

SHIWA Repository User Manual

This manual documents the SHIWA workflow repository. Sections 1-2 describe the key entities, actors and use cases, section 3-7 describe how the provided features can be accessed via the GUI.

Table of Contents

1. Introduction	2
Background	2
Project goal	2
Key Components	3
Links	3
How to register as a user	3
2. Entity, Actor & Use-case Specification	3
2.1 Entity Definitions	3
2.2 Actor Definitions	4
2.3 Repository Model	4
2.4 Use Cases	5
2.5 Implementation Life-cycle	6
3. GUI Structure	6
4. Browse Workflows	7
4.1 List and search published workflows	7
4.1.1 Workflow Browse view	7
4.1.2 Workflow Table view (Figure 6)	8
4.1.3 Workflow details (Figure 7)	9
4.1.4 Workflow Attributes (Figure 8)	10
4.1.5 Workflow files & Download (Figure 9)	11
4.1.6 Workflow Implementations (Figure 10)	12
5. Browse Implementations	12
5.1 List and search implementations	12
5.1.1 Implementation Browse view	12
5.1.2 Implementation Table view	13
5.1.3 Implementations details (Figure 13)	14
5.1.4 Implementation Attributes	15
5.1.5 Implementation files & Download	16
6. Repository metadata graph (Figure 16)	17
7. Limitations	18

1. Introduction

Background

Researchers of all disciplines, from Life Sciences and Astronomy to Computational Chemistry, create and use ever-increasing amounts of complex data, and rely more and more on compute-intensive modelling, simulation and analysis.

Scientific workflows have become a key paradigm for managing complex tasks and have emerged as a unifying mechanism for handling scientific data. Workflows capture the essence of the scientific process, providing a means to describe it via logical data- or work-flows. Workflows are mapped onto concrete Distributed Computing Infrastructures (DCIs) to perform large-scale experiments.

The learning curve for reusing workflows, however, is still steep because workflows typically have their own user interfaces/APIs, description languages, provenance strategies, and enactment engines, which are not standard and do not interoperate. Workflow integration or reuse therefore is currently impractical, thereby inhibiting the growth in uptake and proliferation of workflows in scientific practice.

Project goal

User communities from all around Europe use many kinds of different workflow languages. Communities develop their workflows using one of the workflow engines. Workflow development, testing and validation is a time consuming process and it requires specific expertise. These limit the number of available workflows, so it is important to reuse them. Workflows developed for one workflow system is normally not compatible with workflows of other workflow systems. In the past if two user communities using different workflow systems wanted to collaborate, they had to create the workflows from scratch to transform them to the desired workflow languages. This situation can be resolved by emerging new workflow interoperability technologies. The goal of SHIWA is to develop such technologies.

According to the new SHIWA technologies publicly available workflows can be used by different research communities working on different workflow systems and are enabled to run on multiple distributed computing infrastructures. As a result workflow communities are not locked anymore in to their selected workflow system and its supported distributed computing infrastructure.

Key Components

SHIWA develops, deploys and operates the SHIWA Simulation Platform to offer users production-level services supporting workflow interoperability. As part of the SHIWA Simulation Platform the SHIWA Repository facilitates publishing and sharing workflows, and the SHIWA Portal enables their actual enactment and execution in all the DCIs available in Europe. Use cases targeting various scientific domains will serve to drive and evaluate this platform from a user's perspective.

Links

The SHIWA homepage is <http://www.shiwa-workflow.eu>

The SHIWA Simulation platform can be found at <http://shiwa-portal.cpc.wmin.ac.uk/liferay-portal-6.0.5>

The SHIWA Repository can be found at <http://shiwa-repo.cpc.wmin.ac.uk>

How to register as a user

All the processes described in this user manual can be performed on the Public page by an E-Scientist, without the requirement to login. Should you want to register, to enable workflow development, please send an email to weingan@wmin.ac.uk.

2. Entity, Actor & Use-case Specification

2.1 Entity Definitions

Workflow. This entity represents an abstract workflow. It describes the inputs and outputs and explains what the workflow does in an abstract manner, provides example inputs and outputs (configurations), and some further information.

Implementation. This entity represents an implementation of a workflow. It strictly follows the input and output definitions of the workflow and implements the functionality given in the workflow description. It contains or references (via eg. URLs) all the files and also holds dependencies to run the workflow on its associated workflow engine.

Engine. This entity represents a workflow engine that is able to interpret and execute a given implementation.

User. This entity represents a repository user associated with a user role.

Group. This entity grants read/write/download rights to a particular workflow for a set of users (the members of the group).

Platform. This entity describes in which desktop Grid and/or service Grid environment the implementation can be executed.

Files. This entity contains the files related to workflows and their implementations.

2.2 Actor Definitions

E-scientist. This actor is the consumer of the contents of the repository, i.e. workflow engines and workflows to run their experiments. This actor does not require credentials to login.

