

Workflow Engine Interface Specification

This document describes the command line interface of the Workflow engines available in the GEMLCA service of the SHIWA Simulation Platform. The aim of this document is to help workflow developers who want to make their workflows runnable in the SSP by deploying them to GEMLCA.

Prerequisites:

In order to deploy a workflow implementation to GEMLCA, it has to be uploaded to the SHIWA Repository. In order to do this, please refer to the SHIWA Repository manual: https://www.cpc.wmin.ac.uk/cpcsite/images/e/e0/Admin_Manual.pdf. This manual also describes the new GUI features that enable workflow deployment to GEMLCA.

List of engines discussed in this document:

All engines and workflow implementations runnable in the SSP are to be accessed via the official GEMLCA service of the SSP at: <https://gemlcadevel.cpc.wmin.ac.uk:8443/wsrf/services/uk/ac/wmin/cpc/gemlca/frontend>. The following engines are discussed in this document:

- GIB-TavernaEngine
- GIB-TrianaEngine
- GIB-KeplerEngine
- GIB-MoteurEngine-*
- GIB-LoniWS
- GIB-Galaxy
- GIB-ProActive-PACAGrid
- GWES-medigrid-GIB
- GIB-Python2.4
- GIB-AskalonEngine

Notes:

This document is to be extended to describe all workflow engines available in the SSP via GEMLCA.

GIB-KeplerEngine

The GIB-KeplerEngine is installed at the Westminster Cluster and it executes its jobs locally. All inputs and outputs are files and filenames are hard coded in the workflows. Input and output files are not expected to be represented on the commandline when the engine is invoked, but they are transferred to/from the execution directory.

engGemlcaId: GIB-KeplerEngine, deployed by Tamas Kukla
(kuklat1@staff.westminster.ac.uk)

command line arguments the engine expects:

```
>kepler-1.0-ad-local.sh -w wf_desc.xml
```

workflow descriptor: it is automatically passed to the engine by GEMLCA, only input and output files have to be defined by the workflow developer

files the engine expects and generates: all input file names are fixed, all output file names are fixed, files provided by the user as default values are automatically renamed

Example workflow implementations: to be added

For this engine it is recommended to define all input files it needs and all output files it generates for the abstract workflow. In addition, at least one example dataset providing files for all inputs and outputs is also needed. If this is provided, then all parameters can be defined according to the following table:

Parameter attribute	Inputs	Outputs
Type	INPUT_PORT or DEPENDENCY	OUTPUT_PORT
Command Line	Always false	Always false
Switch Name	Always empty string	Always empty string

If not all expected input and output files are defined or dataset is not provided for the abstract workflow, then CUSTOM parameters can be used for the workflow implementation according to the following table:

Parameter attribute	Inputs	Outputs
Type	CUSTOM	CUSTOM
Default value	String (a file name)	String (a file name)
File	Always true	Always true
Input	Always true	Always false
Command Line	Always false	Always false
Switch Name	Always empty string	Always empty string

GIB-TavernaEngine

The GIB-TavernaEngine is installed at the Westminster Cluster, it executes its jobs locally and it can invoke remote web services. All inputs and outputs should be files and are expected to be represented on the commandline when the engine is invoked and they transferred to/from the execution directory.

engGemlcaId: GIB-TavernaEngine, deployed by Tamas Kukla
(kuklat1@staff.westminster.ac.uk)

command line arguments the engine expects:

```
>taverna-1.7-ad-local.sh -w wf_desc.xml  
[-input in_file1 [...] -input in_file2]
```

workflow descriptor: it is automatically passed to the engine by GEMLCA, only input and output files have to be defined by the workflow developer

files the engine expects and generates: all inputs are to be defined on the command line when the engine is invoked, output file names should not be provided on the command line.

