SCI-BUS gateways for grid and cloud infrastructures

Tamas Kiss
University of Westminster

Peter Kacsuk,
Zoltan Farkas
MTA SZTAKI

VERCE project meeting
1st February 2013, Edinburgh

SCI-BUS is supported by the FP7 Capacities Programme under contract nr RI-283481

Motivations

- There are many user communities who would like to access distributed computing infrastructures (DCIs - grids, clouds etc.) in a transparent way
- They do not want to learn the peculiar features of the used DCIs
- They want to concentrate on their scientific application and run it in the DCI in a transparent way
- Therefore they need a science gateway

How to build a science gateway?

Option 1: Build from scratch

- If the gateway is not extremely simple, it requires long time to develop a robust gateway
- Requires substantial manpower and development cost
- It is very specialized and as users start to use it and come up with new requirements it may be difficult to extend in a scalable way
- It typically represents an isolated development without belonging to an open source community

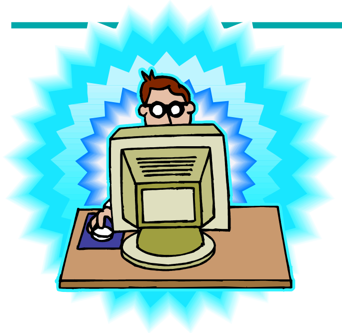
How to build a science gateway?

Option 2: Adapt and customize an existing gateway framework

- Significantly reduces development time
- Requires limited manpower and development cost
- Produces a robust and usable service
- The open source community is driving force for further development and extensions

SCI-BUS provides the required core gateway and customization technology

Who are the members of an e-science community regarding Option 2?



Science Gateway (SG) Framework Developers (5-10)

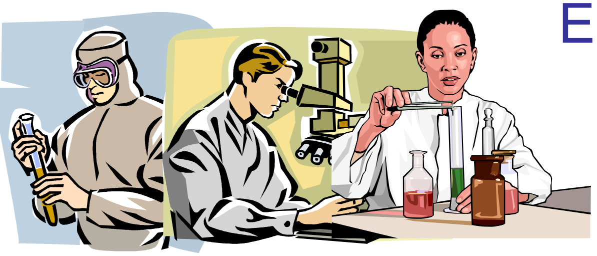
- Develop **generic** SG framework
- SCI-BUS project

SG Instance Developers (50-100)

- Develop **application domain specific** SG instance
- SCI-BUS project

WF Application Developers (500-1.000)

- Develop WF applications
- Publish the completed WF applications for end-users
- SHIWA project



End-users (e-scientists) (50.000-1.000.000)

- Execute the published WF applications with custom input parameters by creating application instances using the published WF applications as templates

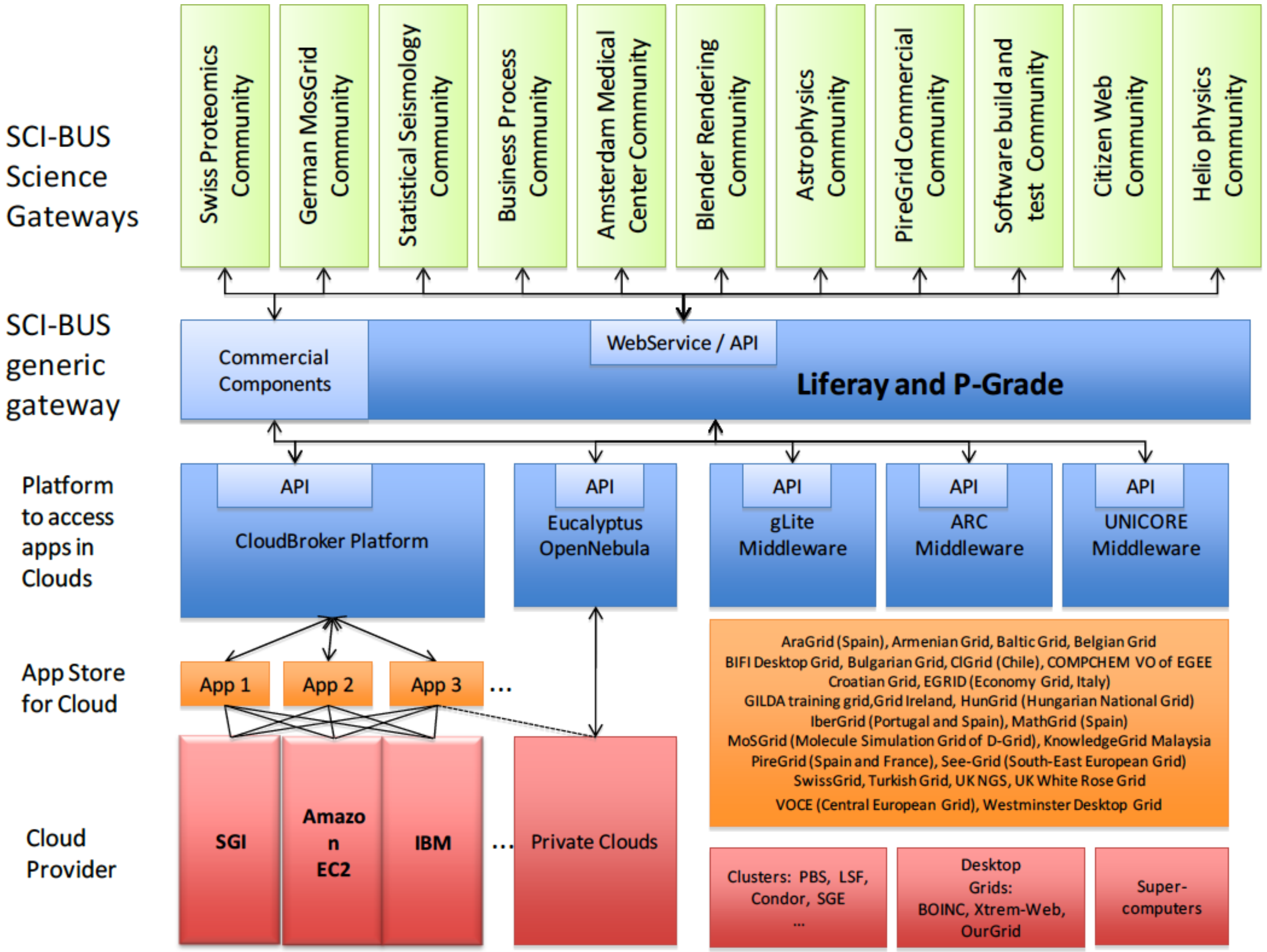
Criteria of selecting an existing gateway framework

1. Robustness
2. Sustainability
   - Will it exist in 3 years time?
   - How big and trustable the community who develops it?
3. Functionalities
4. How easy to adapt for the needs of the new user community?
5. Scalability
6. Extendibility

SCI-BUS EU FP7 project

- SCI-BUS (SCIence gateway Based User Support) provides gateways for both types of user communities
- 3-year project: 1 Oct 2011 – 30 Sep 2014
- Objectives of SCI-BUS
  - Support both WF developers and end-user scientists
  - Create a **generic-purpose science gateway framework**
  - Elaborate a **science gateway instance development technology**
  - Establish **production SG instance services** both for national grids (horizontal user communities) and various science communities (vertical user communities)
  - Develop **business models** to guarantee sustainability and commercial exploitation

SCI-BUS Architecture



SCI-BUS Science Gateways

Swiss Proteomics Community

German MosGrid Community

Statistical Seismology Community

Business Process Community

Amsterdam Medical Center Community

Blender Rendering Community

Astrophysics Community

PireGrid Commercial Community

Software build and test Community

Citizen Web Community

Helio physics Community

SCI-BUS generic gateway

Commercial Components

WebService / API

Liferay and P-Grade

Platform to access apps in Clouds

API

CloudBroker Platform

API

Eucalyptus OpenNebula

API

gLite Middleware

API

ARC Middleware

API

UNICORE Middleware

App Store for Cloud

App 1

App 2

App 3

…

AraGrid (Spain), Armenian Grid, Baltic Grid, Belgian Grid
BIFI Desktop Grid, Bulgarian Grid, ClGrid (Chile), COMPCHEM VO of EGEE
Croatian Grid, EGRID (Economy Grid, Italy)
GILDA training grid,Grid Ireland, HunGrid (Hungarian National Grid)
IberGrid (Portugal and Spain), MathGrid (Spain)
MoSGrid (Molecule Simulation Grid of D-Grid), KnowledgeGrid Malaysia
PireGrid (Spain and France), See-Grid (South-East European Grid)
SwissGrid, Turkish Grid, UK NGS, UK White Rose Grid
VOCE (Central European Grid), Westminster Desktop Grid

