

# A Framework for Distributed Preservation Workflows

Rainer Schmidt  
AIT Austrian Institute of Technology  
rainer.schmidt@ait.ac.at



## Outline

- Overview of the Integrated Environment
  - Main Objectives and Architecture
- Planets Preservation Services
- Digital Objects and Metadata
- Integrating Repositories
- The Workflow Execution Engine (WEE)
  - Workflow Templates
  - WEE Services and the GUI

## Planets Project

- “Permanent Long-term Access through NETworked Services”
- Addresses the problem of digital preservation
  - driven by National Libraries and Archives
- Project instrument: FP6 Integrated Project
- 5. IST Call
- Consortium: 16 organisations from 7 countries
- Duration: 48 months, June 2006 – May 2010
- Budget: 14 Million Euro
- <http://www.planets-project.eu/>

## The Planets Environment

- An integrated System for the development and evaluation of preservation strategies.
- Uniform access mechanisms to a broad range of “commodity” tools, e.g. for *characterization*, *migration*, *emulation*.
- Integration of existing repositories, data/metadata formats.
- Specification, execution, recording of preservation workflows.
- Integration with end-user applications for preservation planning and the evaluation of tools/strategies.
  - PLANETS *Preservation Planning Tool* and *Testbed*
















## Service-Orientated Architecture

- XML Web Services (SOAP, WSDL, WS-\*)
  - Platform, Language, and Location Independence
- Homogeneous interfaces for preservation activities, data management, workflow execution.
  - Remotely access repositories and data.
  - Discover and dynamically utilize tools in a workflow.
- Supports distributed and cross-organizational deployments
  - Shared hardware, software, maintenance
  - Browser-based access to large number of resources

## Preservation Interfaces (the Verbs)

- Define atomic preservation activities (level-one)
  - Concentrates on low-level concepts and actions
    - Bit-stream operations, no data management
  - Designed to be light-weight and easy to implement
- Independent from a specific tool, language, or content type
  - *E.g. Characterize, Migrate, Compare, CreateView*
  - >50 Tools wrapped/provided as Planets Services
- Provides the basic abstractions for assembling workflows.

