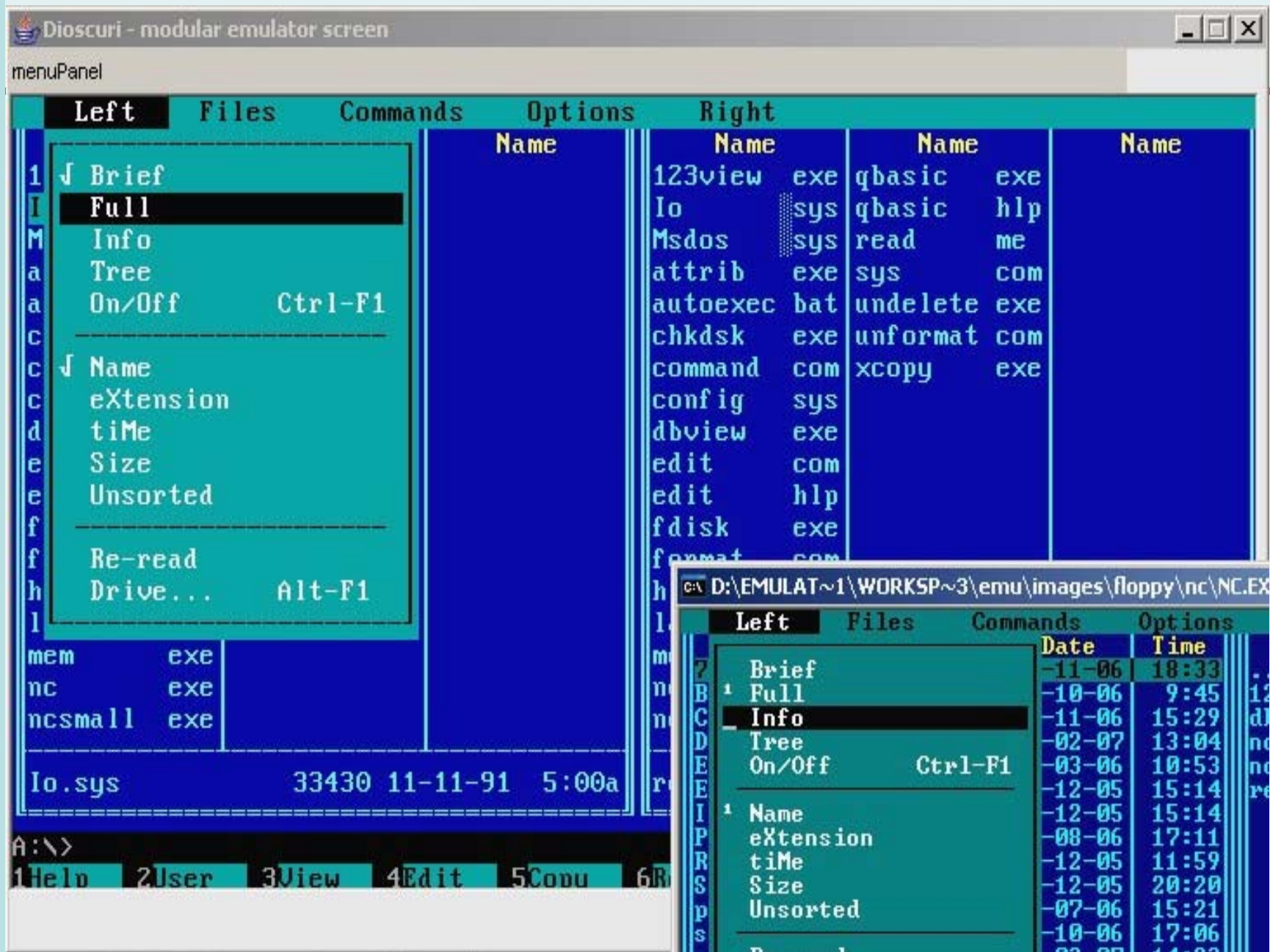
Identify degrees of equality between XCDL documents



Preservation action

- ❑ Inventory of commonly used file formats
- ❑ Gap analysis of current tools
- ❑ Formal language for describing preservation action tools
- ❑ Developing the most wanted tools for digital objects
- ❑ Supporting registry with technical descriptions of tools
- ❑ Migration tools to convert digital objects into newer formats
- ❑ Emulation tools enabling interaction with digital objects using original software applications
 - Dioscuri – the modular emulator
 - Universal Virtual Computer (UVC)

Dioscuri – the modular emulator



The Universal Virtual Machine (UVC)

The screenshot displays the UVC (Universal Virtual Machine) interface, which emulates a virtual machine running Ubuntu. The main window shows the Ubuntu desktop environment with a green landscape background. The top panel includes the Ubuntu logo, the text "Ubuntu VMware Player", and various virtual hardware icons: CD-ROM (IDE 1:0), Floppy, Ethernet, and Sound Adapter. The system clock indicates "Wed Nov 21, 9:01 PM".

In the foreground, there are three windows:

- demo_spreadsheet - Sharp Tools Spreadsheet**: A spreadsheet application window showing a table with renovation expenses.
- uvc@uvc-vmware: ~/uvc_package/spreadsheet**: A terminal window showing the command `cat demo_spreadsheet.csv` and its output, which matches the data in the spreadsheet.
- UVC Spreadsheet**: A window displaying a simplified version of the spreadsheet data.

demo_spreadsheet - Sharp Tools Spreadsheet Data:

	A	B	C	D
1	Expenses renovation	Roof	12000.0	
2		Kitchen	7800.0	
3		Bathroom	3200.0	
4		Garden	2100.0	
5				
6		Total:	25100.0	
7				

Terminal Output:

```
uvc@uvc-vmware: ~/uvc_package/spreadsheet
uvc@uvc-vmware: ~/uvc_package/spreadsheet$ cat demo_spreadsheet.csv
Expenses renovation      Roof      12000.0
      Kitchen 7800.0
      Bathroom      3200.0
      Garden  2100.0

Total:  =C1+C2+C3+C4
```