Cloud Provider

SGI

Amazo
n
EC2

IBM

…

Private Clouds

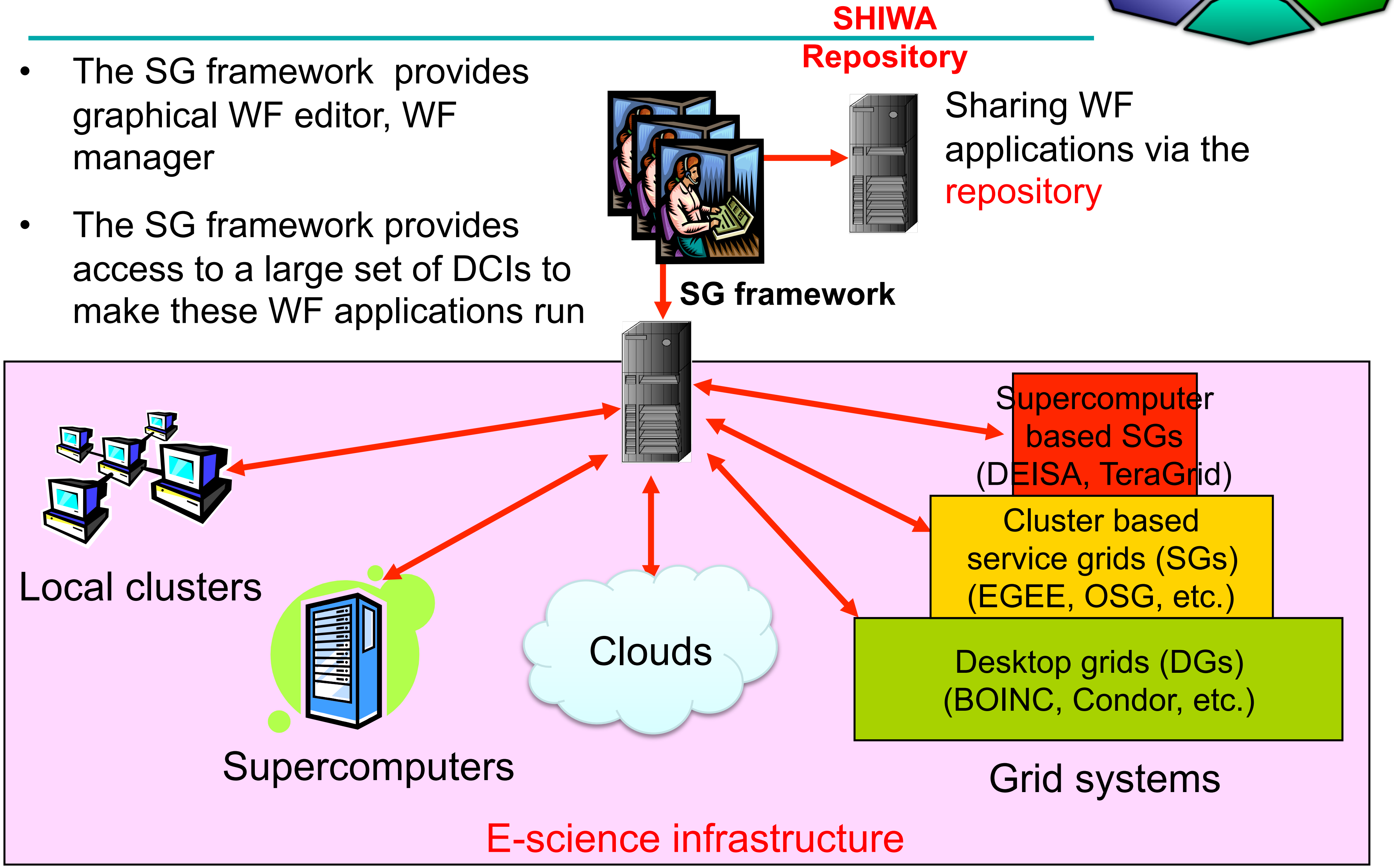
Clusters: PBS, LSF, Condor, SGE …

Desktop Grids: BOINC, Xtrem-Web, OurGrid

Super-computers

Support for workflow developers
(in collaboration with SHIWA and ER-flow)

- The SG framework provides graphical WF editor, WF manager
- The SG framework provides access to a large set of DCIs to make these WF applications run



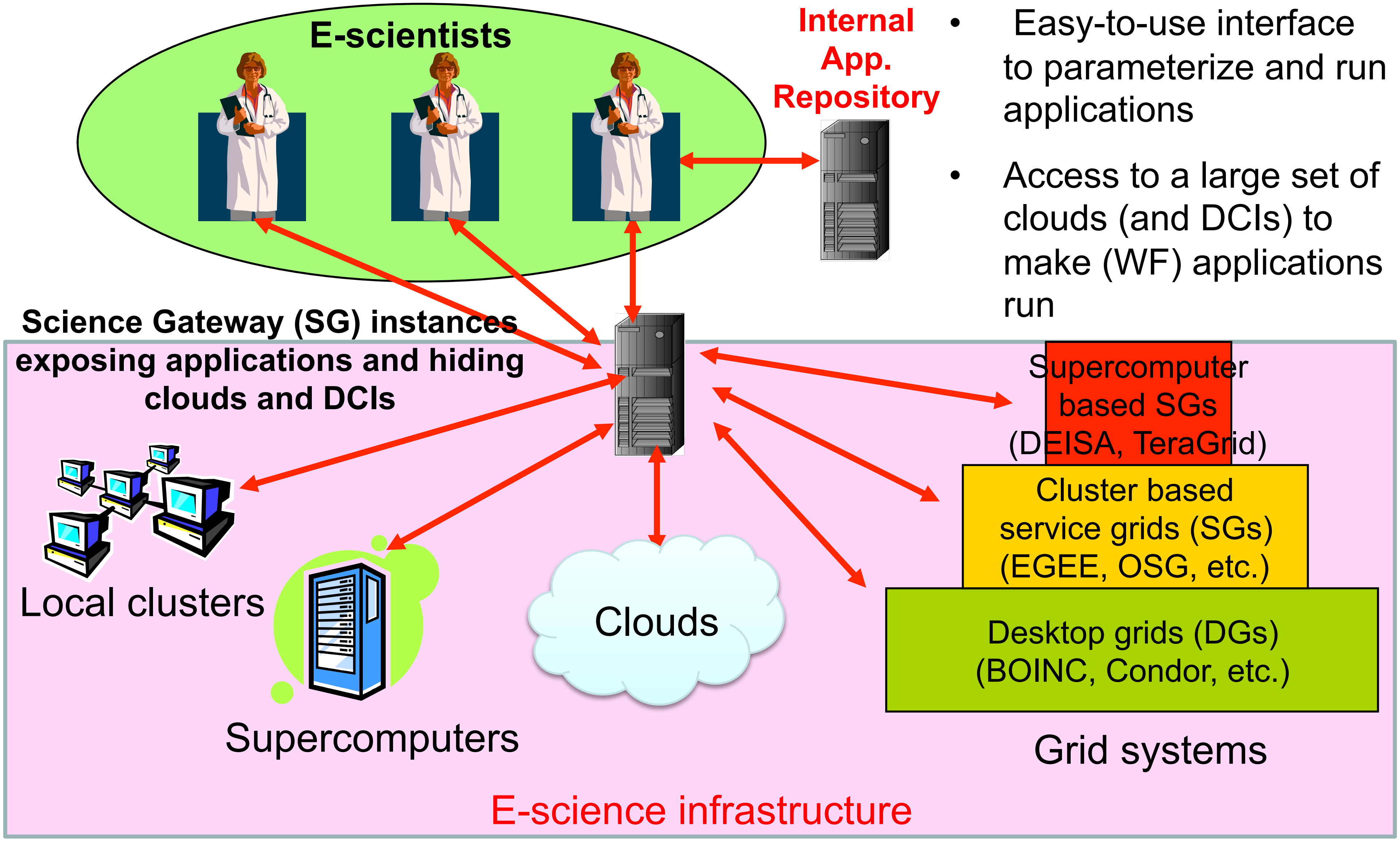
SHIWA Repository

Sharing WF applications via the repository

SG framework

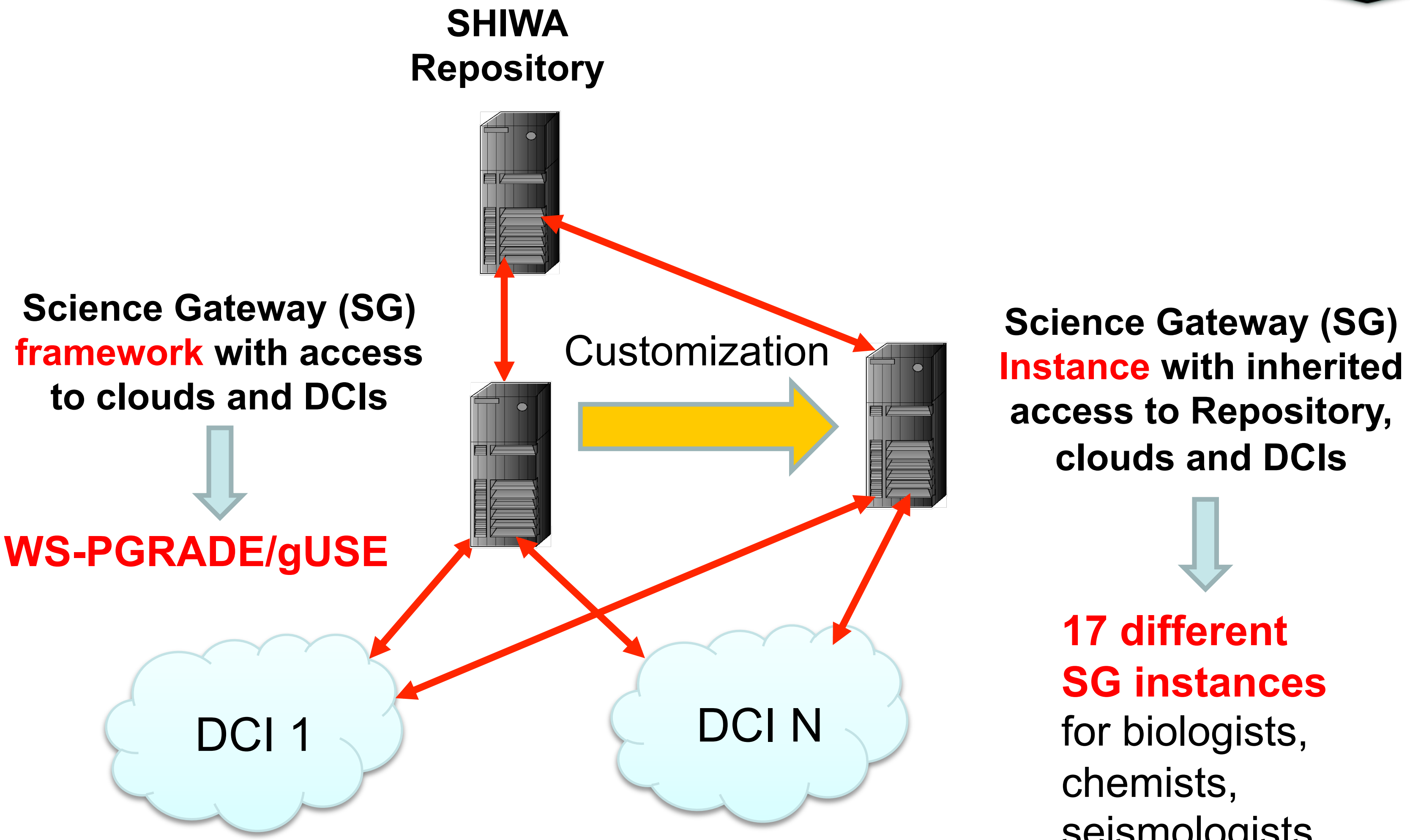
Local clusters
Supercomputers
Clouds
Supercomputer based SGs (DEISA, TeraGrid)
Cluster based service grids (SGs) (EGEE, OSG, etc.)
Desktop grids (DGs) (BOINC, Condor, etc.)
Grid systems
E-science infrastructure

Support for end-user e-scientists


E-scientists
Internal App. Repository
Science Gateway (SG) instances exposing applications and hiding clouds and DCIs
Supercomputer based SGs (DEISA, TeraGrid)
Cluster based service grids (SGs) (EGEE, OSG, etc.)
Desktop grids (DGs) (BOINC, Condor, etc.)
Grid systems
Local clusters
Supercomputers
Clouds
E-science infrastructure

- Easy-to-use interface to parameterize and run applications
- Access to a large set of clouds (and DCIs) to make (WF) applications run

Support for science gateway instance developers



Science Gateway (SG) **framework** with access to clouds and DCIs

WS-PGRADE/gUSE

Science Gateway (SG) **Instance** with inherited access to Repository, clouds and DCIs

17 different SG instances
for biologists,
chemists,
seismologists
etc.