## Planets IF Service Registry

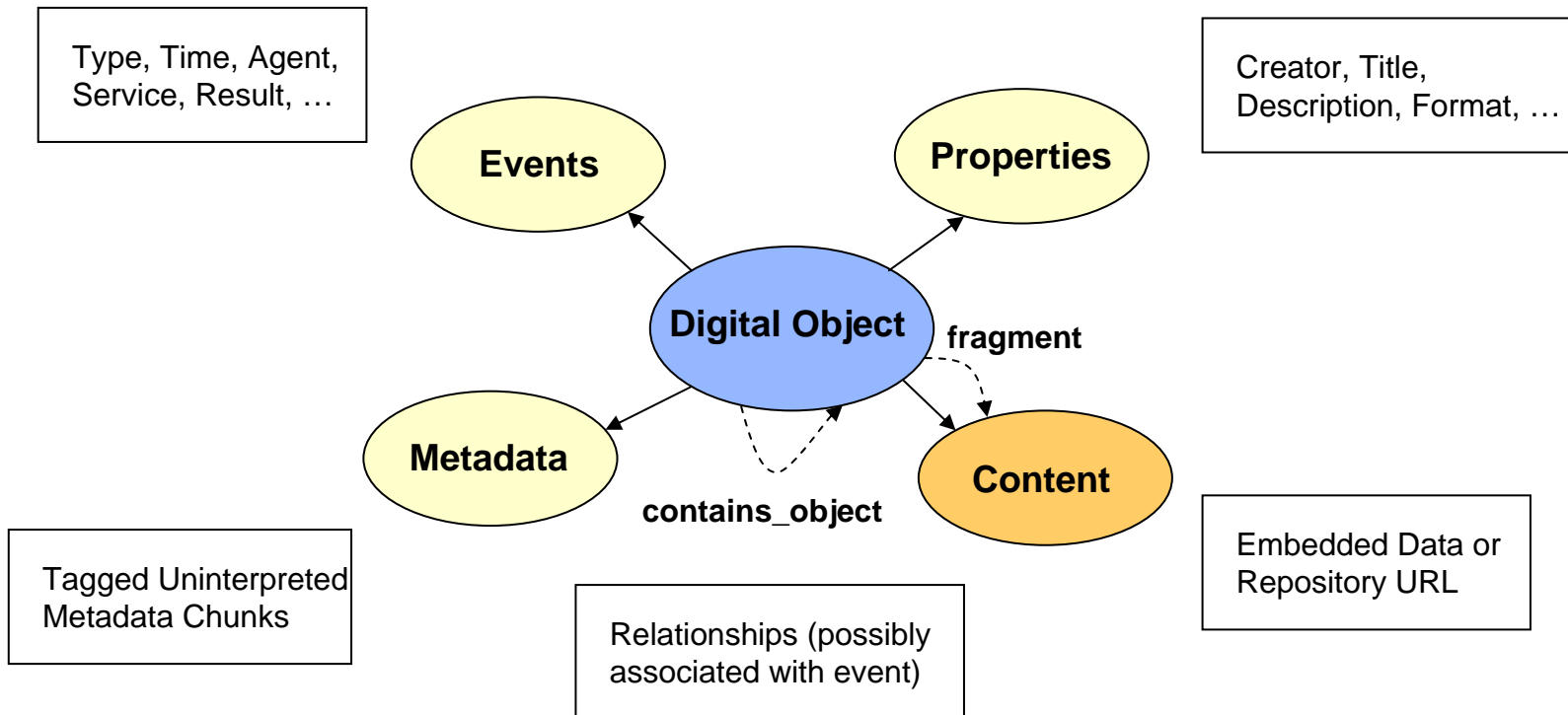
Home		Registered Services	Browse Local Endpoints	Add External Endpoints
All	 Refresh List			Search
Name	Registered	Service Details		
FloppyImageHelperUnix Migrate		<h3>Droid Service Endpoint Information</h3> <p>WSDL: <a href="http://ubuntu.planets-project.arcs.ac.at:80/pserv-pc-droid/Droid?wsdl">http://ubuntu.planets-project.arcs.ac.at:80/pserv-pc-droid/Droid?wsdl</a> Status: OK</p> <h3>DROID Identification Service</h3> <p><b>Type</b> eu.planets_project.services.identify.identify</p> <p><b>Author</b> Carl Wilson, Fabian Steeg</p> <p><b>Provider</b> The Planets Consortium.</p> <p><b>Description</b> Identification service based on Droid (DROID 3.0, Signature File 16).</p> <h3>You are not Authorised to Register Services</h3> <p>Your user account lacks the privileges required to register services. You should contact the IF Administrator and request that they register this service for you.</p> <p>If you believe that this is an oversight, or that you have previously been able to register services then contact the IF Administrator.</p>		
FloppyImageHelperWin Migrate				
FloppyImageModifyWin Modify				
AvidemuxMigration Migrate				
JhoveValidation Validate				
JavalmageIOMigrate Migrate				
DioscuriArjMigration Migrate				
SanselanMigrate Migrate				
PdfBoxMigration Migrate				
JhoveIdentification Identify				
InkscapeMigration Migrate				
Droid Identify				
OpenJpegMigration Migrate				
DioscuriPnmToPngMigration Migrate				

## Digital Objects (the Nouns)

- Generic data abstraction for modeling digital entities.
  - Encapsulates content and metadata
  - Consumed and/or produced by Planets preservation services
- Provides minimal and generic model for data management
  - Stored in Metadata Repository
- Does not prescribe serialization schema
  - May be created from DC/ORE RDF record and be
  - serialized using METS/PREMIS schemas.