Example workflow implementations: to be added

For this engine it is recommended to define all input files it needs and all output files it generates for the abstract workflow. In addition, at least one example dataset providing files for all inputs and outputs is also needed. If this is provided, then all parameters can be defined according to the following table:

Parameter attribute	Inputs	Outputs
Type	INPUT_PORT or DEPENDENCY	OUTPUT_PORT
Command Line	Always true	Always false
Switch Name	-input	Always empty string

If not all expected input and output files are defined or dataset is not provided for the abstract workflow, then CUSTOM parameters can be used for the workflow implementation according to the following table:

Parameter attribute	Inputs	Outputs
Type	CUSTOM	CUSTOM
Default value	String (a file name)	String (a file name)
File	Always true	Always true
Input	Always true	Always false
Command Line	Always true	Always false
Switch Name	-input	Always empty string

GIB-TrianaEngine

The GIB-TrianaEngine is installed at the Westminster Cluster and it executes its jobs locally. All inputs can be files or strings, while outputs can be only files. Inputs and outputs are expected to be represented on the commandline when the engine is invoked and they are transferred to/from the execution directory.

engGemlcaId: GIB-TrianaEngine, deployed by Tamas Kukla
(kuklat1@staff.westminster.ac.uk)

command line arguments the engine expects:

```
> triana-3.2.3-ad-local.sh -w wf_desc.xml  
[jobId.inPortId input1 [...] jobId.inPortId input2]  
[jobId.outPortId* output1 [...] jobId.outPortId* output2]
```

workflow descriptor: it is automatically passed to the engine by GEMLCA, only input and output files have to be defined by the workflow developer

files the engine expects and generates: all inputs and outputs are to be defined on the command line when the engine is invoked

Example workflow implementations: to be added

For this engine it is recommended to define all inputs it needs and all output files it generates for the abstract workflow. In addition, at least one example dataset providing files for all inputs and outputs is also needed. If this is provided, then all parameters can be defined according to the following table:

Parameter attribute	Inputs	Outputs
Type	INPUT_PORT or DEPENDENCY	OUTPUT_PORT
Command Line	Always true	Always true
Switch Name	JobId.inPortId*	JobId.outPortId*

If not all expected inputs and outputs are defined or dataset is not provided for the abstract workflow, then CUSTOM parameters can be used for the workflow implementation according to the following table:

Parameter attribute	Inputs	Outputs
Type	CUSTOM	CUSTOM
Default value	String	String (a file name)
File	No restriction	Always true
Input	Always true	Always false
Command Line	Always true	Always true
Switch Name	JobId.inPortId*	JobId.outPortId*

* Note that in Triana 3.2.3 each job and port has an identifier. These have to be defined on the command line when the engine is invoked in order to map them to workflow inputs and outputs.

GIB-MoteurEngine-*

The GIB-MoteurEngines are installed remotely. GEMLCA invokes a Moteur client that communicates with the given Moteur service. Non-file inputs can be defined and passed to the engine client in an xml based input document. Input files are not transferred; users have to make sure that they are provided where the workflow expects them. The following engines are registered in GEMLCA:

engGemlcaId	Contact	Moteur service url
GIB-MoteurEngine-Dirac	Tamas Kukla kuklat1@staff.westminster.ac.uk	https://ui.egee.creatis.insa-lyon.fr/cgi-bin/m2Server-prod/moteur_server
GIB-MoteurEngine-gLite		https://ui.egee.unice.fr/cgi-bin/m2Server-gLite/moteur_server
GIB-MoteurEngine-gLite2		https://ui.egee.creatis.insa-lyon.fr/cgi-bin/m2Server-gLite/moteur_server
GIB-MoteurEngine-Vlmed	Vladimir Korkhov v.korkhov@amc.uva.nl	http://orange.ebioscience.amc.nl/WorkflowWS/MoteurService

command line arguments the engine client expects:

```
>workflow-execute.sh wf_desc.xml moteur_service_url  
input_doc.xml
```

workflow descriptor and moteur service url: they are both automatically passed to the engine client by GEMLCA, only the input document has to be defined by the workflow developer

files the engine expects and generates: the engine client expects a single input file, an input document xml, that describes all non-file inputs. The user should provide the input files where the workflow expects it and should be able to get the output files manually.

Example workflow implementations: to be added

For these engines it is recommended to define a single input that represents the input document xml file as a CUSTOM parameter. Standard output and error will be automatically generated as workflow outputs.