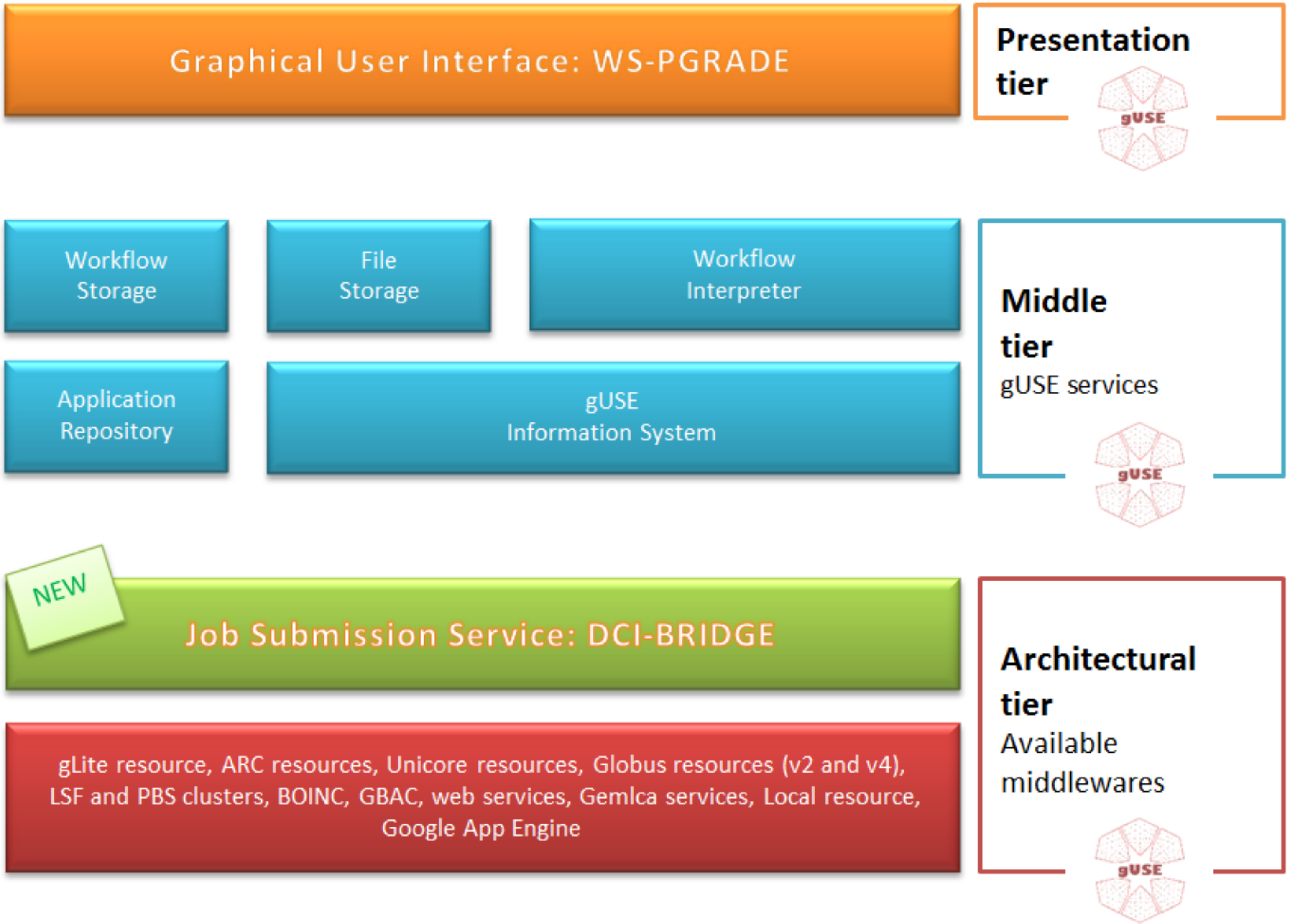
WS-PGRADE/gUSE

Generic-purpose gateway framework

- Based on Liferay
- General purpose
- Workflow-oriented portal framework
- Supports the development and execution of workflow-based applications
- Enables the multi-cloud and multi-DCI execution of any WF
- Supports the fast development of SG instances by a customization technology
- Provides access to
  - internal repository
  - external SHIWA Repository

gUSE – grid User Support Environment

Scalable architecture based on collaborating services



Graphical User Interface: WS-PGRADE

Presentation tier

Workflow Storage

File Storage

Workflow Interpreter

Application Repository

gUSE Information System

Middle tier

gUSE services

NEW

Job Submission Service: DCI-BRIDGE

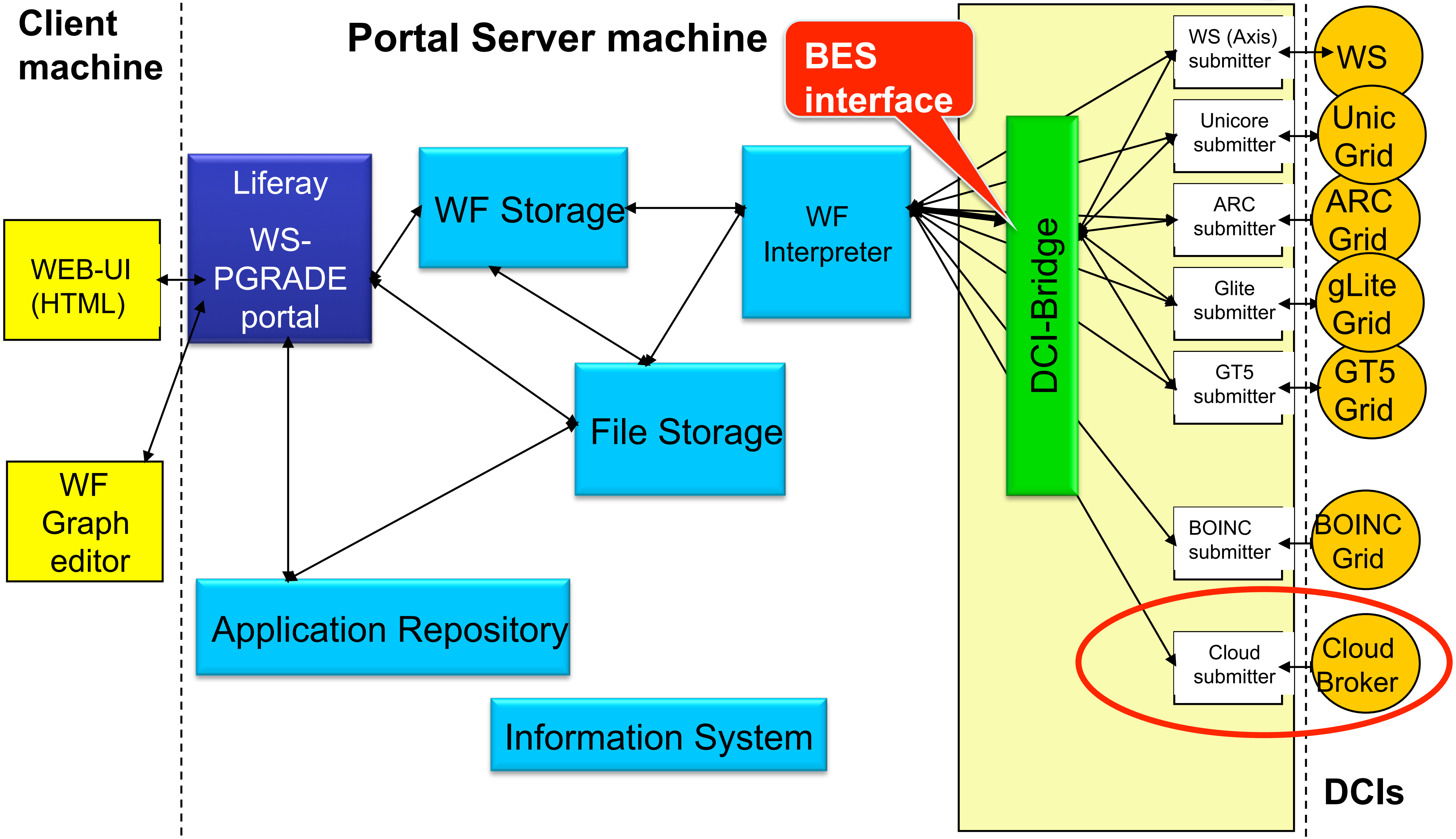
gLite resource, ARC resources, Unicore resources, Globus resources (v2 and v4), LSF and PBS clusters, BOINC, GBAC, web services, Gemlca services, Local resource, Google App Engine