Workflow Developer. This actor is the creator and maintainer of the contents of the repository, i.e. workflows and their implementations and configurations.

Repository Administrator. This actor is a system administrator has the highest role among the actors. His task is to maintain the SHIWA repository and to support all other users.

Validator. This actor is a computer scientist who wants to test workflows and implementations created by workflow developers and give feedback. The actor should find non-validated workflows submitted by workflow developers, download packages and sample inputs and attempt to run the workflows and implementations. After successful validation he gives a feedback about the workflows and implementations and makes them available for E-scientists by marking them as validated.

2.3 Repository Model

Users represent all actors (e-scientists, workflow developers, validators and repository administrators) see Figure 1. We distinguish among actors based on their roles and the corresponding access rights. Workflow developers may own workflows and their implementations. E-scientists can browse validated workflows and their validated implementations, download and run them on the SHIWA Simulation Platform.

The repository also enables workflow developers and repository administrator to create and manage groups. Groups support controlled access to workflows and their implementations. Workflows have implementations and attributes.

Implementations have attributes and files. They are also associated with workflow engines.

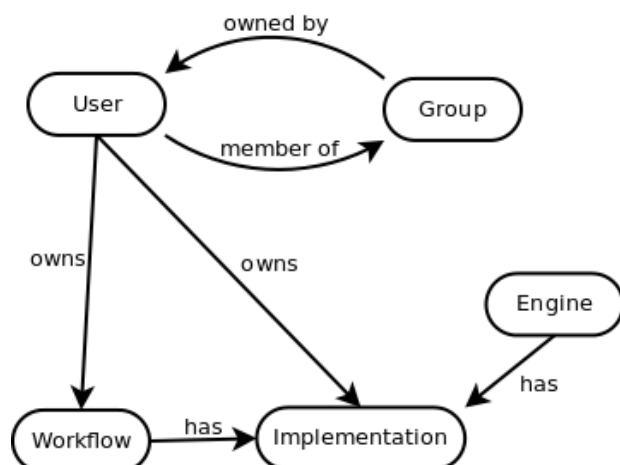


Figure 1: Repository model

2.4 Use Cases

This user manual only covers the use case of the E-scientist. Use cases of all actors can be found in the administrator manual. E-scientist are the consumers of the contents of the repository and can access the following functionality (see Figure 2).

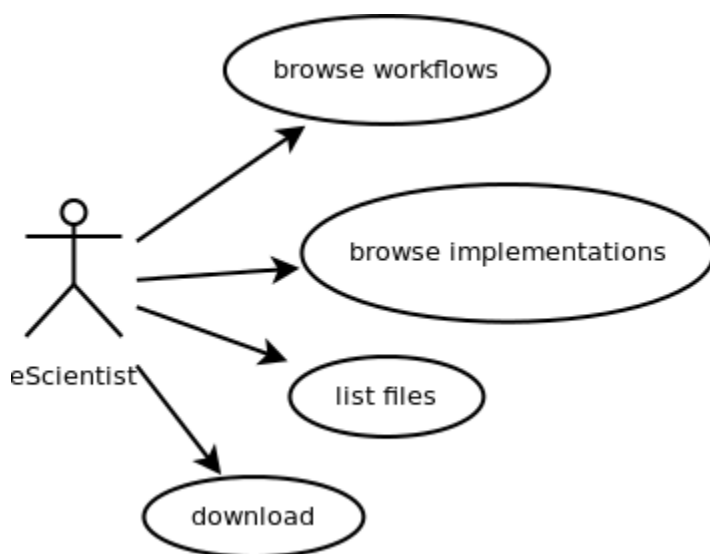


Figure 2: Use Case

- **Browse workflows (see section 4. Browse Workflows)**
Browsing includes searching and listing workflows based on their metadata.
- **Browse implementations (see section 5. Browse Implementations)**
E-scientists can browse implementations of the workflows selected by the "Browse workflows" operation.
- **List files (see section 5.1.5 Implementation files & Download)**
E-scientists can list files belonging to workflow implementations selected by the "Browse implementation" operation.

- **Download (see sections 4. Browse Workflows and 5. Browse Implementations)**
Users can download workflows and their related entities (implementations, configurations and files).