UVC Spreadsheet Data:

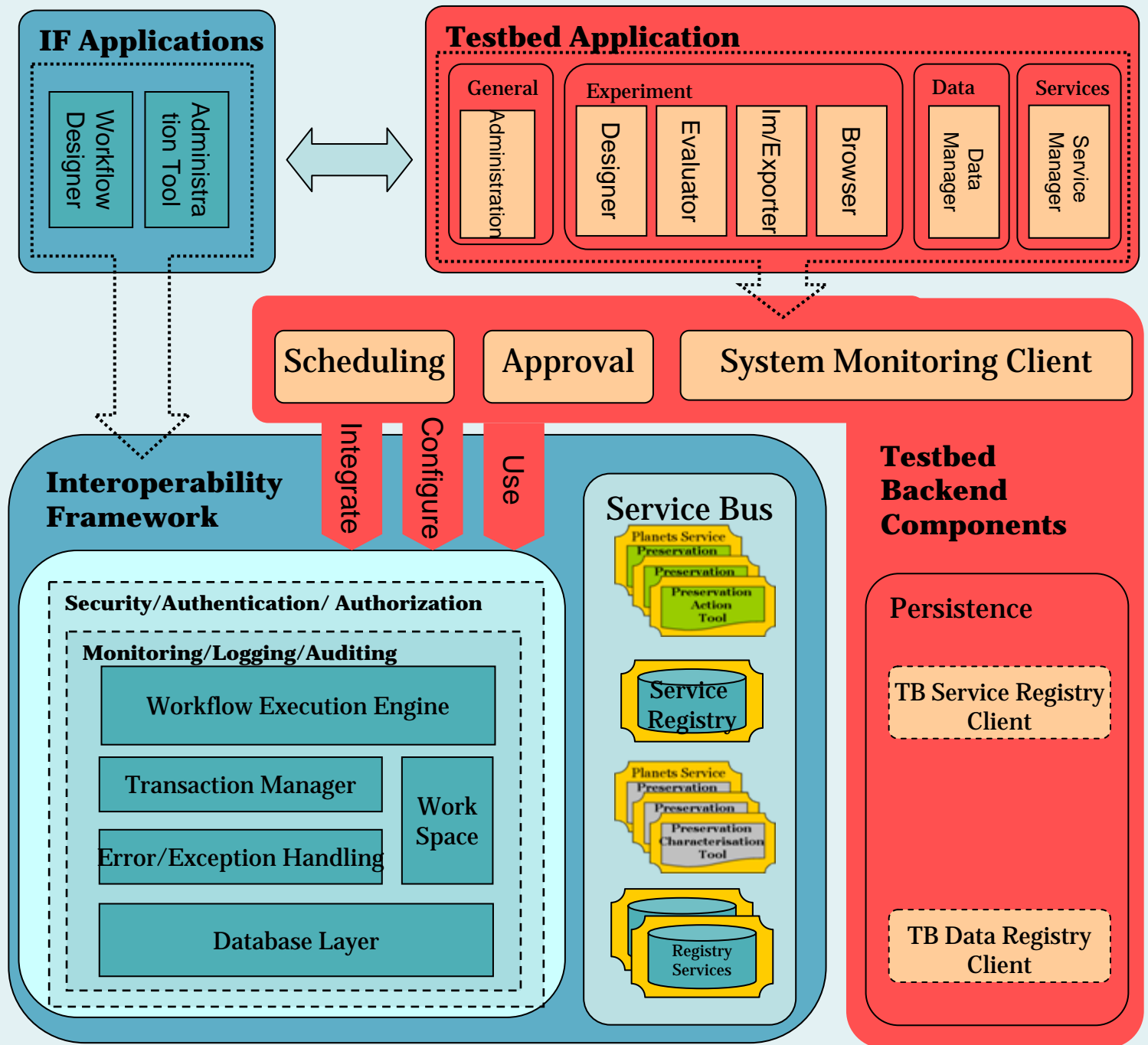
	A	B	C
1	Expenses renovation	Roof	12000.0
2		Kitchen	7800.0
3		Bathroom	3200.0
4		Garden	2100.0
5			
6		Total:	25100.0
7			

The bottom status bar shows the following open windows: [System Monitor], uvc@uvc-vmware: ~/uvc_..., Untitled - Sharp Tools Spre..., and UVC Spreadsheet.