Architectural tier

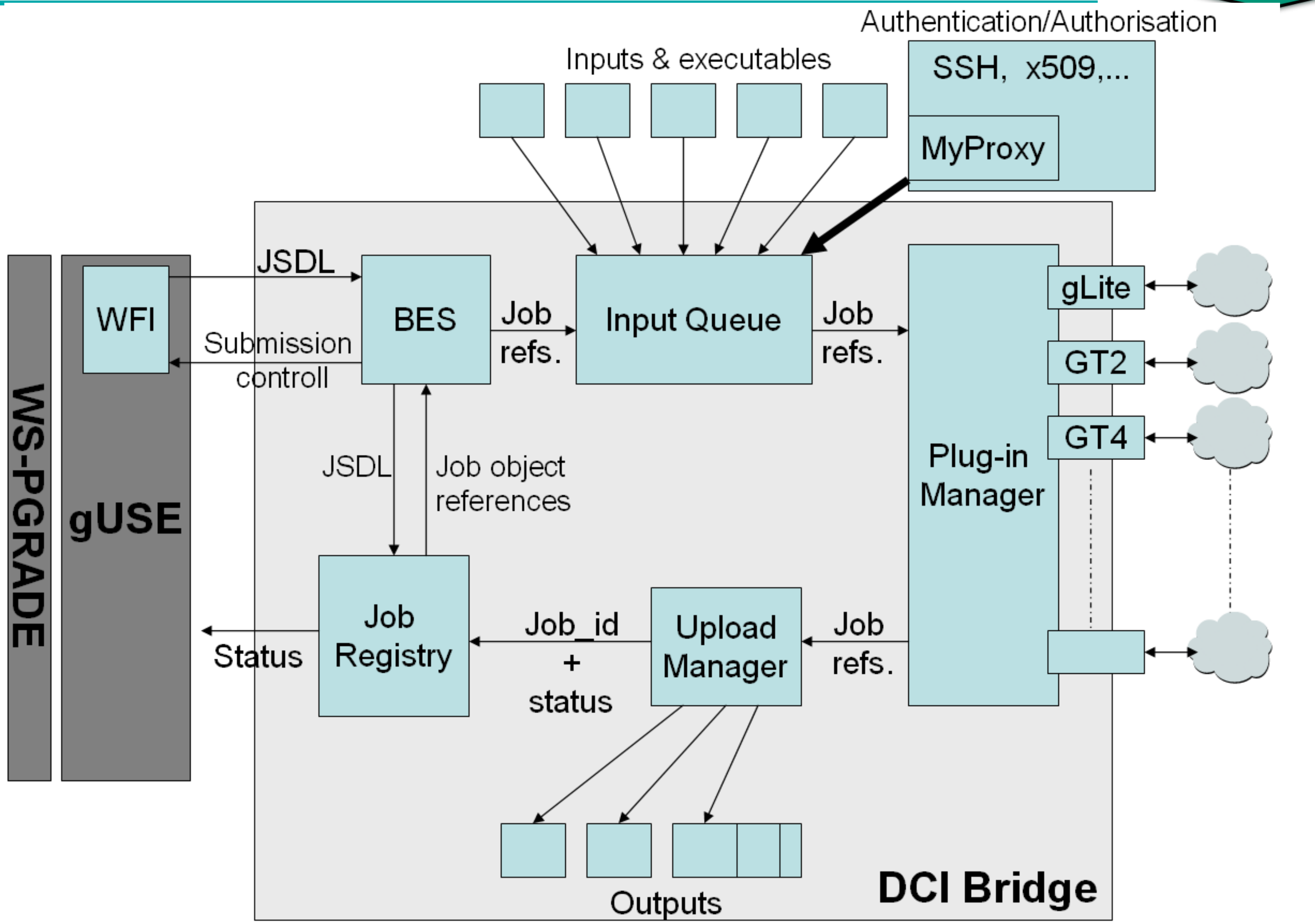
Available middlewares

gUSE – grid User Support Environment

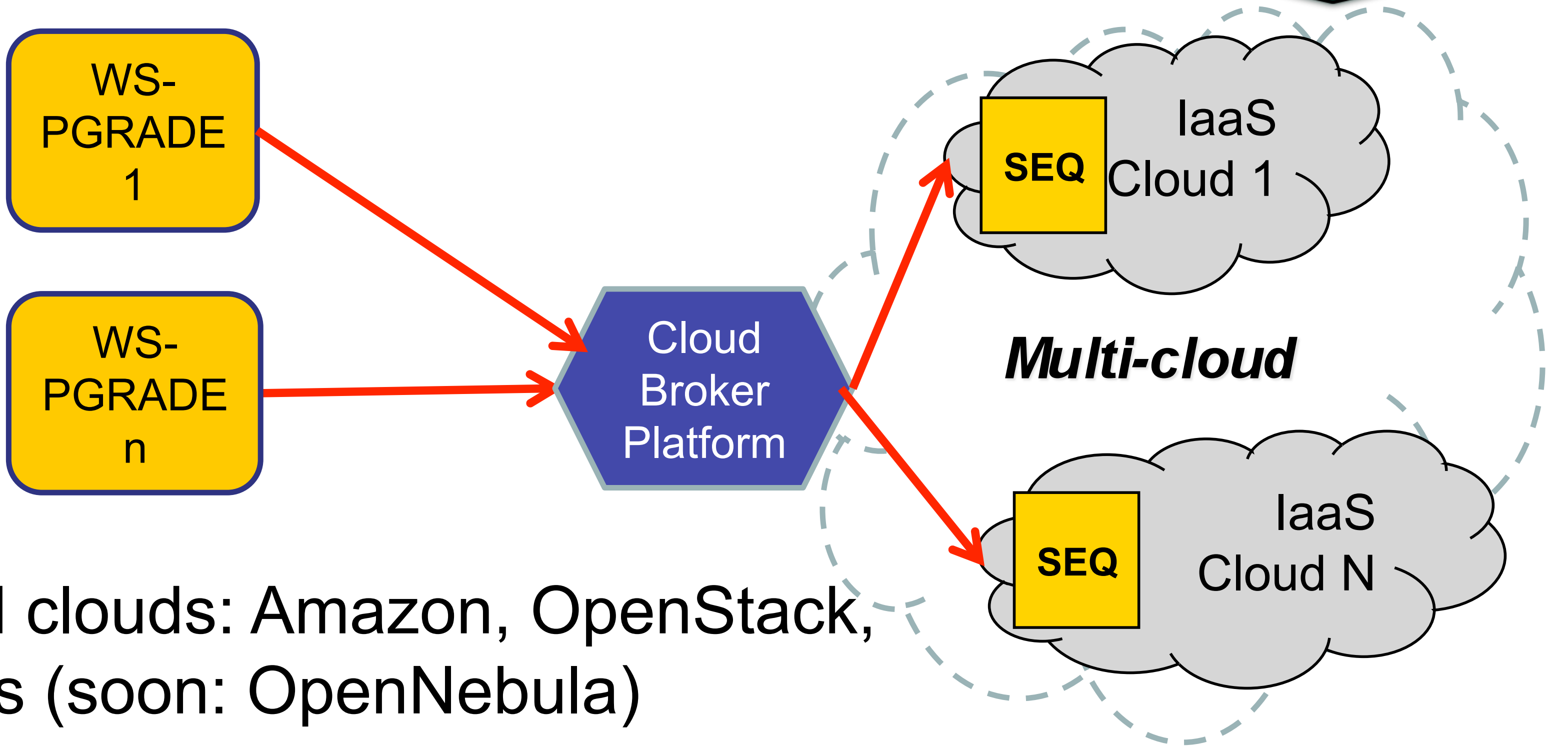
Seamless access to various DCIs



DCI Bridge


Authentication/Authorisation
Inputs & executables
SSH, x509,...
MyProxy
JSDL
WFI
Submission controll
BES
Job refs.
Input Queue
Job refs.
Plug-in Manager
gLite
GT2
GT4
WS-PGRADE
gUSE
JSDL
Job object references
Status
Job Registry
Job_id + status
Upload Manager
Job refs.
Outputs
DCI Bridge

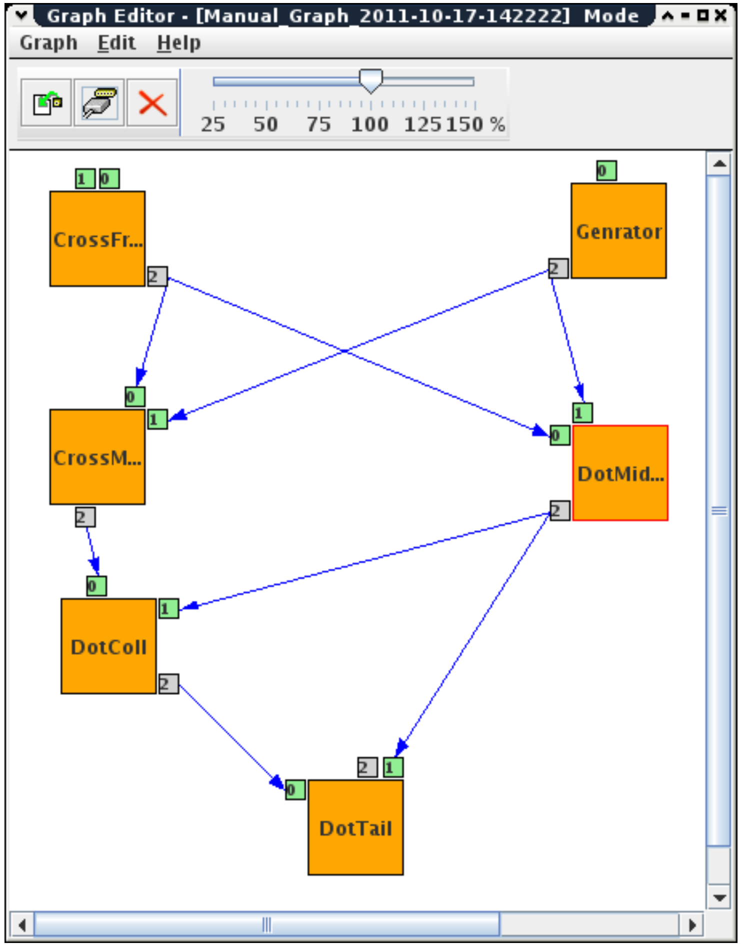
Integrated WS-PGRADE/CloudBroker Platform to access multi-clouds



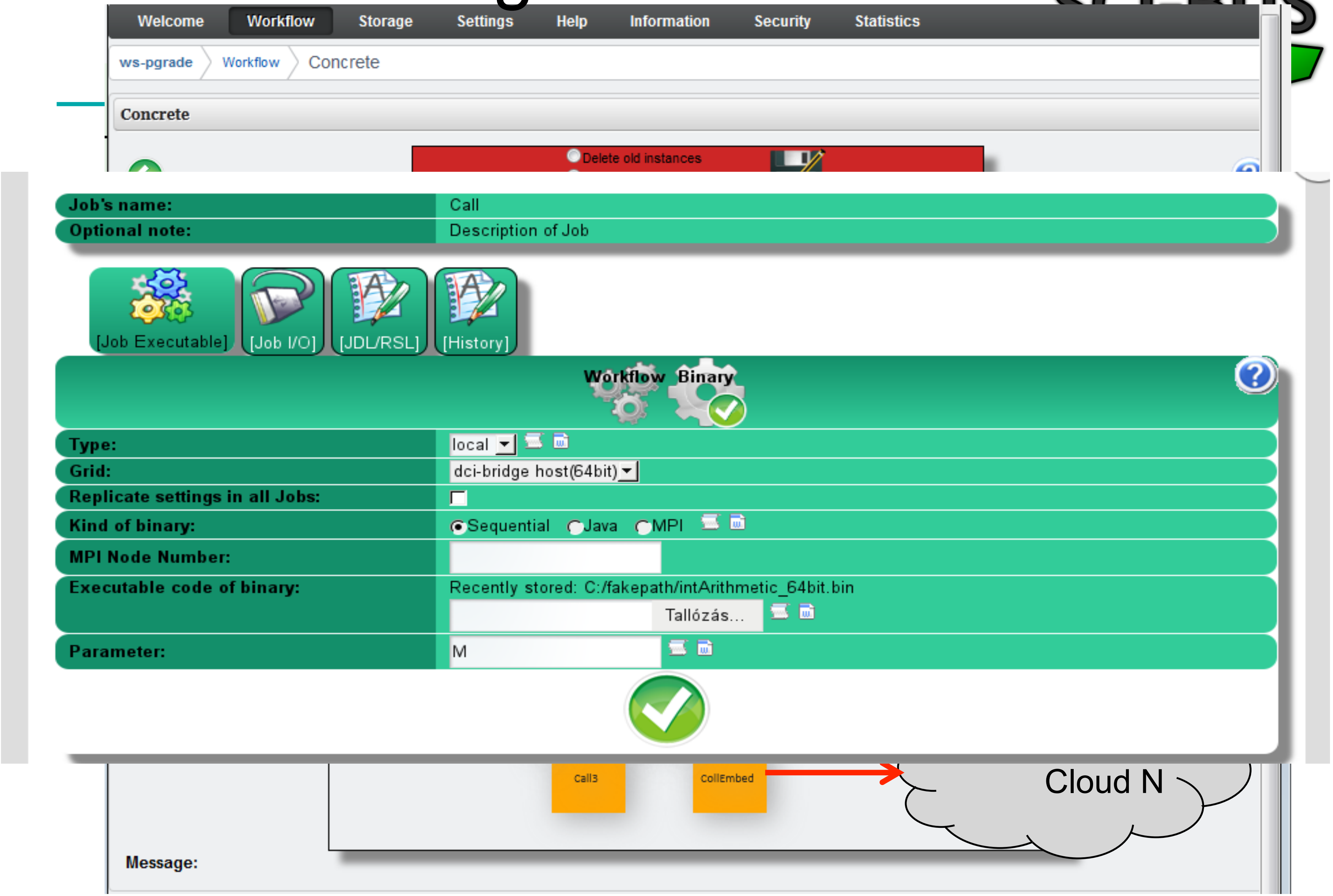
- Supported clouds: Amazon, OpenStack, Eucalyptus (soon: OpenNebula)
- **SaaS solution**:
  - Preregistered services/jobs can run from WS-PGRADE Supported from gUSE 3.5.0
- **IaaS solution**:
  - any services/jobs (e.g. BoT jobs) can run from WS-PGRADE Supported from gUSE 3.5.1