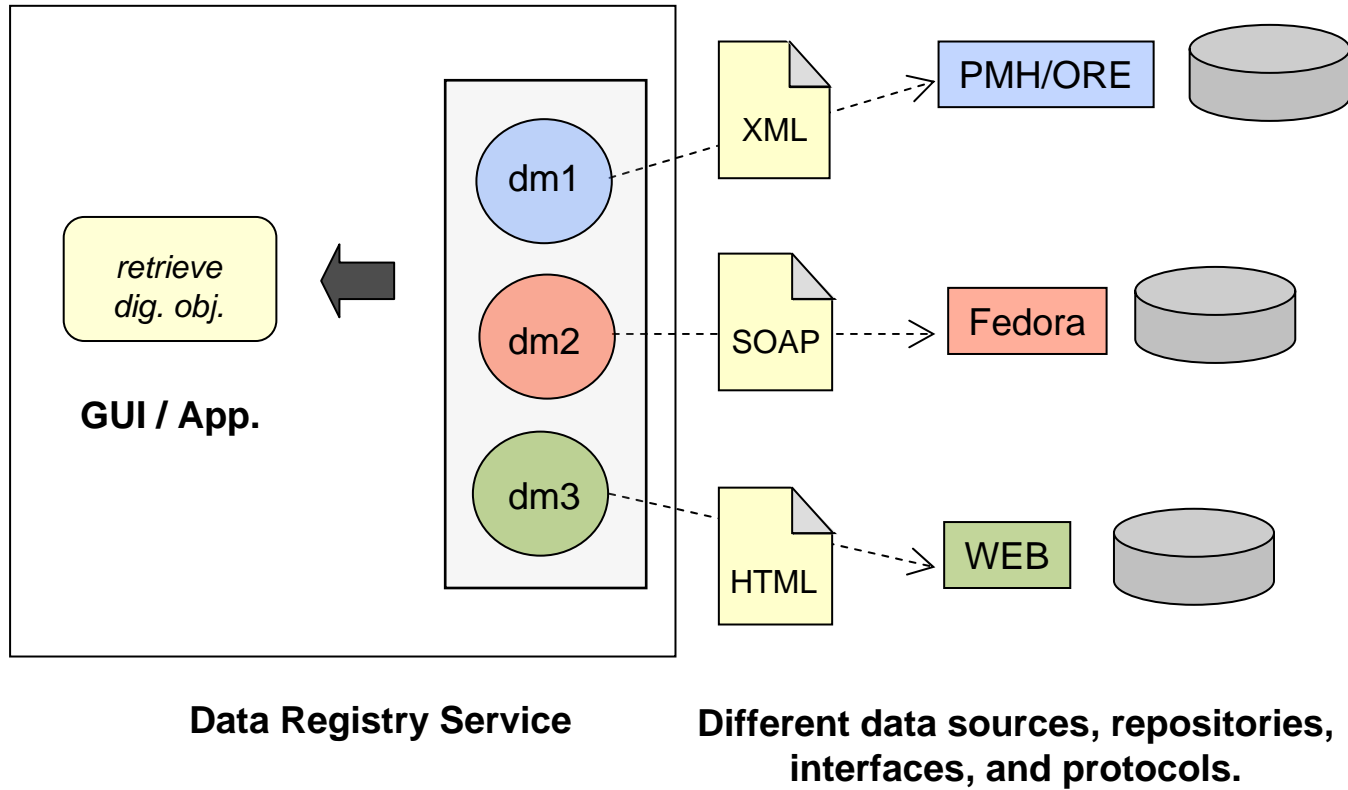
# Digital Objects (the Nouns)



## Digital Object Managers

- Individual adapters for retrieving (& storing) Planets DOs
  - Provide access to existing repositories.
  - Map metadata records to Planets DOs
  - Ingest digital objects to Planets data repositories
- Current implementation for
  - retrieving OAI-PMH records, BL digitized newspaper, Web resources, Amazon S3 buckets, ...
- Planets Data Registry services (ingesting DOs) based on Apache Jackrabbit and Fedora Commons.

# Digital Object Managers





PLANETS Testbed - Browse Data

PLANETS FTP AREA

You may upload data to the PLANETS FTP area using your wiki username and password and the following host name and port number:

Host name:  
**ftp://www.planets-project.eu**

RESTRICT RESULT SET

The option to restrict the data based on supplied criteria will be added in a later version.

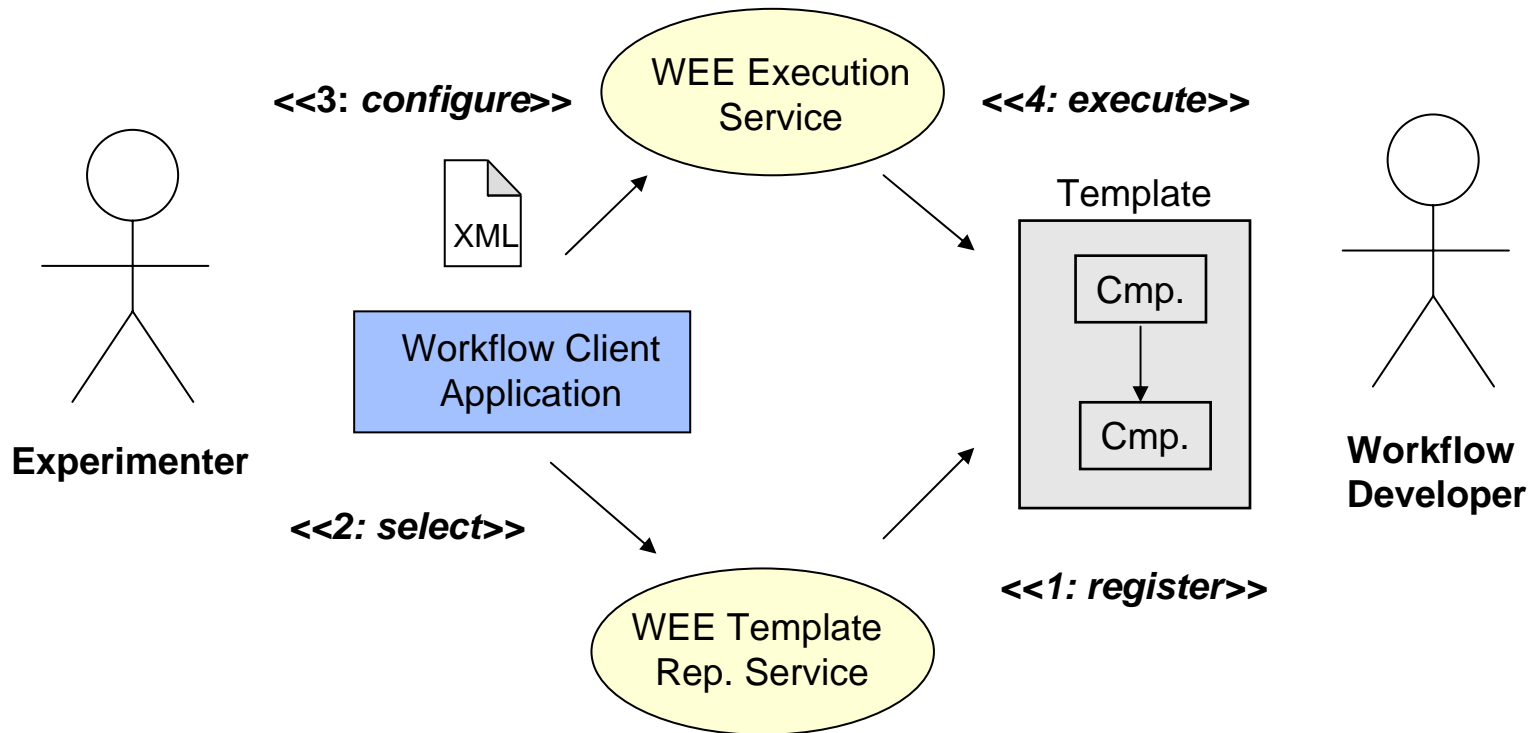
- ONB-OAI
- planets-ftp
- planets-s3
- bl-newspaper
- nasa
  - world\_2km
  - world\_500m
  - world\_8km
  - world\_big