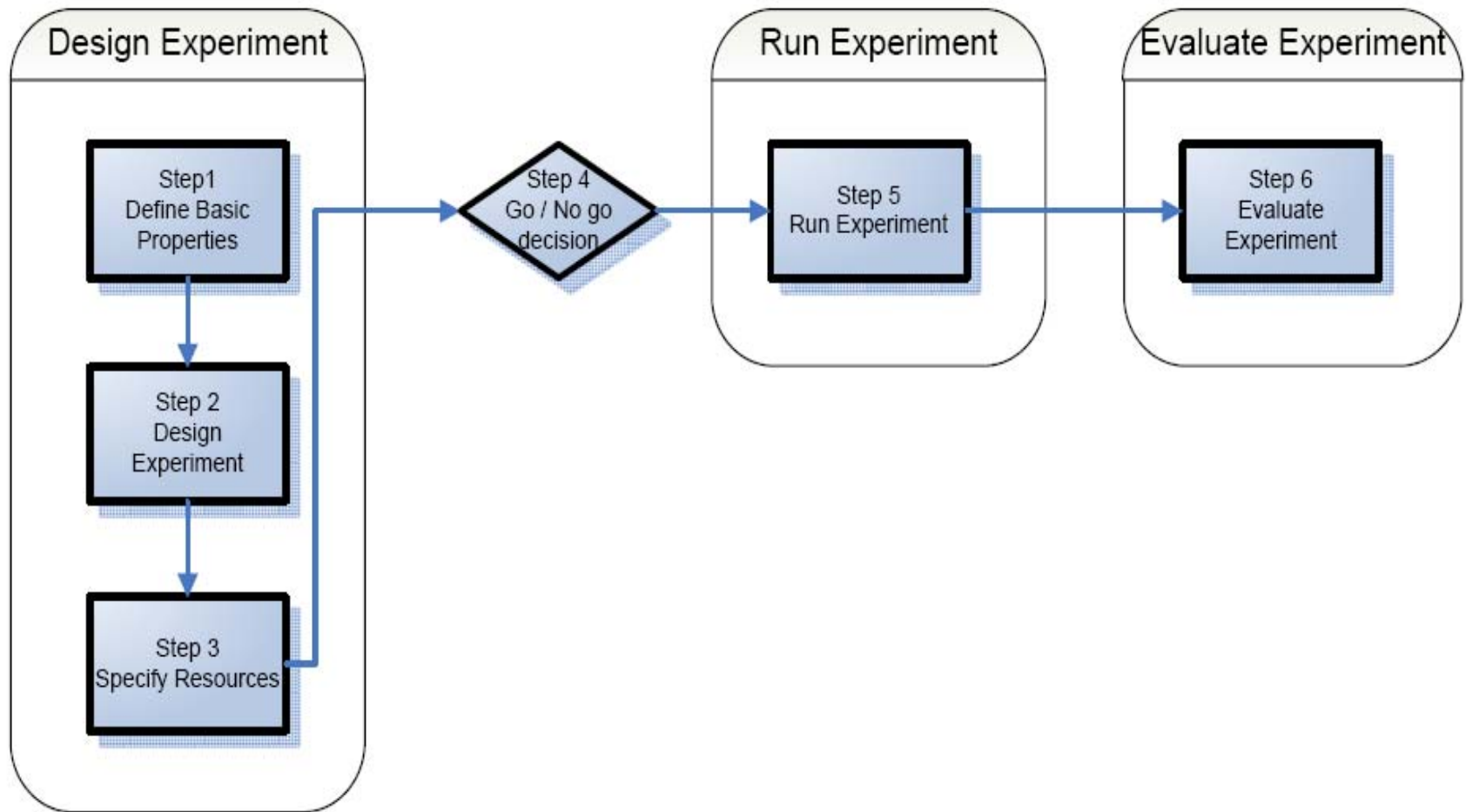
Planets Testbed: learn what works

- ❑ “A controlled environment with metrics and benchmark content that allows the objective comparison of preservation tools and strategies through experimentation”
- ❑ A dedicated scientific research environment
- ❑ Consists of
 - Data storage, hardware, Planets software, Testbed application
 - Benchmark and other content
- ❑ Use well known and widely spread technologies
- ❑ Available for external organisations in early 2009

Planets Testbed



Testbed experiment process



Prototype Testbed

Welcome, Hans [Contact](#) | [Log out](#)

[Home](#) [My Experiments](#) [New Experiment](#) [Administrate Experiments](#) [User Administration](#)

Steps:

1. Define Basic Properties
2. Design Experiment
3. Specify Resources
4. Go / No-go decision
5. Run Experiment
6. Evaluate Experiment

Define Basic Properties * Indicates required field

* **Name:**

Code:

Status:

Summary:

Type:

Objects:

☐ Text ☐ Image ☐ Sound

☐ Video ☐ Mixed

Description:

[Purpose](#) [Focus](#) [References](#) [Approach](#) [Scope](#) [Considerations](#)

Scope of experiment:

Planets Software: Vision

- ❑ A single downloadable software package
- ❑ Simple to install, configure and administer
- ❑ When deployed, a **Planets instance** is created, in which
 - an administrator can
 - create user accounts
 - deploy and browse services
 - browse registries
 - a preservation expert can
 - define service workflows (Workflow Design Tool)
 - define and evaluate preservation plans (Preservation Planning Application)
 - define and run experiments (Testbed)
 - a librarian or archivist can
 - define and test preservation plans
 - execute preservation processes on a repository (Online Design Tool)



Planets Software: motivation

- ❑ Provide commonly required functions
 - Data persistence
 - User management
 - Authentication and Authorisation
 - Monitoring, Logging, and Messaging
- ❑ Meet non-functional requirements on the infrastructure
 - Robust
 - Scalable
 - Distributed