Creating workflow graphs

- Create WF structure
- Add new jobs
- Add ports to jobs
- Define connections between jobs

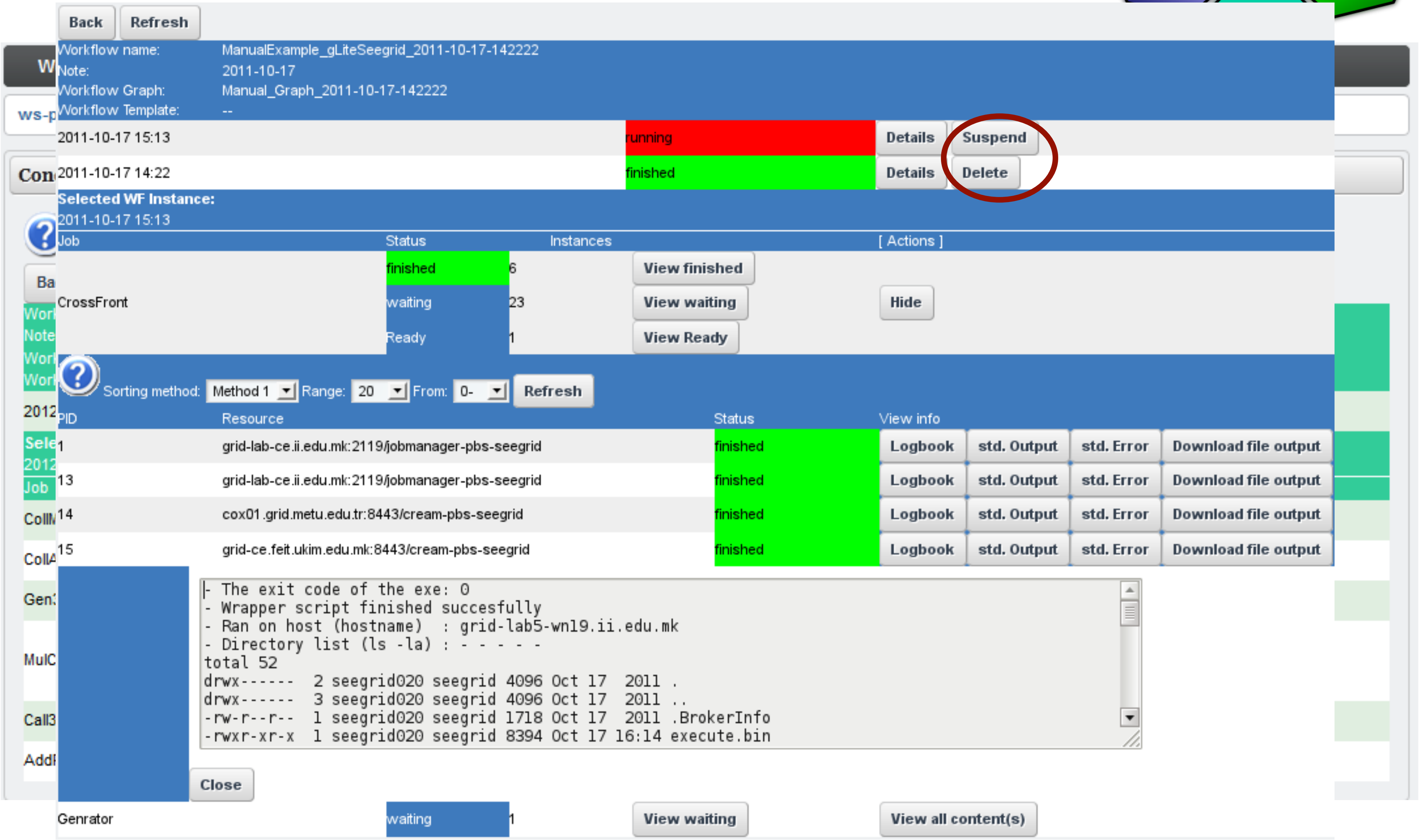


Workflow configuration



Job's name:	Call
Optional note:	Description of Job
Type:	local
Grid:	dci-bridge host(64bit)
Replicate settings in all Jobs:	□
Kind of binary:	● Sequential ○ Java ○ MPI
MPI Node Number:	
Executable code of binary:	Recently stored: C:/fakepath/intArithmetic_64bit.bin
Parameter:	M

Running and monitoring workflow instances



Workflow name: ManualExample_gLiteSeegrid_2011-10-17-142222
Note: 2011-10-17
Workflow Graph: Manual_Graph_2011-10-17-142222
Workflow Template: --

2011-10-17 15:13	running	Details	Suspend
2011-10-17 14:22	finished	Details	Delete

Selected WF Instance:
2011-10-17 15:13

Job	Status	Instances		[Actions]
CrossFront	finished	6	View finished	Hide
	waiting	23	View waiting	
	Ready	1	View Ready	

Sorting method: Method 1 Range: 20 From: 0- Refresh

PID	Resource	Status	View info			
1	grid-lab-ce.ii.edu.mk:2119/jobmanager-pbs-seegrid	finished	Logbook	std. Output	std. Error	Download file output
13	grid-lab-ce.ii.edu.mk:2119/jobmanager-pbs-seegrid	finished	Logbook	std. Output	std. Error	Download file output
14	cox01.grid.metu.edu.tr:8443/cream-pbs-seegrid	finished	Logbook	std. Output	std. Error	Download file output
15	grid-ce.feit.ukim.edu.mk:8443/cream-pbs-seegrid	finished	Logbook	std. Output	std. Error	Download file output

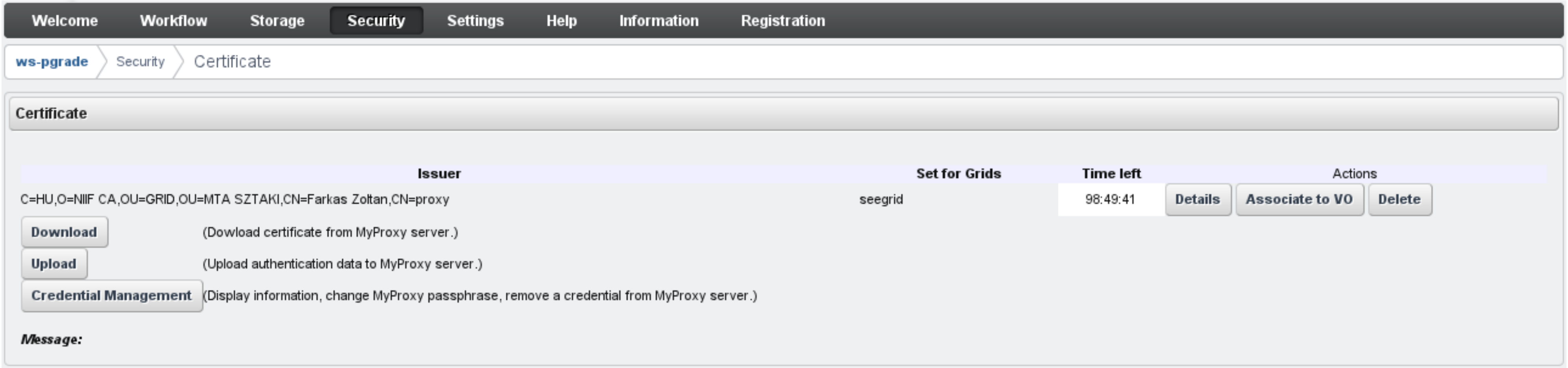
```
- The exit code of the exe: 0
- Wrapper script finished succesfully
- Ran on host (hostname)  : grid-lab5-wn19.ii.edu.mk
- Directory list (ls -la) : - - - - -
total 52
drwx------  2 seegrid020 seegrid 4096 Oct 17  2011 .
drwx------  3 seegrid020 seegrid 4096 Oct 17  2011 ..
-rw-r--r--  1 seegrid020 seegrid 1718 Oct 17  2011 .BrokerInfo
-rwxr-xr-x  1 seegrid020 seegrid 8394 Oct 17 16:14 execute.bin
```

Genrator	waiting	1	View waiting	View all content(s)

Certificate management

Manage X.509 certificates:

- Upload to MyProxy server
- Get new proxy from a MyProxy server
- Manage proxies on MyProxy servers



Certificate

Issuer	Set for Grids	Time left	Actions
C=HU,O=NIIF CA,OU=GRID,OU=MTA SZTAKI,CN=Farkas Zoltan,CN=proxy	seegrid	98:49:41	Details Associate to VO Delete

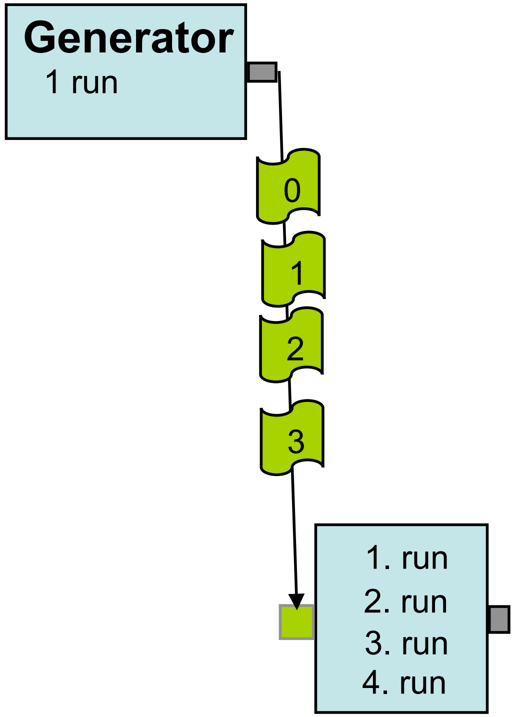
Download (Dowload certificate from MyProxy server.)