2.5 Implementation Life-cycle

Figure 3 depicts the life-cycle of an implementation

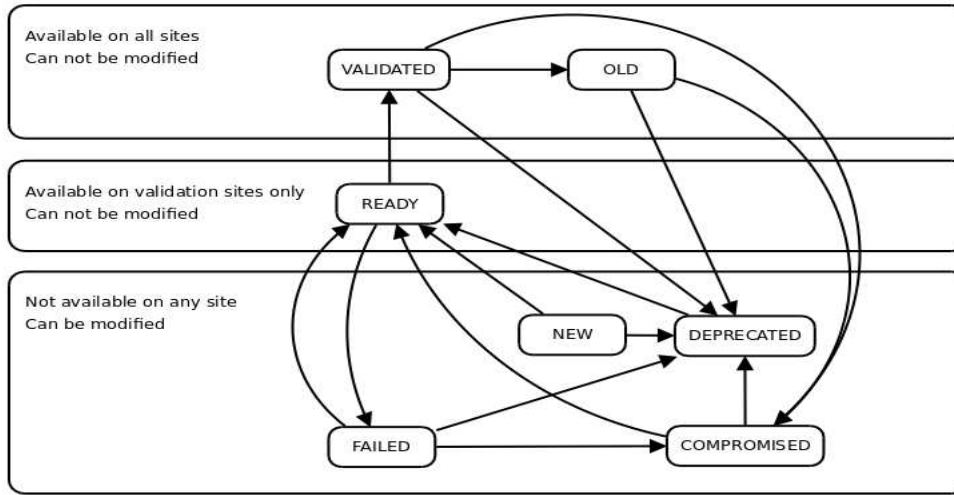


Figure 3: Implementation Life-cycle

3. GUI Structure

Repository features can be accessed using the main menu on the top. The following chapters go through the Workflows and Implementations tabs and describe the provided functionality. Information related to the selected tab is displayed in a table. Rows of a table can be filtered by entering text into the text field below any column title. Actions can be initiated using an “Actions” tab on the right. See illustration in Figure 4. This is the GUI structure of the public view, which can be viewed without any login credentials.

★ Welcome Workflows Implementations ? Documentation Log in				
Browse view Table view Actions				
25 (1 of 1)				
Name	Owner	Group	Status	Description
SolidBwaAlignment	vladimir	BWA	validated	Burrows-Wheeler Aligner (BWA) workflo...
simriMPI	fredj	CreatisSimri	validated	SIMRI is an MR simulator based on t...
simriMetaWorkflow	fredj	CreatisSimri	validated	This meta-workflow shows how to use d...
PreparationGateWF	fredj	GateWF	validated	This workflow splits a GATE simulatio...
mergeWF	fredj	GateWF	validated	The merge process of GATE simulation.
metaGATE	fredj	GateWF	validated	The meta workflow of GATE simulation...

Figure 4 : GUI Structure

4. Browse Workflows

4.1 List and search published workflows

4.1.1 Workflow Browse view

Figure 5 bellow, displays the Workflow Browse view. This view displays most of the information about published workflows and validated implementations in a convenient form.

Inputs, Outputs and Datasets can be expanded, to display more information.

Find Workflows

Find Workflows interface showing a search bar with a dropdown menu set to "All Domains" and a "Fetch" button. Below the search bar is a pagination bar indicating "(1 of 1)" results. The main content area displays a workflow card titled "Workflow: FetchImages" with a "Details" button. The card is divided into two columns: "Workflow Summary" and "Implementation Preview (1)".

Workflow Summary:

- Domain:** Demonstration
- Application:** Shiwa Image Manipulation Demo
- Owner:** Tamas Kukla
- Group:** shiwaExampleWfs
- Status:** validated
- Keywords:** web service, images
- Description:** This workflow fetches images using a web service and puts them in a zip file. Used for demonstration purposes.

Below the summary are expandable sections for:

- Inputs (1)
- Outputs (1)
- Data sets (1)

Implementation Preview (1):

FetchImagesTaverna1.7

- Engine:** Taverna(1.7)
- Version:** 1.0
- DCIs:** SHIWA VO
- Keywords:** Taverna, Images, Web Service
- Description:** Taverna 1.7 implementation of the workflow that is designed...

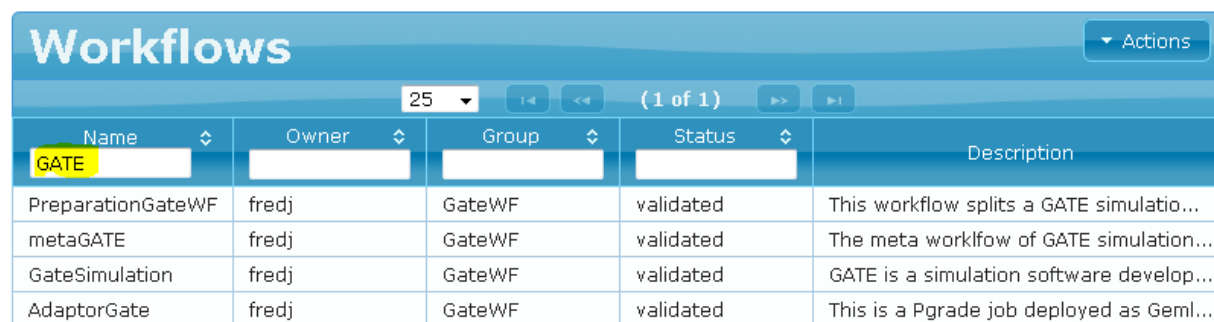
A second pagination bar is located at the bottom of the interface.