Planets software: benefits

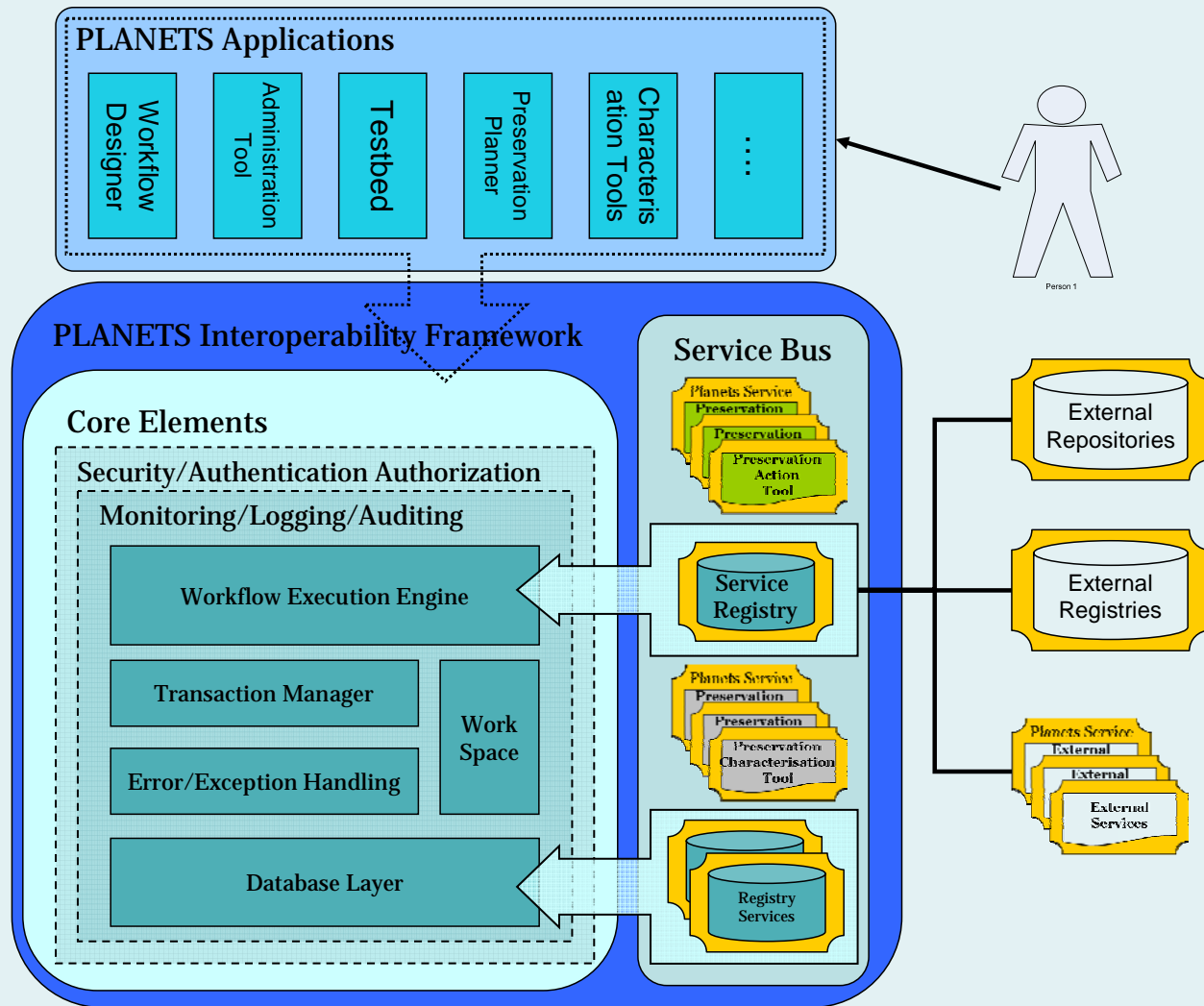
❑ Efficiency

- Components can be shared and only need to be developed once
- Optimise the number of components: IF provides a single database for all components, so only one database need be installed

❑ Interoperability

- Common components help ensure interoperability between various applications
- Support access to remote and distributed 3rd party characterisation and migration services by enforcing Web Service standards

Planets software: architecture



Interoperability Framework components

- ❑ Application server (JBoss)
- ❑ Relational Database (Apache Derby)
- ❑ Service Registry (jUDDI)
- ❑ Data Registry (JCR implementation on Jackrabbit)
- ❑ Workflow Design Tool
 - Workflow expressed in BPEL
 - Expert design tool based on the Eclipse BPEL Plugin

Single sign-on service

https://localhost:8443/security/login?service=https%3A%2F%2Flocalhost%3A8443%2Fadmin%2F%2F_scegl_cas_security_check

PLANETS SINGLE SIGN ON SERVICE

Planets Central Authentication Service

You have requested access to a site that requires authentication.

Enter your ID and password below; then click on the Login button to continue.

Requested Service : Administration User Interface

Be wary of any program or web page that asks you for your ID and password. Planets Distributed Digital Archive Services web pages that ask you for your ID and password will generally have URLs that begin with "https://www.planets-project.eu" or "https://localhost". In addition, your browser should visually indicate that you are accessing a secure page.

For the time being, cookies must be enabled to access PLANETS services

Username:

Password:

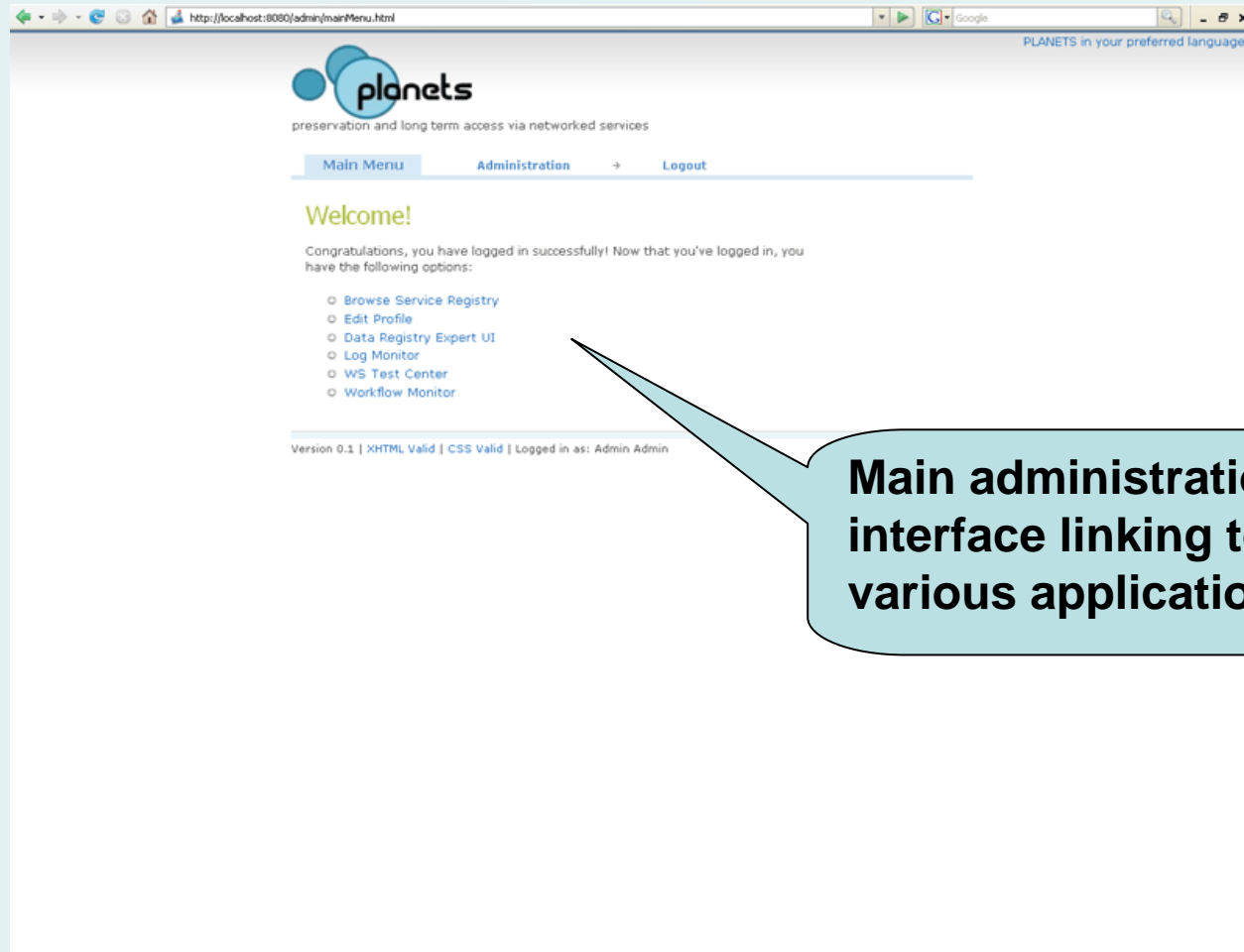
Login

For security reasons, quit your web browser when you are done accessing services that require authentication!

planets

Initial login directed to the Single Sign On Service

Administration interface



Main administration interface linking to various applications...

Planets Web Service registry



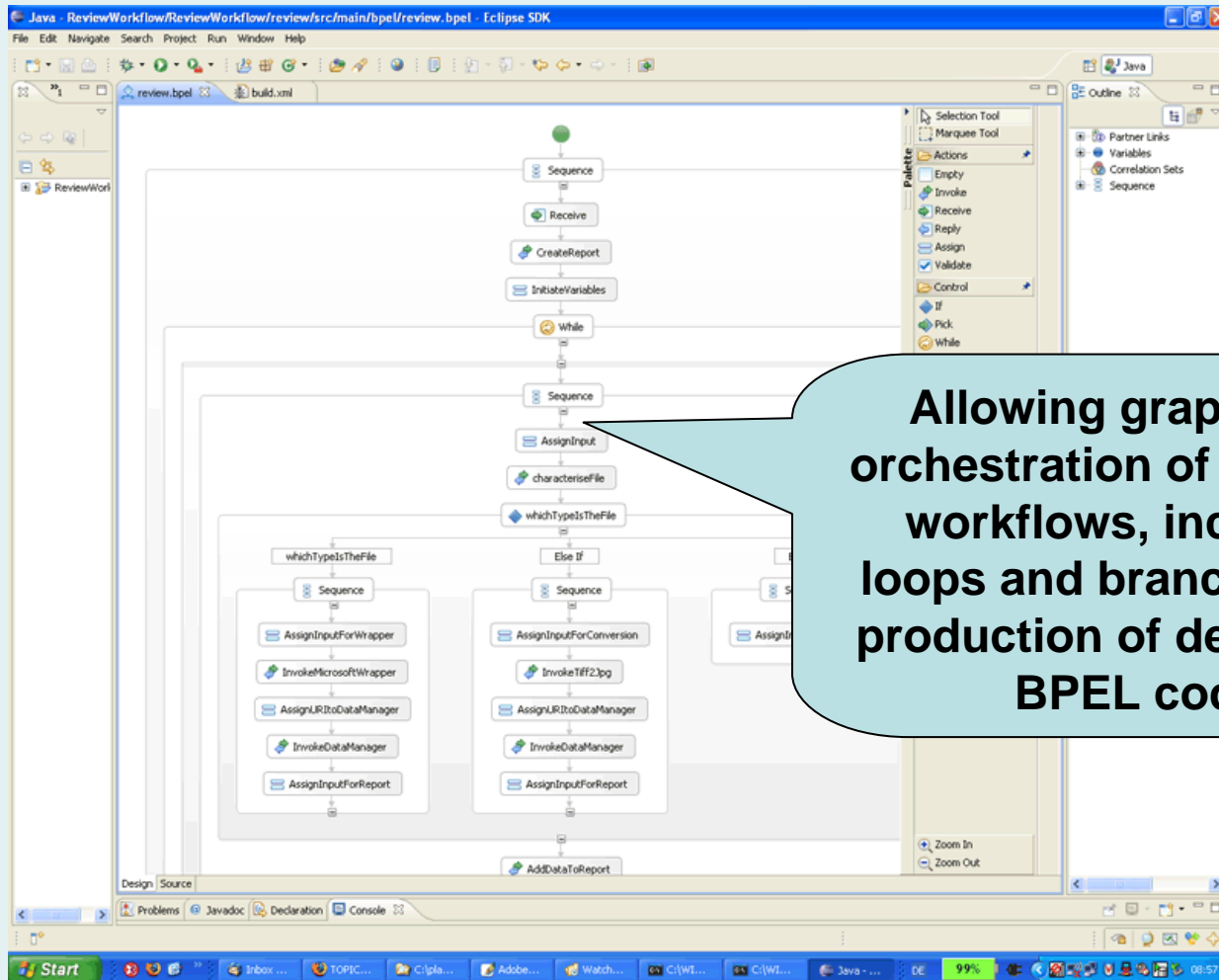
The screenshot shows a web browser window with the URL `http://localhost:8080/admin/FindOrg.html`. The page title is "PLANETS Service Registry - Brow...". The Planets logo is at the top left, with the tagline "preservation and long term access via networked services". A navigation bar includes "Main Menu", "Administration", and "Logout". The main heading is "Browse Service Registry". Below it, a text box explains: "Enter the partial or complete name of an organization to search for. The '%' symbol can be used as a wildcard that matches any character string. Press the 'Search' button to start the search." There is a dropdown menu set to "Businesses". A "Search criteria" section contains a text input field with "Ple%" and three buttons: "Search", "Reset", and "Cancel". The footer shows "Version 0.1 | XHTML Valid | CSS Valid | Logged in as: Admin Admin" and a copyright notice "© 2004".