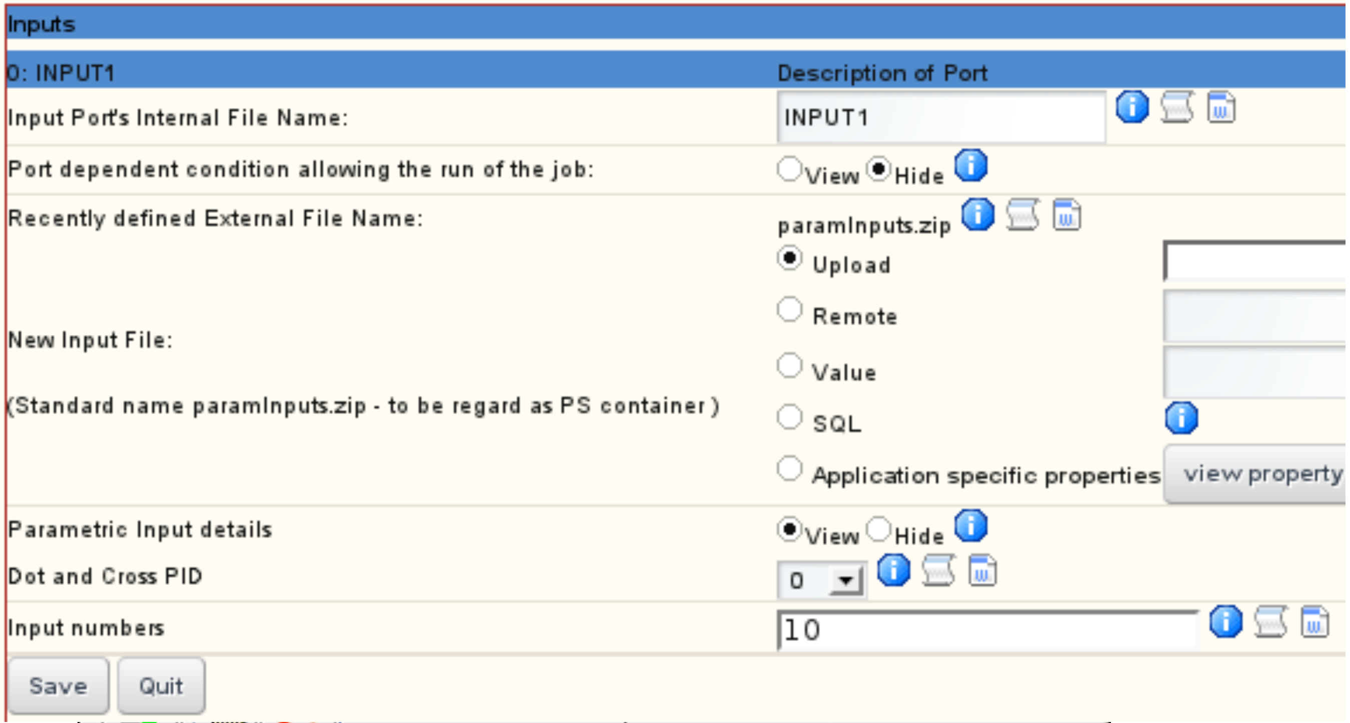
Upload (Upload authentication data to MyProxy server.)

Credential Management (Display information, change MyProxy passphrase, remove a credential from MyProxy server.)

Message:

Advanced features – Parameter sweep workflows




Inputs

0: INPUT1 — Description of Port

Input Port's Internal File Name: INPUT1

Port dependent condition allowing the run of the job: ○ View ● Hide

Recently defined External File Name: paramInputs.zip

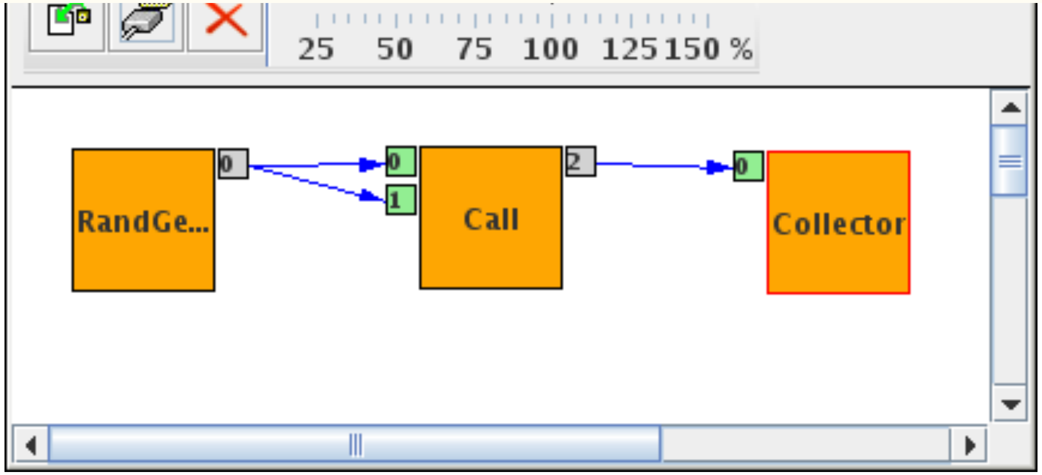
New Input File: ● Upload ○ Remote ○ Value ○ SQL ○ Application specific properties

(Standard name paramInputs.zip - to be regard as PS container)

Parametric Input details: ● View ○ Hide

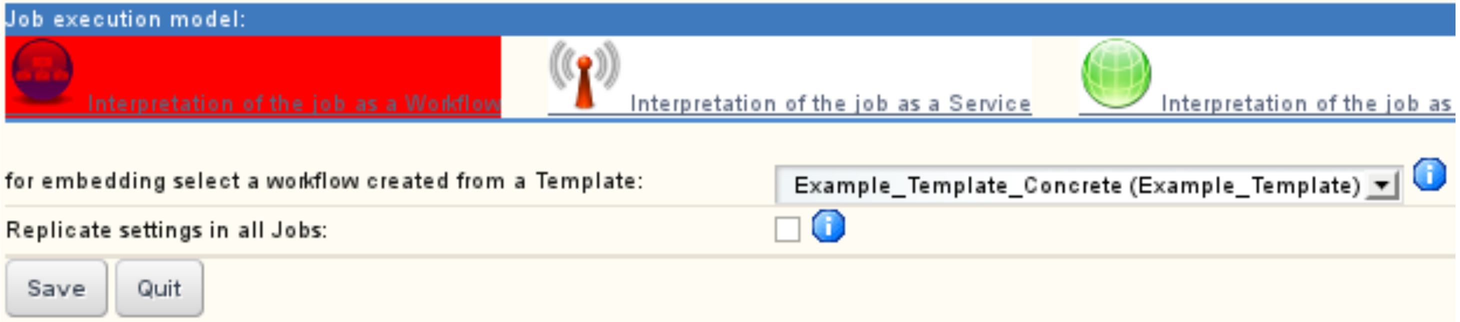
Dot and Cross PID: 0

Input numbers: 10



Advanced features – Embedded workflows

- A job can be a workflow!
- Rule: a concrete workflow created from a template may be embedded



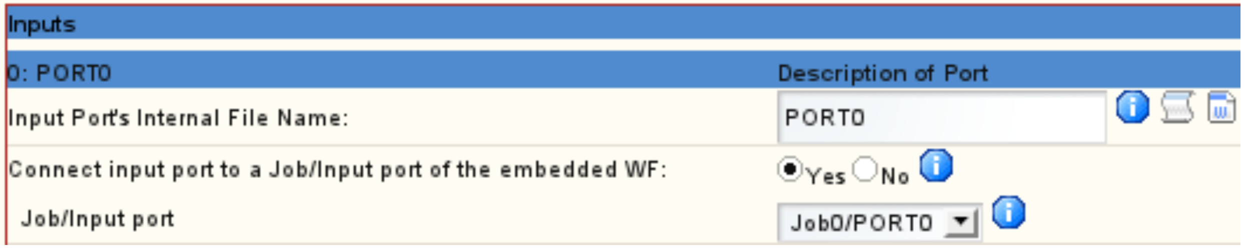
Job execution model:

Interpretation of the job as a Workflow
Interpretation of the job as a Service
Interpretation of the job as

for embedding select a workflow created from a Template: Example_Template_Concrete (Example_Template)

Replicate settings in all Jobs: □

- Port assignments have to be set:



Inputs

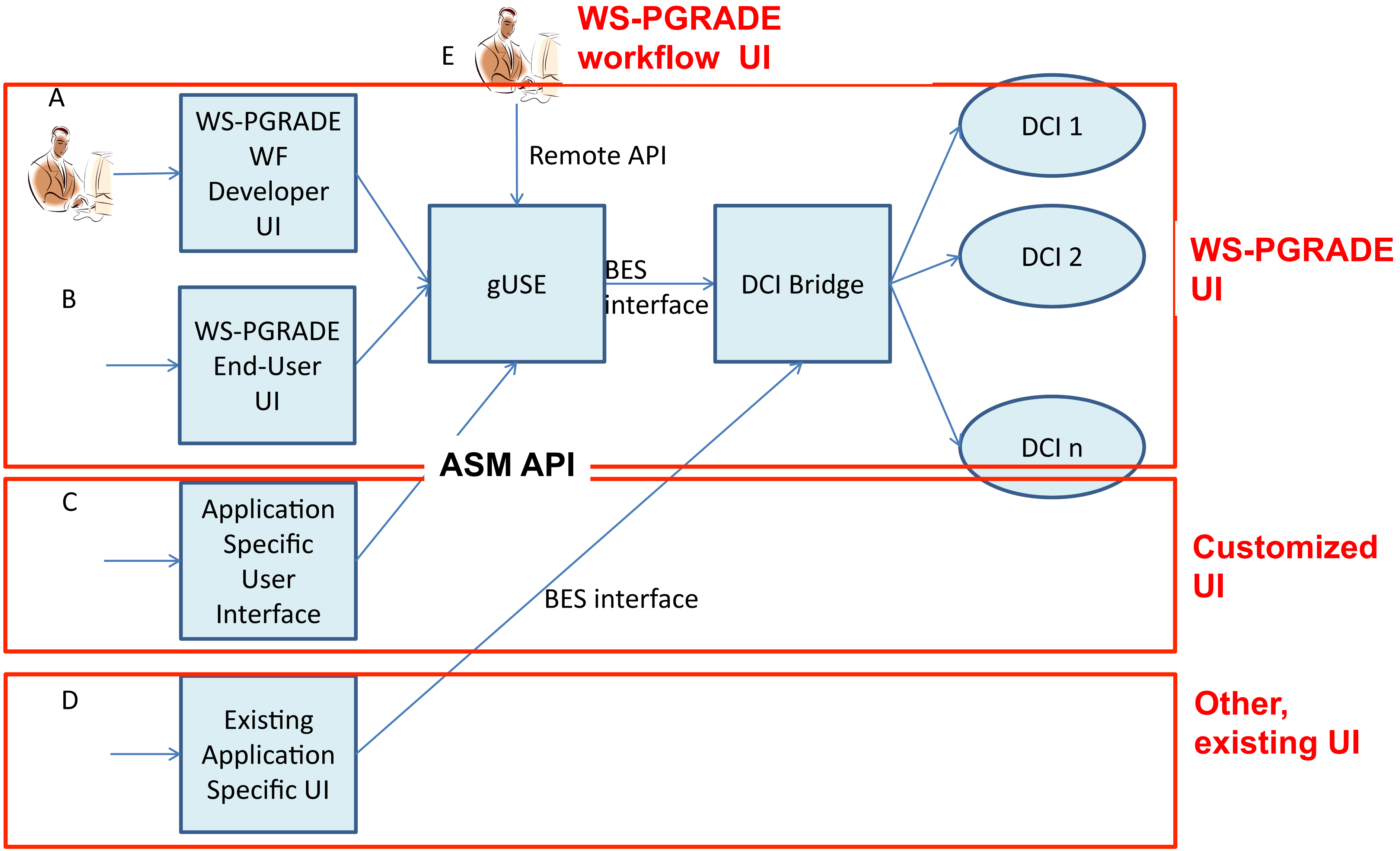
0: PORT0 Description of Port

Input Port's Internal File Name: PORT0

Connect input port to a Job/Input port of the embedded WF: ● Yes ○ No

Job/Input port: Job0/PORT0

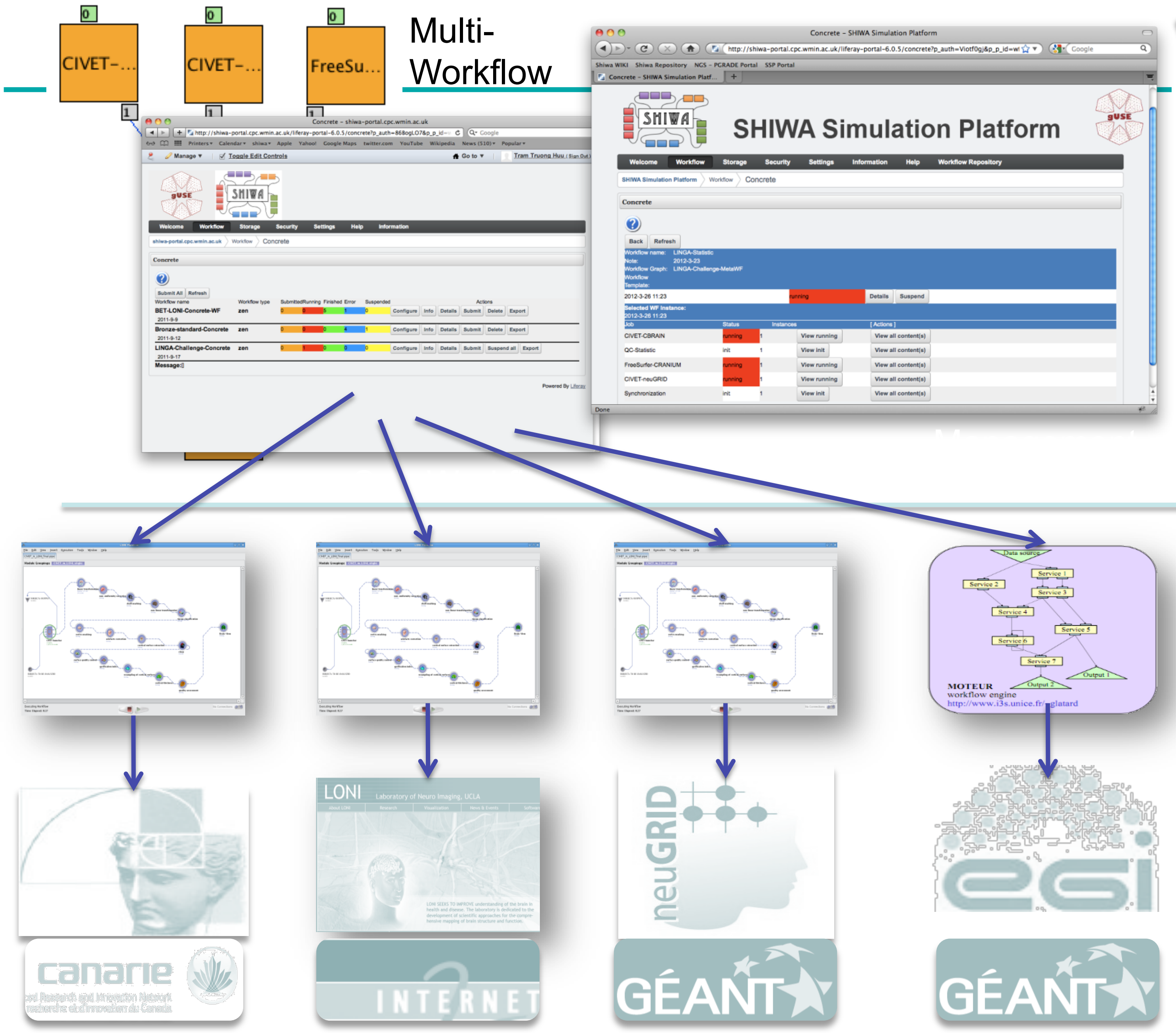
User access modes



Types of gateways to be developed from the core WS-PGRADE/gUSE framework

1. Generic purpose gateways for grids/clouds
   - Core WS-PGRADE/gUSE
2. Generic purpose gateway for specific technologies
   - SHIWA gateway for workflow sharing and interoperation
3. Application-oriented science gateway instance
   - Autodock gateway
   - Statistical seismology gateway

SHIWA solution for LINGA



Concrete – shiwa-portal.cpc.wmin.ac.uk

http://shiwa-portal.cpc.wmin.ac.uk/liferay-portal-6.0.5/concrete?p_auth=868ogLO7&p_p_id=

Manage ▾ ✓ Toggle Edit Controls Go to ▾ Tram Truong Huu (Sign Out)

Welcome Workflow Storage Security Settings Help Information

shiwa-portal.cpc.wmin.ac.uk › Workflow › Concrete

Concrete

Submit All Refresh

Workflow name	Workflow type	Submitted	Running	Finished	Error	Suspended	Actions
BET-LONI-Concrete-WF 2011-9-9	zen	0	0	5	1	0	Configure Info Details Submit Delete Export
Bronze-standard-Concrete 2011-9-12	zen	0	0	0	4	1	Configure Info Details Submit Delete Export
LINGA-Challenge-Concrete 2011-9-17	zen	0	1	0	0	0	Configure Info Details Submit Suspend all Export

Message:[]

Powered By Liferay

Concrete – SHIWA Simulation Platform

http://shiwa-portal.cpc.wmin.ac.uk/liferay-portal-6.0.5/concrete?p_auth=Viotf0gj&p_p_id=wi

Shiwa WIKI Shiwa Repository NGS – PGRADE Portal SSP Portal

SHIWA Simulation Platform

Welcome Workflow Storage Security Settings Information Help Workflow Repository

SHIWA Simulation Platform › Workflow › Concrete

Concrete

Back Refresh

Workflow name: LINGA-Statistic
Note: 2012-3-23
Workflow Graph: LINGA-Challenge-MetaWF
Workflow Template:

2012-3-26 11:23 running Details Suspend

Selected WF Instance:
2012-3-26 11:23

Job	Status	Instances		[Actions]
CIVET-CBRAIN	running	1	View running	View all content(s)
QC-Statistic	init	1	View init	View all content(s)
FreeSurfer-CRANIUM	running	1	View running	View all content(s)
CIVET-neuGRID	running	1	View running	View all content(s)
Synchronization	init	1	View init	View all content(s)

Done

Support for end-user scientists

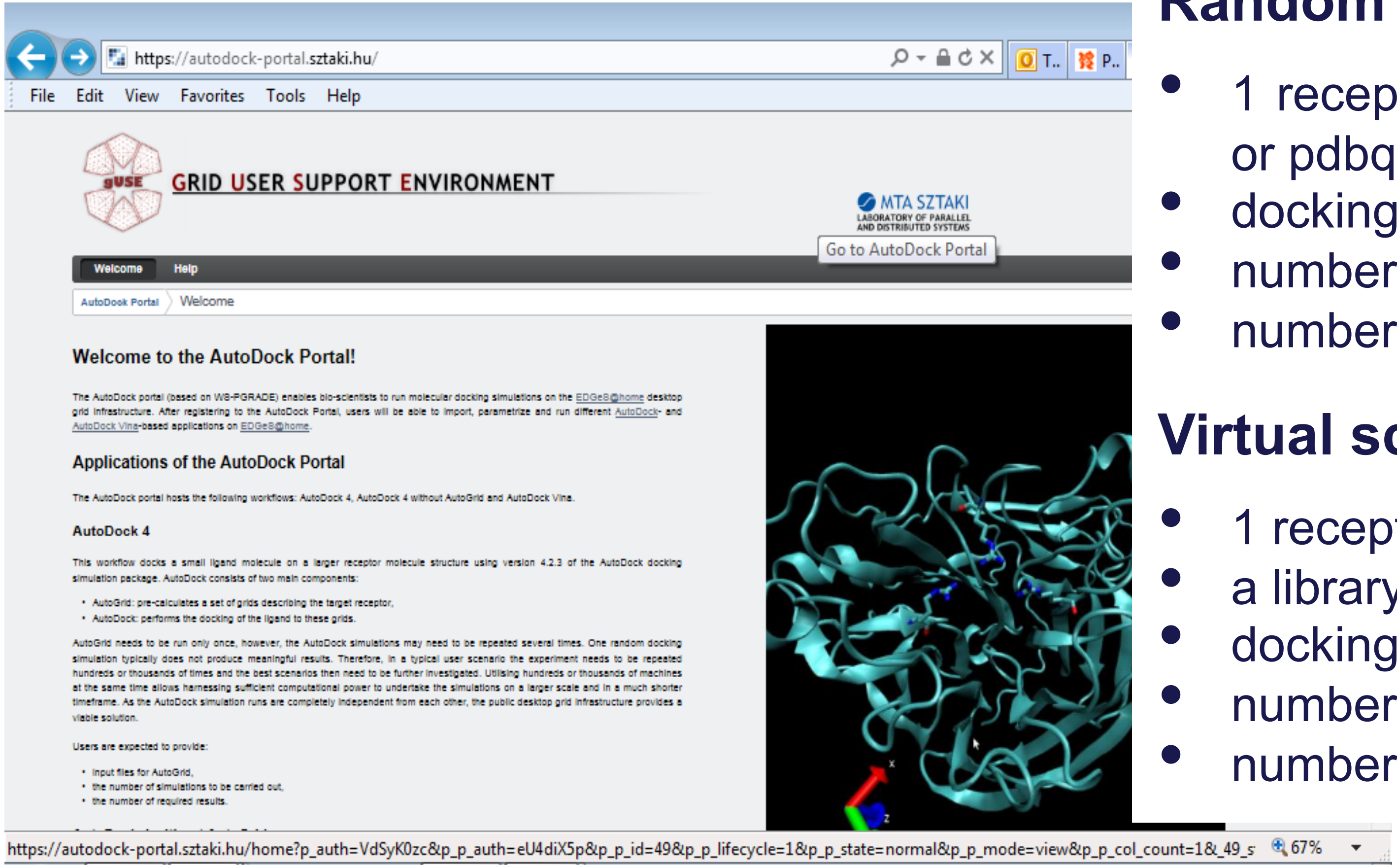
SCI-BUS provides technology to easily and rapidly create application-oriented science gateway instances

Two methods:

- End-user mode: Autodock gateway
- Customization via ASM: Statistical Seismology gateway

Example: Autodock gateway

- Free public access to pre-deployed molecular docking scenarios for bio-scientist end-users
- Workflows executed on public desktop grids
- 70+ registered users, over 1 millions jobs


https://autodock-portal.sztaki.hu/
T..
P..
File
Edit
View
Favorites
Tools
Help
gUSE
GRID USER SUPPORT ENVIRONMENT
MTA SZTAKI
LABORATORY OF PARALLEL AND DISTRIBUTED SYSTEMS
Go to AutoDock Portal
Welcome
Help
AutoDock Portal
Welcome
x
z
https://autodock-portal.sztaki.hu/home?p_auth=VdSyK0zc&p_p_auth=eU4diX5p&p_p_id=49&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&p_p_col_count=1&_49_s
67%

Welcome to the AutoDock Portal!