Figure 5: Workflows Browse view

Textual searches in workflow or implementation records can be performed using the search box. These searches can be restricted to domains of research. By clicking the Details button, more information is displayed about the Workflow and its validated implementations.

4.1.2 Workflow Table view (Figure 6)

The Workflows Table view can be used to list and filter published workflows by Name, Owner or Group.



Name	Owner	Group	Status	Description
GATE				
PreparationGateWF	fredj	GateWF	validated	This workflow splits a GATE simulatio...
metaGATE	fredj	GateWF	validated	The meta workflow of GATE simulation...
GateSimulation	fredj	GateWF	validated	GATE is a simulation software develop...
AdaptorGate	fredj	GateWF	validated	This is a Pgrade job deployed as Geml...

Figure 6: Workflows table view

All available information about workflows and their validated implementations can be viewed by selecting the workflow from this view (see sections 4.1.3 Workflow details (Figure 7), 4.1.4 Workflow Attributes (Figure 8), 4.1.5 Workflow files & Download (Figure 9) and 4.1.6 Workflow Implementations (Figure 10))

4.1.3 Workflow details (Figure 7)

Selected workflow: FetchImages



DetailsOwnerAccessAttributesFilesImplementations

ID:

1654

Description:

This workflow fetches images using a web service and puts them in a zip file. Used for demonstration purposes.

Figure 7: Workflow details page

4.1.4 Workflow Attributes (Figure 8)

Selected workflow: FetchImages

Details Owner Access Attributes Files Implementations		
Attributes		Actions
Name	Value	Expand/Collapse
▼ inputs		Reload
▼ port0001		
datatype	file	
description	This file contains the data request for fetching images from a given web service.	
title	Input zip	
▼ outputs		
▼ port0002		
datatype	file	
description	This file contains the gathered images.	
title	Output zip	
▼ datasets		
▼ dataset0001		
description	An example dataset providing an input	
port0001	example_input.zip	Download
tasktype	demonstration	
application	Shiwa Image Manipulation Demo	
domain	Demonstration	
keywords	web service, images	

Figure 8: Workflow attributes page

The Expand/Collapse button in the Action control can be used to display all the attributes of the workflow.

Workflow attributes can be listed by clicking on the Attributes tab of a particular Workflow.

Table 1 describes the metadata structure of the attributes, and provides example values Figure 16 presents the Workflow metadata structure. These attributes allow straightforward categorisation of workflows and improve the browsing and search operations significantly. The input and output attributes with their sub-attributes define inputs and outputs of the Workflow. The configuration attribute specifies values of input parameters passed to workflow inputs, and they can also specify example outputs.

SHIWA Workflow Repository – User Manual

Workflow metadata			Example value	Description	Table	Type	Mapping to SHIWA Desktop
id			1001	workflow identifier	workflow	int	
name			Factorial	workflow name	workflow	string	
owner_id			1008	workflow owner id	workflow	int	
group_id			exampleGroup	user group for defining access rights	workflow	int	
group_read			TRUE	whether group members can see wf/impl. data	workflow	bool	
group_download			TRUE	whether group members can download wf/impl. files	workflow	bool	
group_modify			TRUE	whether group members can modify wf/impl. data and upload files	workflow	bool	
others_read			TRUE	whether registered users can see wf/impl. data	workflow	bool	
others_download			TRUE	whether registered users can download wf/impl. files	workflow	bool	
published			TRUE	whether unregistered users can see wf/impl. data and download files	workflow	bool	
created			6/5/2011 13:12	workflow creation time	workflow	timestamp	
modified			6/7/2011 16:59	time of last modification	workflow	timestamp	
application			GATE	name of the application which the wf is part of	workflow_attr.	string	workflow->shiwa:application
description			This workflow ...	workflow description	workflow_attr.	string	
domain			Mathematics	scientific domain	workflow_attr.	string	workflow->shiwa:domain
keywords			factorial, integer	workflow keywords	workflow_attr.	string	
tasktype			demo	type of task the workflow represents	workflow_attr.	string	workflow->shiwa:tasktype
inputs	\			list of workflow inputs			
		port0001		first input port			shiwa:inport
		title	PositiveInteger	name of the port	workflow_attr.	string	shiwa:inport->dc:title
		datatype	file	data type of port	workflow_attr.	string	shiwa:inport->rdf:datatype
		description	this file contains an integer	port description	workflow_attr.	string	shiwa:inport->dc:description
outputs	\			list of output ports			
		file0002		first output port			shiwa-outport
		title	Factorial	name of the port	workflow_attr.	string	shiwa:outport->dc:title
		datatype	file	data type of port	workflow_attr.	string	shiwa:outport->rdf:datatype
		description	this file contains the factorial of the input integer	port description	workflow_attr.	string	shiwa:outport->dc:description
configurations	\			List of input/output configurations			
		conf0001		First configuration			
		description	This configuration...	A description of the configuration	workflow_attr.	string	shiwa:configuration->de:description
		port0001	input.dat	example value for port0001	workflow_attr.	string	shiwa:portref->rdf:value
		port0002	output.dat	example value for port0002	workflow_attr.	string	shiwa:portref->rdf:value