Parameter attribute	Input
Type	CUSTOM
Default value	String (a file name)
Title	Input doc XML
File	Always true
Input	Always true
Command Line	Always false
Switch Name	Always empty string

GIB-LoniWS

The GIB-LoniWS service is installed remotely. GEMMLCA invokes a LoniWS client that communicates with the given Loni service. Non-file inputs can be defined and passed to the engine client in an xml based input document. Input files are not transferred; users have to make sure that they are provided where the workflow expects them. The following engines are registered:

engGemlcaId: GIB-LoniWS, deployed by Tamas Kukla
(kuklat1@staff.westminster.ac.uk)

command line arguments the engine client expects:

```
>moteur-gLite.sh wf_desc.xml input_doc.xml credential.xml  
loni_service_url output.txt
```

workflow descriptor: it is automatically passed to the engine client by GEMMLCA. The input document, credential file has to be defined by the workflow developer

files the engine expects and generates: the engine client expects the input document file, the credential file and the loni service url on the command line. There is a single output file the engine passes back to the client. This file is to be defined as the last parameter on the command line. If the workflow expects/generates further input/output files on remote locations, the user has to ensure that the inputs are provided and should be able to get the output files manually.

Example workflow implementations: to be added

For this engine it is recommended to define a single input that represents the input document xml file and a single output as CUSTOM parameters. Standard output and error will be automatically generated as workflow outputs.

Parameter attribute	Input	Input	Input	Output
Parameter id	Param0001	Param0002	Param0003	Param0004
Type	CUSTOM	CUSTOM	CUSTOM	CUSTOM
Default value	String (a file name)	String (a file name)	String (service url)	String (a file name)
Title	Input doc XML	Credentials XML	Loni Service URL	Output text
File	Always true	Always true	Always false	Always true
Input	Always true	Always true	Always true	Always false
Command Line	Always true	Always true	Always true	Always true
Switch Name	Always empty string	Always empty string	Always empty string	Always empty string

If the workflow developer wants to allow the execution of the workflow implementation always on the same loni service, then Param0003 can be set to fixed. In this case it will be hidden from the user, but will be always present when the client is invoked by GEMMLCA.

GIB-Galaxy

The GIB-Galaxy engine is installed at the Westminster Cluster and it executes its jobs locally. All inputs and outputs should be files and are expected to be represented on the commandline when the engine is invoked and they transferred to/from the execution directory.

engGemlcaId: GIB-Galaxy, deployed by Tamas Kukla
(kuklat1@staff.westminster.ac.uk)

command line arguments the engine expects:

```
>run_workflow.sh -w wf_desc.ga  
[-i in_port1 in_type1 in_file1 [...] -i in_port2 in_type2 in_file2]  
[-o out_port1 out_file1 [...] -o out_port2 out_file2]  
  
where  
    in_port: represents the id of a workflow input  
    in_type: represents the type of a workflow input (i.e: "bed")  
    out_port represents the id of a workflow output
```

workflow descriptor: it is automatically passed to the engine by GEMLCA, only input and output files, their types and workflow in/output id have to be defined by the workflow developer

files the engine expects and generates: all inputs and outputs are to be defined on the command line when the engine is invoked

Example workflow implementations in the SHIWA repository:
Exons-Galaxy-central-1.0

For this engine it is recommended to define all input files it needs and all output files it generates for the abstract workflow. In addition, at least one example dataset providing files for all inputs and outputs is also needed. If this is provided, then all parameters can be defined according to the following table:

Parameter attribute	Inputs	Outputs
Type	INPUT_PORT	OUTPUT_PORT
Command Line	Always true	Always true
Switch Name	-i in_port in_type	-o out_port

GIB-ProActive-PACAGrid

The GIB-ProActive-PACAGrid is installed remotely. GEMMLCA invokes a ProActive client that communicates with the PACAGrid ProActive service. All inputs and outputs are files and filenames do not need to be specified. Input and output files are not expected to be represented on the commandline when the engine is invoked, but they are transferred to/from the execution directory.

engGemlcaId: GIB-ProActive-PACAGrid, deployed by Noam Weingarten
(weingan@westminster.ac.uk)

command line arguments the engine expects:

```
>ProActive_Scheduler3.1.1.test.sh -w workflow.jobarch
```

workflow descriptor: it is automatically passed to the engine by GEMMLCA, only input and output files have to be defined by the workflow developer

files the engine expects and generates: all inputs and outputs are to be defined in the Execution, at the stage of the workflow's deployment to GEMMLCA.