Admin interface linking to Planets Web Service registry, where services can be manually registered and searched

Workflow design tool



Workflow design tool



**Allowing graphically
orchestration of complex
workflows, including
loops and branches, and
production of deployable
BPEL code**

Workflow execution

PLANETS - PC/4 CHARACTERISATION TOOL

Home Test WS Check WS Deploy Undeploy

Test a WebService - Create and Send Request

WSDL URL:

Selected Service Name:

Selected Operation Name:

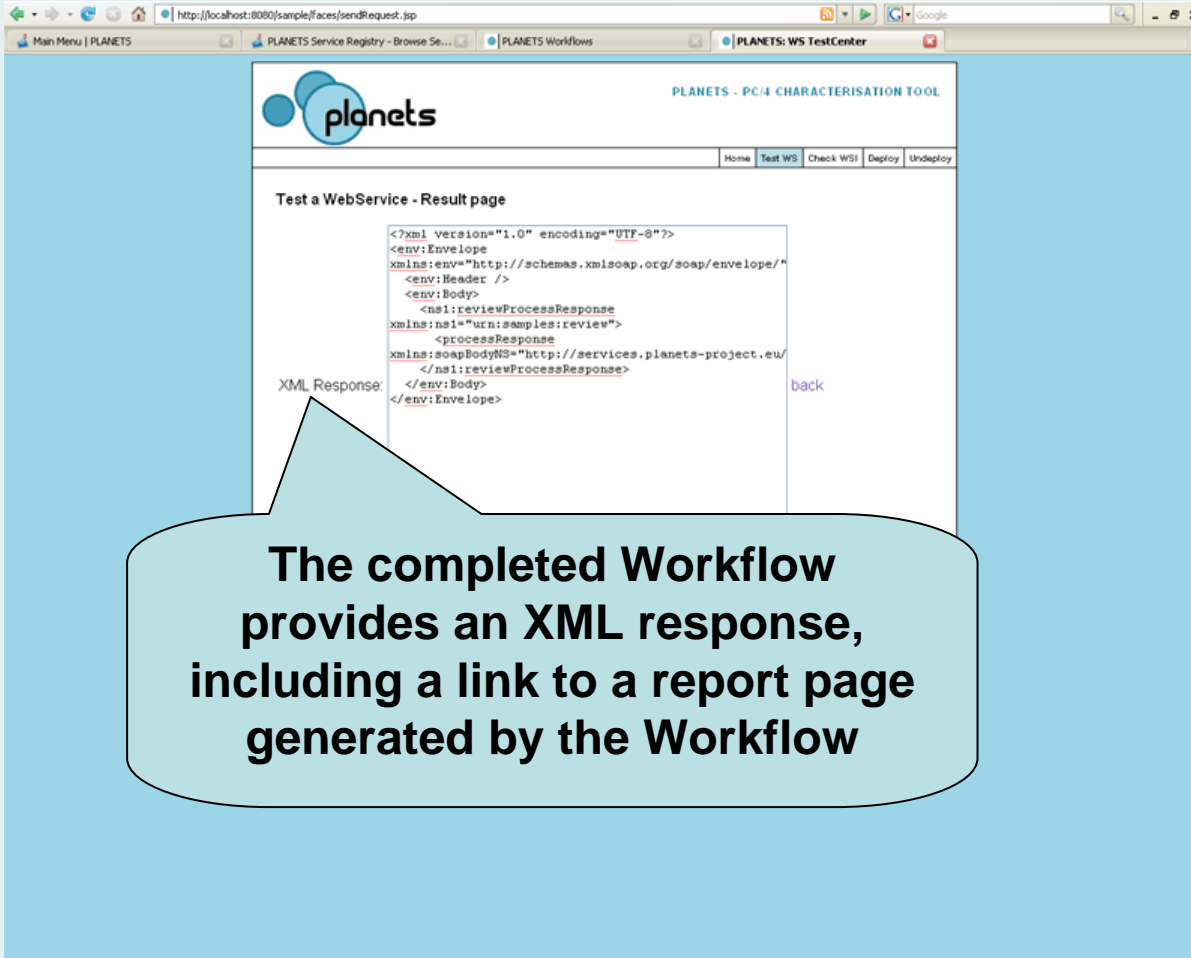
XML Request:

```
<ns1:reviewProcess xmlns:ns1="urn:samples:review">
  <processRequest>
    <ns1:item>C:/planets-dev/docs/adaptive_query_build
    <ns1:item>C:/planets-dev/docs/dme_flyer_co_2006012
    <ns1:item>C:/planets-dev/ConvertPics/test1.tif</ns1
    <ns1:item>C:/planets-dev/ConvertPics/test2.tif</ns1
    <ns1:item>C:/planets-dev/ConvertPics/test3.tif</ns1
    <ns1:item>C:/planets-dev/ConvertPics/test4.tif</ns1
    <ns1:item>C:/planets-dev/ConvertPics/test5.tif</ns1
    <ns1:item>C:/planets-dev/ConvertPics/test6.tif</ns1
    <ns1:item>C:/planets-dev/ConvertPics/test7.tif</ns1
  </processRequest>
</ns1:reviewProcess>
```

[back to top](#) | Date Published: 6 June 2006 | Last Modified: 6 June 2006 | © Planets 2006 unless otherwise stated.

Executing service request...

Workflow execution



The screenshot shows a web browser window with the URL `http://localhost:8080/sample/faces/sendRequest.jsp`. The browser tabs include "Main Menu | PLANETS", "PLANETS Service Registry - Browse Se...", "PLANETS Workflows", and "PLANETS: WS TestCenter". The page title is "PLANETS - PC/4 CHARACTERISATION TOOL". The navigation bar has links: Home, Test WS, Check WS, Deploy, and Undeploy. The main content area is titled "Test a Webservice - Result page" and displays an XML response. A callout bubble points to the XML response with the text: "The completed Workflow provides an XML response, including a link to a report page generated by the Workflow".