Table 1: Workflow metadata attributes

4.1.5 Workflow files & Download (Figure 9)

Selected workflow: FetchImages

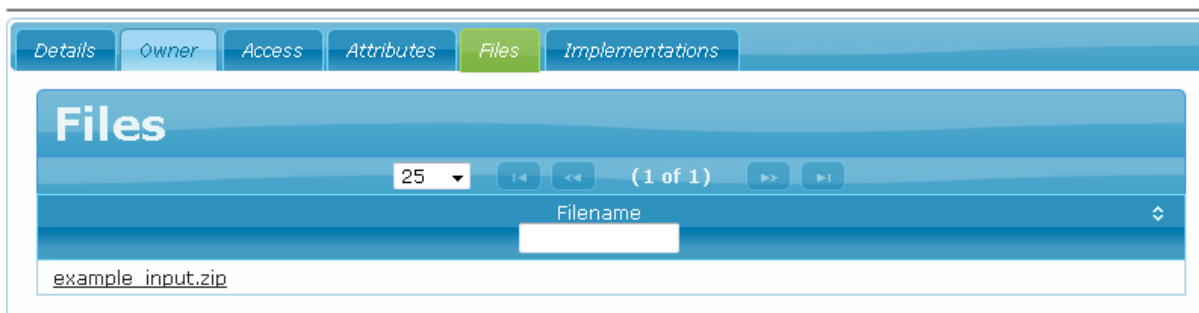


Figure 9: Workflows files page

Files associated with the abstract workflow definition can be downloaded from this page by right-clicking the files and selecting “Save as” or appropriate, as per your browser.

If files are not permitted to be downloaded, only their names will be displayed, but no live-links for download will be generated.

4.1.6 Workflow Implementations (Figure 10)

Selected workflow: FetchImages

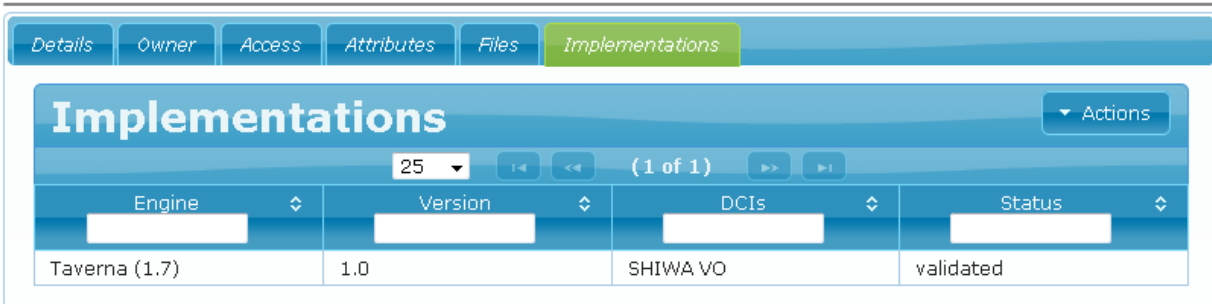


Figure 10: Workflows Implementations page

This lists the implementations of the selected workflow.

Selecting any of the validated implementations will direct you to the Implementation details of that Implementation (see section 5.1.3 Implementations details (Figure 13)).

5. Browse Implementations

5.1 List and search implementations

5.1.1 Implementation Browse view

Figure 11 below, displays the Implementation Browse view. This view displays most of the information about validated implementations in a convenient form.

Dependencies and Configurations can be expanded, to display more information.

Find Implementations

All Domains
Search
Show All
Refresh

(1 of 1)
1
10

Workflow: FetchImages | Engine: Taverna(1.7) | Implementation version: 1.0

Graph

Implementation summary

Title: FetchImagesTaverna1.7
Workflow: [FetchImages](#)
Engine: Taverna(1.7)
Version: 1.0
Status: validated
Validator: [Tamas Kukla](#)
Language: Scufl
Domain: Demonstration
DCIs: SHIWA VO
GEMMLCA id: Taverna-WF2g
Licence: Creative Commons Attribution 3.0 Unported License
Definition: [fetchImages.xml](#)
Keywords: Taverna, Images, Web Service

Description: Taverna 1.7 implementation of the workflow that is designed to be executed on ngs worker nodes.