Example workflow implementations: to be added

For this engine it is recommended to define all input files it needs and all output files it generates for the abstract workflow. In addition, at least one example dataset providing files for all inputs and outputs is also needed. If this is provided, then all parameters can be defined according to the following table:

Parameter attribute	Inputs	Outputs
Type	INPUT_PORT or DEPENDENCY	OUTPUT_PORT
Command Line	Always false	Always false
Switch Name	Always empty string	Always empty string

If not all expected input and output files are defined or dataset is not provided for the abstract workflow, then CUSTOM parameters can be used for the workflow implementation according to the following table:

Parameter attribute	Inputs	Outputs
Type	CUSTOM	CUSTOM
Default value	String (a file name)	String (a file name)
File	Always true	Always true
Input	Always true	Always false
Command Line	Always false	Always false
Switch Name	Always empty string	Always empty string

GWES-medigrid-GIB

The GWES-medigrid-GIB is installed remotely. GEMLCA invokes a GWES client that communicates with the MediGrid GWES service. Non-file inputs can be defined and passed to the engine client in the workflow descriptor xml based input document. Input files are not transferred; users have to make sure that they are provided where the workflow expects them. The following engines are registered:

engGemlcaId: GWES-medigrid-GIB, deployed by Noam Weingarten
(weingan@westminster.ac.uk)

command line arguments the engine expects:

```
>gwes-0.1-medigrid.izbi.sh -w workflow.gwdl
```

workflow descriptor and moteur service url: they are both automatically passed to the engine client by GEMLCA, only the input document has to be defined by the workflow developer

files the engine expects and generates: the engine client expects a single input file, an input document xml, that describes all non-file inputs. The user should provide the input files where the workflow expects it and should be able to get the output files manually.

Example workflow implementations: to be added

For this engine no input or output files can be defined. Standard output and error will be automatically generated as workflow outputs.

GIB-Python2.4

The GIB-Python2.4 is installed at the Westminster Cluster and it executes its jobs locally. All inputs and outputs are files and filenames do not need to be specified. Input and output files are not expected to be represented on the commandline when the engine is invoked, but they are transferred to/from the execution directory.

engGemlcaId: GIB-Python2.4, deployed by Noam Weingarten
(weingan@westminster.ac.uk)

command line arguments the engine expects:

```
>python.sh -w workflow.py
```

workflow descriptor: it is automatically passed to the engine by GEMLCA, only input and output files have to be defined by the workflow developer

files the engine expects and generates: all inputs and outputs are to be defined in the Execution, at the stage of the workflow's deployment to GEMLCA.

Example workflow implementations: to be added

For this engine it is recommended to define all input files it needs and all output files it generates for the abstract workflow. In addition, at least one example dataset providing files for all inputs and outputs is also needed. If this is provided, then all parameters can be defined according to the following table:

Parameter attribute	Inputs	Outputs
Type	INPUT_PORT or DEPENDENCY	OUTPUT_PORT
Command Line	Always false	Always false
Switch Name	Always empty string	Always empty string

If not all expected input and output files are defined or dataset is not provided for the abstract workflow, then CUSTOM parameters can be used for the workflow implementation according to the following table:

Parameter attribute	Inputs	Outputs
Type	CUSTOM	CUSTOM
Default value	String (a file name)	String (a file name)
File	Always true	Always true
Input	Always true	Always false
Command Line	Always false	Always false
Switch Name	Always empty string	Always empty string

GIB-AskalonEngine

The GIB-AskalonEngine service is installed remotely at `karwendel.dps.uibk.ac.at`. GEMLCA invokes an Askalon client that communicates with the given Askalon service.. Input files are not transferred; users have to make sure that they are provided where the workflow expects them. The following engines are registered:

engGemlcaId: GIB-AskalonEngine, deployed by Tamas Kukla
(kuklat1@staff.westminster.ac.uk)

command line arguments the engine client expects:

```
>moteur-gLite.sh wf_desc.xml
```

workflow descriptor: it is automatically passed to the engine client by GEMLCA.

files the engine expects and generates: If the workflow expects/generates further input/output files on remote locations, the user has to ensure that the inputs are provided and should be able to get the output files manually. The engine GEMLCA retrieves the STDOUT and STDERR as text files.

Example workflow implementations: to be added

For this engine inputs and outputs should not be defined. The workflow, STDOUT, STDERR are passed automatically.