XML Response:

```
<?xml version="1.0" encoding="UTF-8"?>
<env:Envelope
  xmlns:env="http://schemas.xmlsoap.org/soap/envelope/"
  <env:Header />
  <env:Body>
    <ns1:reviewProcessResponse
      xmlns:ns1="urn:samples:review">
      <processResponse
        xmlns:soapBodyNS="http://services.planets-project.eu/
      </ns1:reviewProcessResponse>
    </env:Body>
  </env:Envelope>
```

[back](#)

Workflow execution

The screenshot displays the PLANETS - REPORT GENERATION SERVICE interface. The main content area is titled "Workflow Report - 0" and lists seven file conversion tasks. Each task entry includes the file path, format, status, and converted URI. The status for all files is "File successfully converted".

File Path	File Format Information	Conversion Status	Converted File URI
C:/planets-dev/docs/adaptive_query_building.doc	application/msword	File successfully converted	/reviewWorkflow/migrateTaskfile1
C:/planets-dev/docs/dme_flyer_co_20060126_v5.doc	application/msword	File successfully converted	/reviewWorkflow/migrateTaskfile2
C:/planets-dev/ConvertPics/test1.tif	image/tiff	File successfully converted	/reviewWorkflow/migrateTaskfile3
C:/planets-dev/ConvertPics/test2.tif	image/tiff	File successfully converted	/reviewWorkflow/migrateTaskfile4
C:/planets-dev/ConvertPics/test3.tif	image/tiff	File successfully converted	/reviewWorkflow/migrateTaskfile5
C:/planets-dev/ConvertPics/test4.tif	image/tiff	File successfully converted	/reviewWorkflow/migrateTaskfile6
C:/planets-dev/ConvertPics/test5.tif	image/tiff	File successfully converted	/reviewWorkflow/migrateTaskfile7
C:/planets-dev/ConvertPics/test6.tif	image/tiff	File successfully converted	/reviewWorkflow/migrateTaskfile8

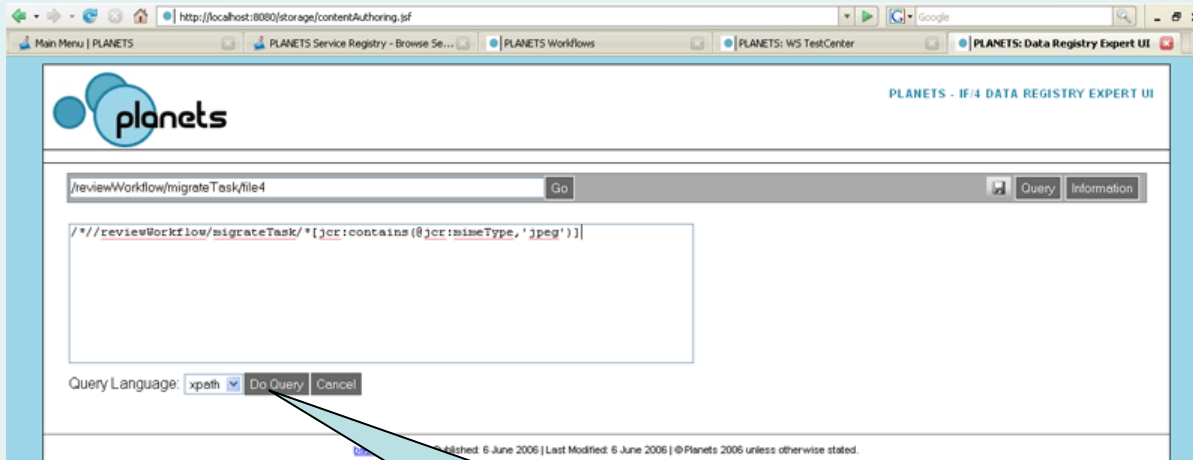
**Results of the workflow
for each submitted file.
Migrated files are written
to the data registry**

Data registry

The screenshot displays the PLANETS Data Registry Expert UI. The browser address bar shows `http://localhost:8080/storage/contentAuthoring.jsf`. The page title is "PLANETS - IF/4 DATA REGISTRY EXPERT UI". The main content area is titled "reviewWorkflow/migrateTask". On the left, a file tree lists "file1" through "file22". The main panel shows details for the "migrateTask" node: "Current node: migrateTask", "Primary NodeType: nt:unstructured", "Number of versions: This node is not versionable", and "Index: 1". Below this, it states "This node has no properties." and provides a list of actions: "Add mixin" (set to "mix:versionable"), "Remove mixin", "Add named property", "Add wildcard property", "Add Node" (set to "nt:base"), "Move node to:", "Copy node to:", "Export this node" (with "Recursive" checked and "Export System View" selected), "Locking" (with "Lock is deep" checked and "Lock is session-scoped" unchecked), "Import data", and "Delete Node". A callout bubble points to the "migrateTask" node in the file tree and contains the text: "It is possible to directly browse the node structure of the JCR-based Data Repository".

It is possible to directly browse the node structure of the JCR-based Data Repository

Data registry



The screenshot shows a web browser window with the URL `http://localhost:8080/storage/contentAuthoring.jsf`. The browser's tab bar includes links for 'Main Menu | PLANETS', 'PLANETS Service Registry - Browse Se...', 'PLANETS Workflows', 'PLANETS: WS TestCenter', and 'PLANETS: Data Registry Expert UI'. The page header features the 'planets' logo and the text 'PLANETS - IF/4 DATA REGISTRY EXPERT UI'. The main content area has a search bar with the text `/reviewWorkflow/migrateTask/file4` and a 'Go' button. To the right of the search bar are 'Query' and 'Information' buttons. Below the search bar is a large text area containing the Xpath query `//*[reviewWorkflow/migrateTask/*[jcr:mimeType,'jpeg']]`. At the bottom, there is a 'Query Language:' dropdown menu set to 'xpath', followed by 'Do Query' and 'Cancel' buttons. A small footer note at the bottom right of the interface reads: 'Published: 6 June 2006 | Last Modified: 6 June 2006 | © Planets 2006 unless otherwise stated.'