Dependencies (3)

Name	Type	Description
ImageWebService	Service	A web service for gathering images.
Parameter Mapping	Other	This file maps files of the input zip to workflow ports.
VO for execution in SSP	DCI	The VO the user has to be member of to execute this workflow in the SSP

Configurations (1)

Configuration 1

Dependency Name	Value
ImageWebService	http://moby.ucalgary.ca/moby/MOBY-Central.pl
Parameter Mapping	example_params.map
VO for execution in SSP	SHIWA VO

Figure 11: Browse Implementation view

Textual searches in workflow records can be performed using the search box. These searches can be restricted to specific domains of research.

5.1.2 Implementation Table view

The Implementations Table View can be used to list and filter validated Implementations by Workflow, Engine, Version or DCI.

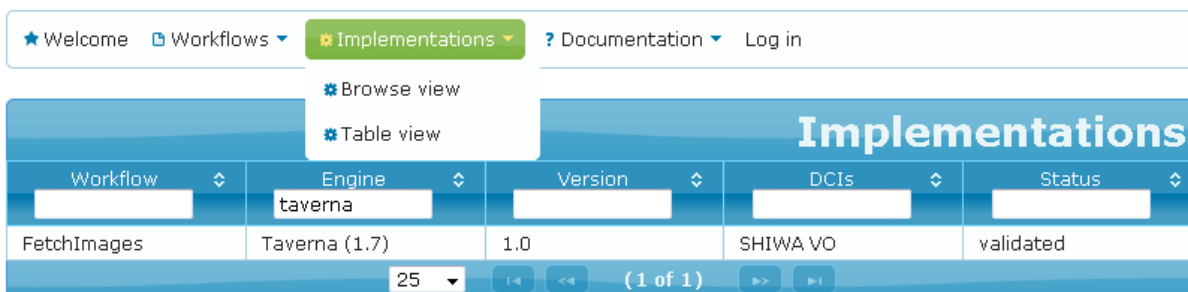


Figure 12: Implementations Table view

All available information about Implementations can be viewed by selecting the Implementation from this view (see 5.1.3 Implementations details (Figure 13), 5.1.4 Implementation Attributes and 5.1.5 Implementation files & Download)

5.1.3 Implementations details (Figure 13)

Workflow: FetchImages

Engine: Taverna(1.7)

Implementation version: 1.0

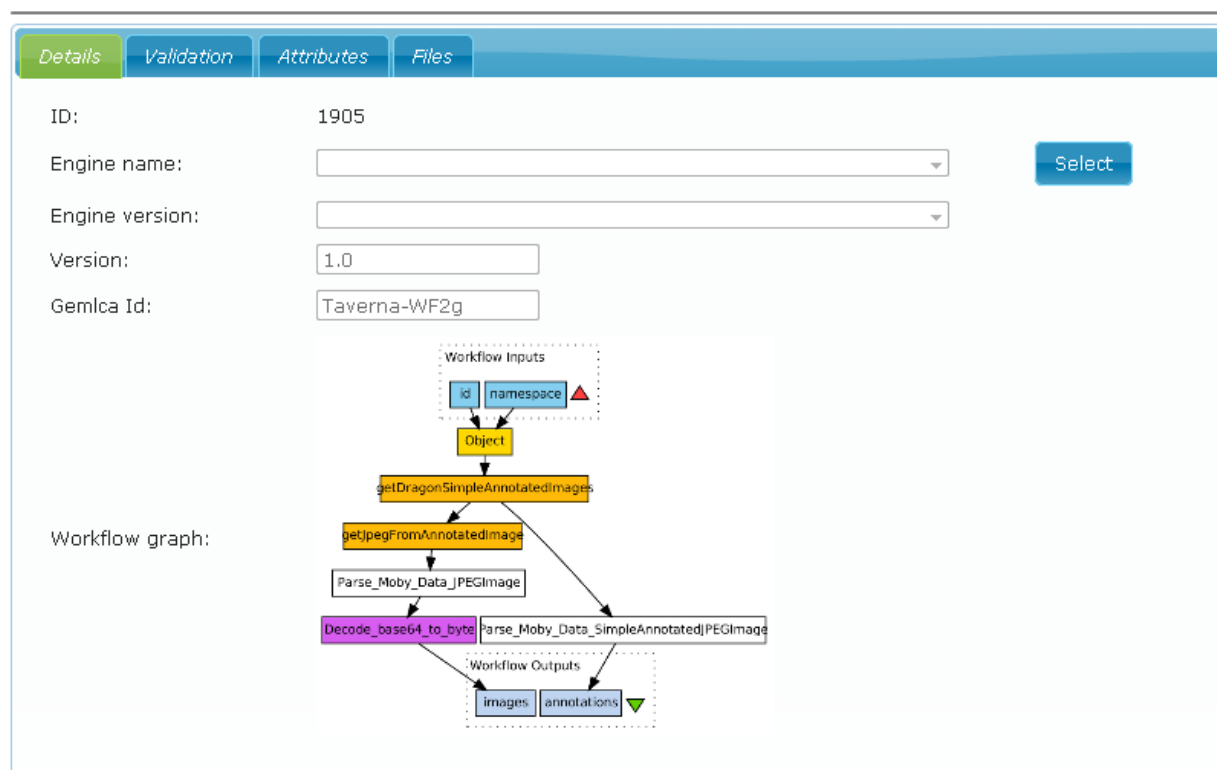


Figure 13: Implementation Details page

The *Validation* tab can be used to see the validation life-cycle status.
The *Attributes* tab can be used to open the attributes of the selected Implementation.