	Name	Size	Type
<input type="checkbox"/>	world.200401.3x21600x10800.jpg		File
<input type="checkbox"/>	world.200401.3x21600x10800.png		File
<input type="checkbox"/>	world.200402.3x21600x10800.jpg		File
<input type="checkbox"/>	world.200402.3x21600x10800.png		File
<input type="checkbox"/>	world.200403.3x21600x10800.jpg		File
<input type="checkbox"/>	world.200403.3x21600x10800.png		File
<input type="checkbox"/>	world.200404.3x21600x10800.jpg		File
<input type="checkbox"/>	world.200404.3x21600x10800.png		File
<input checked="" type="checkbox"/>	world.200405.3x21600x10800.jpg		File
<input checked="" type="checkbox"/>	world.200405.3x21600x10800.png		File
<input checked="" type="checkbox"/>	world.200406.3x21600x10800.jpg		File
<input type="checkbox"/>	world.200406.3x21600x10800.png		File
<input type="checkbox"/>	world.200407.3x21600x10800.jpg		File
<input type="checkbox"/>	world.200407.3x21600x10800.png		File

## Workflow Definitions / Templates

- Separation of concerns:
  - Fragments of complex workflow logic are implemented by <<*workflow developers*>>
  - <<*Experimenters*>> selected from predefined templates, configure them, and execute individual processes.
- Templates implement abstract and reusable processes definitions based on *level-on* operations (API) and decision logic.
- Execute in *trusted* environment (level-two)
  - handle digital objects in metadata repository and
  - basis for recording provenance and preservation information

# Workflow Execution Engine (WEE) Service



## Workflow Control Panel

Manage Templates Load Workflow Edit Workflow Select Digital Objects Workflow Status

## Edit an XML Workflow description

Services				Parameters		
Service ID	Type	Endpoint	Name/Value Pairs		Add Parameter	
identify	Identify	DROID Identification Service	Select an Endpoint...	Name	Value	Edit Parameters
migrate	Migrate	GraphicsMagickMigrate	Select an Endpoint...	Name	Value	Edit Parameters

Save Workflow

Clear Workflow

Export XML

Reset

- Select an Endpoint...
- Select an Endpoint...
- None
- GraphicsMagickMigrate
- ImageMagickMigrate
- Mdb2SiardMigrate

Name	Value
compressionType	None, BZip, Fax, Group4, JPEG, Lossless, LZW, RLE, Zip
imageQuality	0-100

## Active Template

eu.planets\_project.ifr.core.wee.impl.templates.IdentifyMigrateTemplate

## Active Workflow

none (in memory)

## Input Objects Selected

false

Submit Workflow

Error Messages

## Summary

- Research infrastructure for
  - integrating variety of tools and repositories
  - executing defined preservation operations
  - recording provenance and preservation metadata
- Not necessary an “out-of-the-box” solution
  - Extensible network of services,
  - Public deployment,
  - Allows sharing of resources and results.
- Downloadable package available for local installation of selected preservation tools/services.



Fin