**One can also execute Xpath
or standard SQL queries on
the repository**

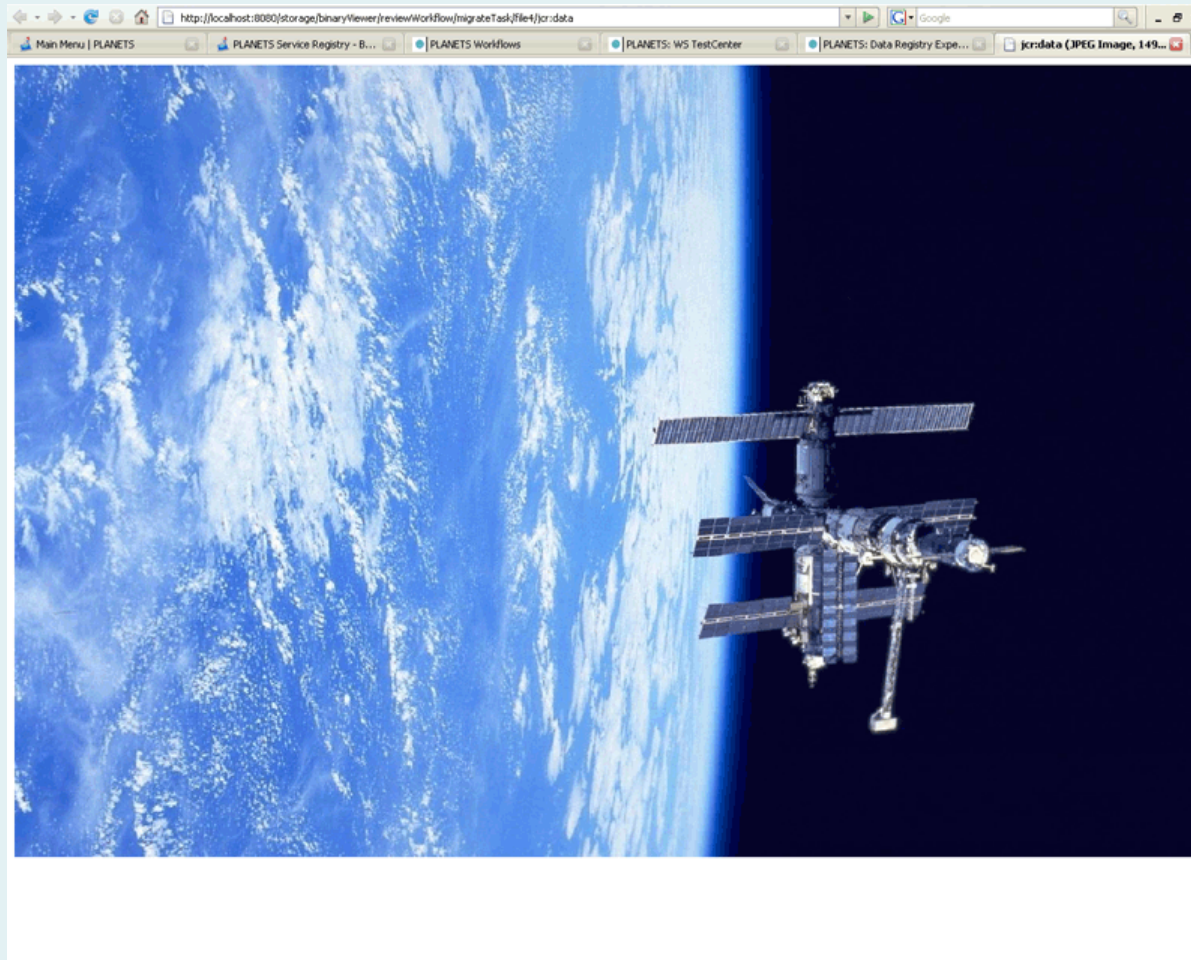
Data registry

The screenshot displays the PLANETS Data Registry Expert UI. The browser address bar shows `http://localhost:8080/storage/contentAuthoring.jsf`. The page title is "PLANETS - IF/4 DATA REGISTRY EXPERT UI". The main content area shows a workflow migration task for a node named "file4". The node is of type "nt:resource" and is not versionable. It has one version with index 1. The node's properties are listed in a table:

Name	Type	MA	MU	PR	AU	Decl. NT	Value
jcr:uuid	String	X		X	X	mix:referenceable	6a7247d3-7aa8-4961-a025-ec781708b493
jcr:data	Binary	X				nt:resource	Open binary value... Browse...
jcr:lastModified	Date	X				nt:resource	27/5/2008
jcr:mimeType	String	X				nt:resource	image

Below the table, there are several actions available: "Apply changes", "Add mixin", "Remove mixin", "Add named property", "Add wildcard property", "Add Node", "Move node to:", "Copy node to:", "Export this node", "Locking", and "Import data". A callout box points to the "jcr:data" property, stating: "... and observe the content (in this case a TIFF image migrated to the JPEG format)".

Data registry



Conclusion

- ❑ Planets will help organisations diagnose and treat problems with their digital collections
- ❑ High levels of automation and scalable components will reduce costs and improve quality
- ❑ Empirical data will enable improved decision making

Find out more about Planets

- ❑ **Contact us:** info@planets-project.eu
- ❑ **Website:** www.planets-project.eu
- ❑ **Planets news:** <http://www.planets-project.eu/news/>
- ❑ **Workshops, conferences and training events:**
<http://www.planets-project.eu/events/>
- ❑ **Quarterly newsletters:**
[http://www.planets-project.eu/publications/?search\[0\]=10](http://www.planets-project.eu/publications/?search[0]=10)
- ❑ **Scientific publications:**
[http://www.planets-project.eu/publications/?search\[0\]=12](http://www.planets-project.eu/publications/?search[0]=12)