SHIWA Workflow Repository – User Manual

The *Files* tab can be used to list and download files held in the repository for the selected implementation.

5.1.4 Implementation Attributes

Implementation attributes can be opened by clicking on the attributes tab of a given implementation as illustrated in Figure 14. The left column of the attributes table contains attribute names, while the right column contains attribute values.

Similarly to workflows, the metadata template is used to help the definition of most common attributes. The three key attributes are: definition, dependencies and configurations. The definition attribute, is the workflow definition file i.e. the executable to be interpreted by the workflow engine. The dependency attribute can be any requirement of the particular implementation. These can include for instance files, executables, libraries or VO memberships required for execution. Configuration attributes resolve these dependencies.

Table 2 describes each attribute and provides example values. Figure 16 illustrates the metadata structure.

Implementation metadata			Example value	Description	Table	Type	Mapping to SHIWA Desktop
id			1002	implementation identifier	implementation	int	
workflow_id			1001	identifier of the abstract workflow that the impl. implements	implementation	int	
engine_id			1005	workflow engine identifier	implementation	int	
version			1.01	implementation version	implementation	string	
created			6/15/2011 4:12	workflow creation time	implementation	timestamp	workflow->dcterms:created
modified			6/21/2011 11:24	time of last modification	implementation	timestamp	workflow->dcterms:modified
state			VALIDATED	implementation status	implementation	enum	
uuid			1234-1234-1234	uuid of implementation	imp_attr.	string	workflow->dc:identification
title			FetchImagesTaverna	title of the implementation	imp_attr.	string	workflow->dc:title
description			this implementation is	implementation description	imp_attr.	string	workflow->dc:description
definition			workflow.xml	workflow descriptor file	imp_attr.	string	workflow->shiwa:definition
graph			workflow.png	workflow graph screenshot	imp_attr.	string	
language			SCUFL	language of the workflow descriptor	imp_attr.	string	workflow->shiwa:language
rights			© SHIWA	copyright information	imp_attr.	string	workflow->dc:rights
licence			Demo licence	licence information	imp_attr.	string	workflow->dcterms:licence
keywords			Taverna, Images, Web Service	keywords used for searching	imp_attr.	string	
dependencies	\	dep0001	\	List of dependencies: files needed for executing factorial.sh. It can be empty in the case of DGs.	imp_attr.		
				first dependency	imp_attr.		shiwa:dependency
			title	Title of the first dependency	imp_attr.	string	shiwa:dependency->dc:title
			description	A web service for gathering images.	imp_attr.	string	shiwa:dependency->dc:description
			dep0002	Description of first dependency	imp_attr.		shiwa:dependency
			title	Title of the second dependency	imp_attr.	string	shiwa:dependency->dc:title
configurations	\	conf0001	\	This file maps files of the input zip to workflow ports.	imp_attr.	string	shiwa:dependency->dc:description
				Description of second dependency	imp_attr.		
				List of dependency configurations. A configuration resolves all dependencies of the executable. It can be empty if no dependencies.	imp_attr.		
			description	This configuration...	imp_attr.	string	shiwa:configuration->dc:description
			dep0001	http://moby.ucalgary.ca/	imp_attr.	string	shiwa:dependencyref->rdf:value
			dep0002	example_params.map	imp_attr.	string	shiwa:dependencyref->rdf:value

Table 2: Implementation metadata attributes

Workflow: FetchImages
Engine: Taverna(1.7)
Implementation version: 1.0

Attributes		
Name	Value	Actions
▼ dependencies		Expand/Collapse Reload
▼ dep0001		
type	Service	
description	A web service for gathering images.	
title	ImageWebService	
▼ dep0002		
type	Other	
description	This file maps files of the input zip to workflow ports.	
title	Parameter Mapping	
▼ dep0003		
type	DCI	
description	The VO the user has to be member of to execute this workflow in the SSP	
title	VO for execution in SSP	
▼ configurations		
▼ conf0001		
dep0001	http://moby.ucalgary.ca/moby/MOBY-Central.pl	
dep0002	example_params.map	Download
dep0003	SHIWA VO	
title	FetchImagesTaverna1.7	
description	Taverna 1.7 implementation of the workflow that is designed to be executed on ngs worker nodes.	
definition	fetchImages.xml	Download
graph	Taverna_wf2.png	Download
language	Scufl	
rights		
licence	Creative Commons Attribution 3.0 Unported License	
keywords	Taverna, Images, Web Service	
uuid		

Figure 14: Implementation attribute table

The definition file can be downloaded from this page by clicking on the *download* live-link.

5.1.5 Implementation files & Download

As with Workflow files, these can be downloaded from the Implementation Files view, by clicking on the respective live-link (see Figure 15).

Workflow: FetchImages
Engine: Taverna(1.7)
Implementation version: 1.0

Files		
25	(1 of 1)	
Filename		
Taverna_wf2.png		
example_params.map		
fetchImages.xml		

Figure 15: Implementation Files and download

6. Repository metadata graph (Figure 16)

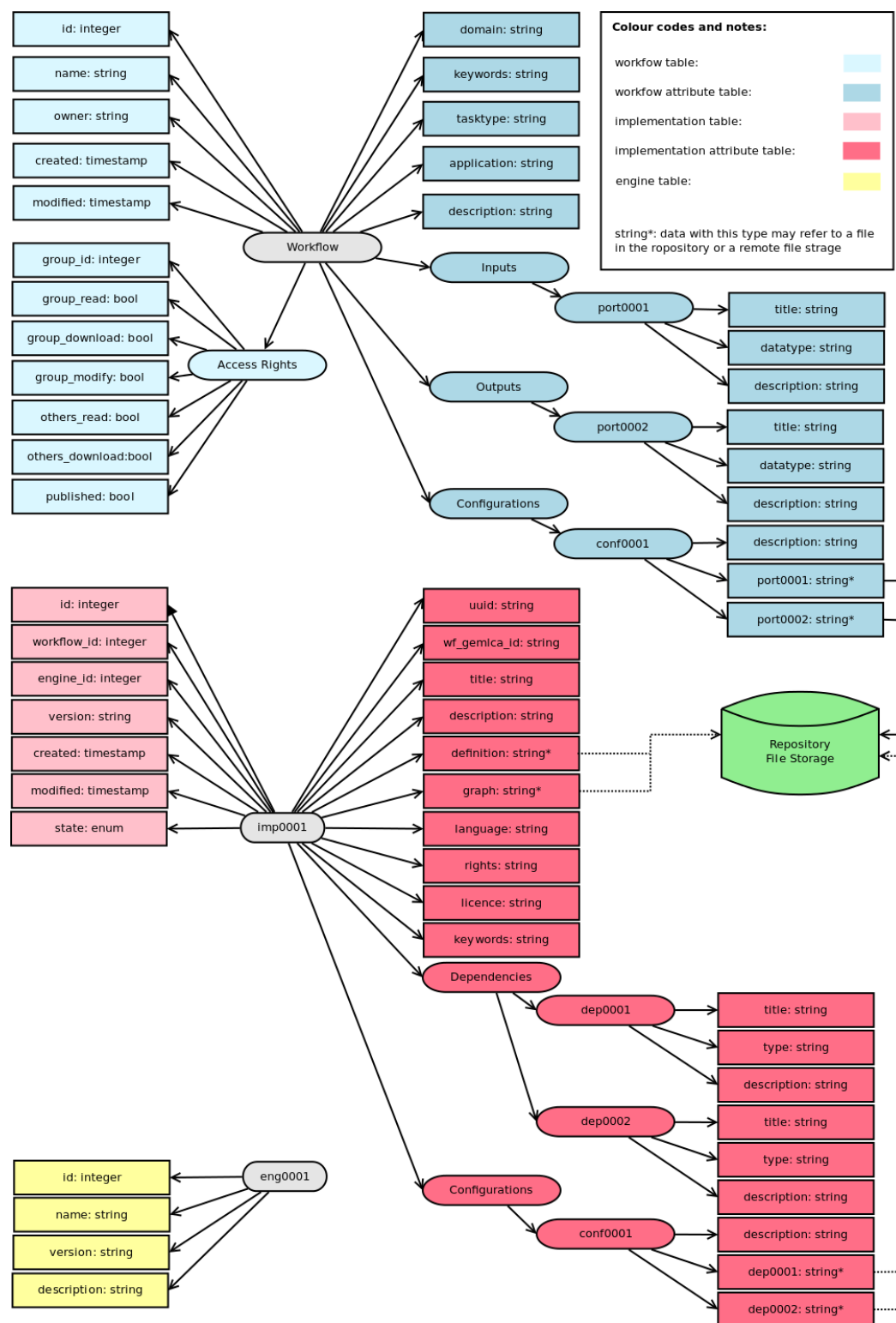


Figure 16: SHIWA Repository metadata graph

7. Limitations

- It is not recommended to open the repository in multiple browser tabs.