The AutoDock portal (based on WS-PGRADE) enables bio-scientists to run molecular docking simulations on the EDGeS@home desktop grid infrastructure. After registering to the AutoDock Portal, users will be able to import, parametrize and run different AutoDock- and AutoDock Vina-based applications on EDGeS@home.

Applications of the AutoDock Portal

The AutoDock portal hosts the following workflows: AutoDock 4, AutoDock 4 without AutoGrid and AutoDock Vina.

AutoDock 4

This workflow docks a small ligand molecule on a larger receptor molecule structure using version 4.2.3 of the AutoDock docking simulation package. AutoDock consists of two main components:

- AutoGrid: pre-calculates a set of grids describing the target receptor,
- AutoDock: performs the docking of the ligand to these grids.

AutoGrid needs to be run only once, however, the AutoDock simulations may need to be repeated several times. One random docking simulation typically does not produce meaningful results. Therefore, in a typical user scenario the experiment needs to be repeated hundreds or thousands of times and the best scenarios then need to be further investigated. Utilising hundreds or thousands of machines at the same time allows harnessing sufficient computational power to undertake the simulations on a larger scale and in a much shorter timeframe. As the AutoDock simulation runs are completely independent from each other, the public desktop grid infrastructure provides a viable solution.

Users are expected to provide:

- input files for AutoGrid,
- the number of simulations to be carried out,
- the number of required results.

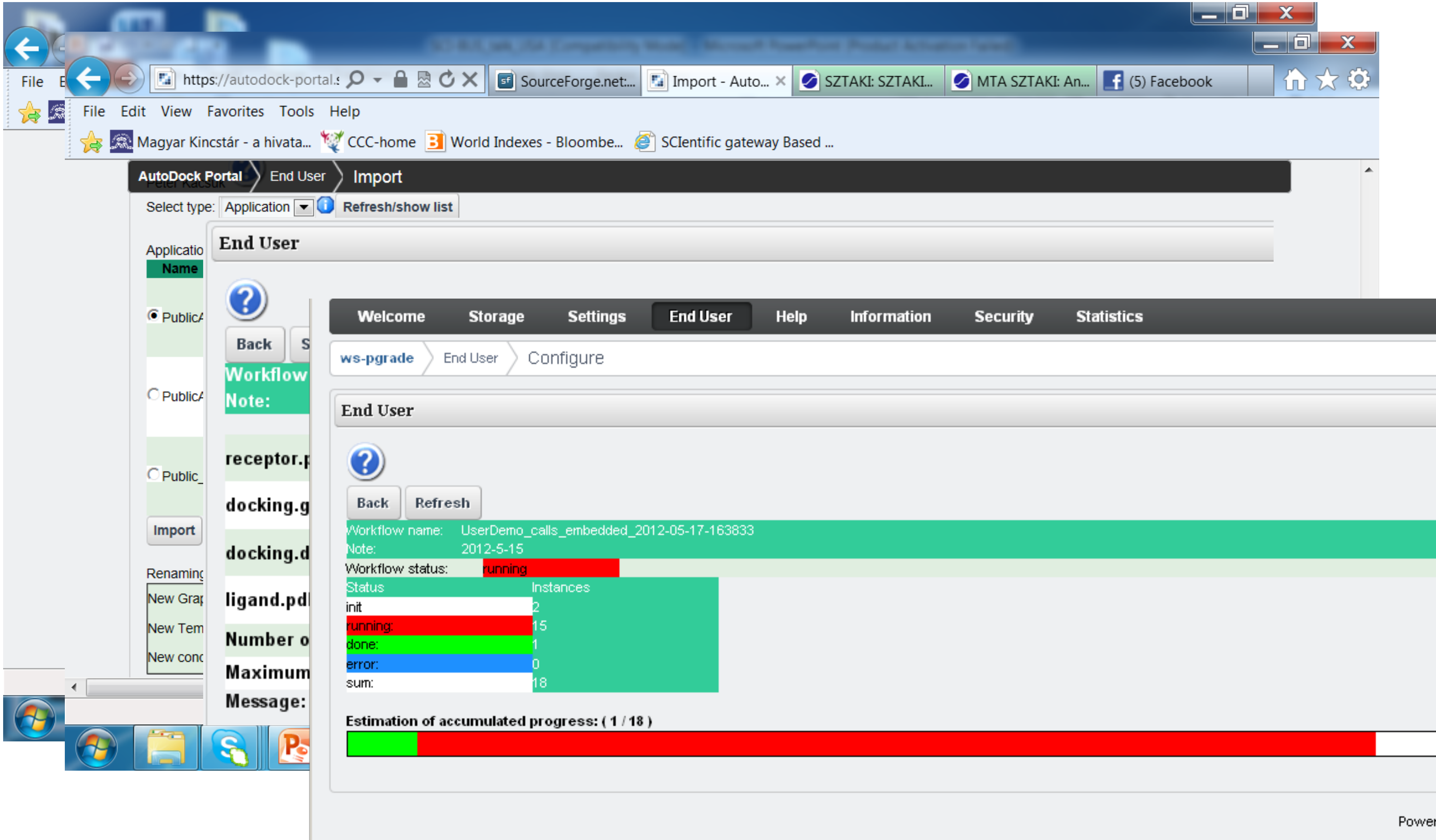
Random blind docking:

- 1 receptor and 1 ligand file (pdb or pdbqt)
- docking parameter files
- number of iterations
- number of lowest energy results

Virtual screening:

- 1 receptor file
- a library of ligands
- docking parameter files
- number of work units
- number of lowest energy results

Example: Autodock gateway



Workflow name: UserDemo_calls_embedded_2012-05-17-163833

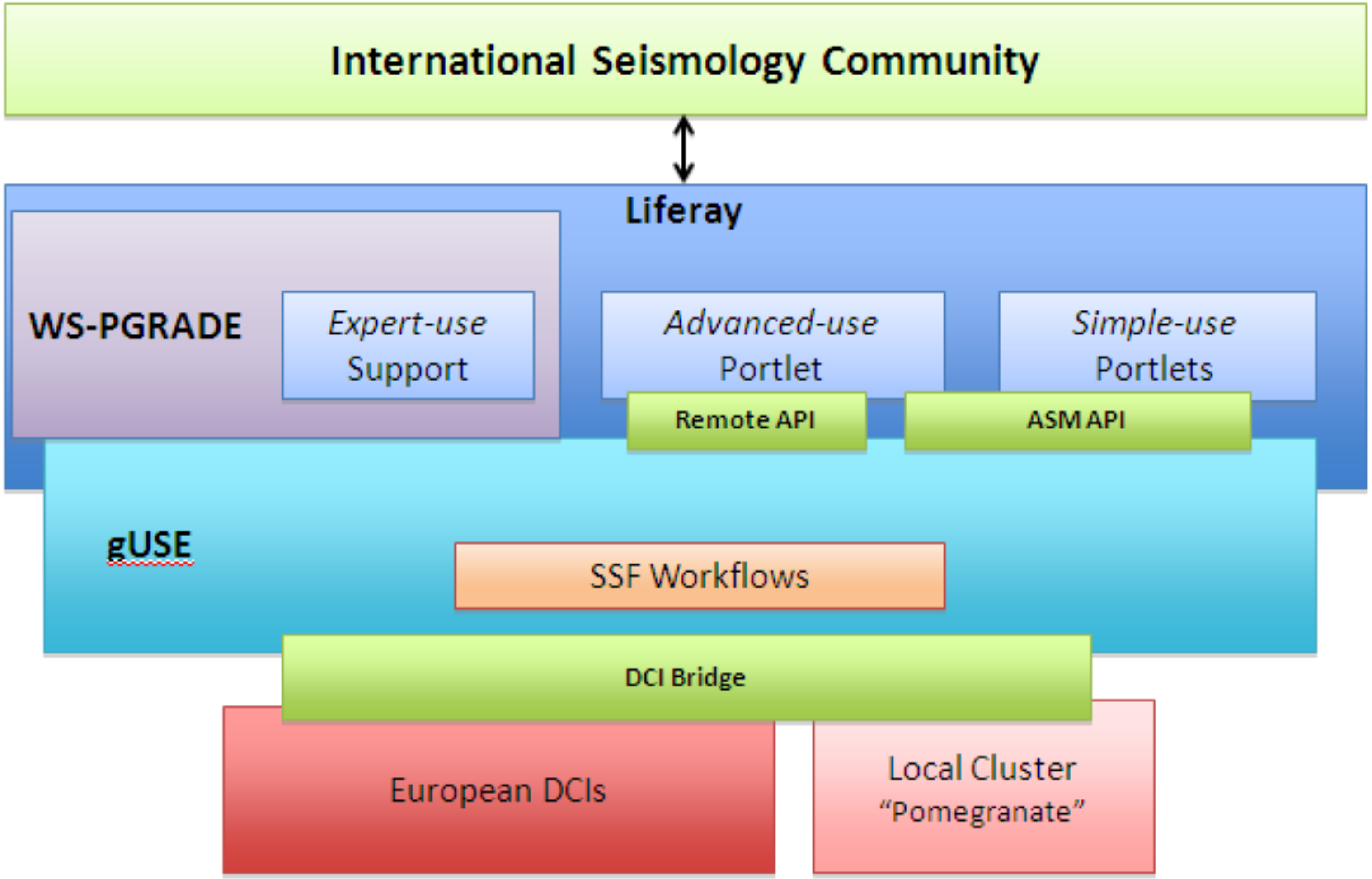
Note: 2012-5-15

Workflow status: running

Status	Instances
init	2
running:	15
done:	1
error:	0
sum:	18

Estimation of accumulated progress: (1 / 18)

Seismology Gateway Architecture



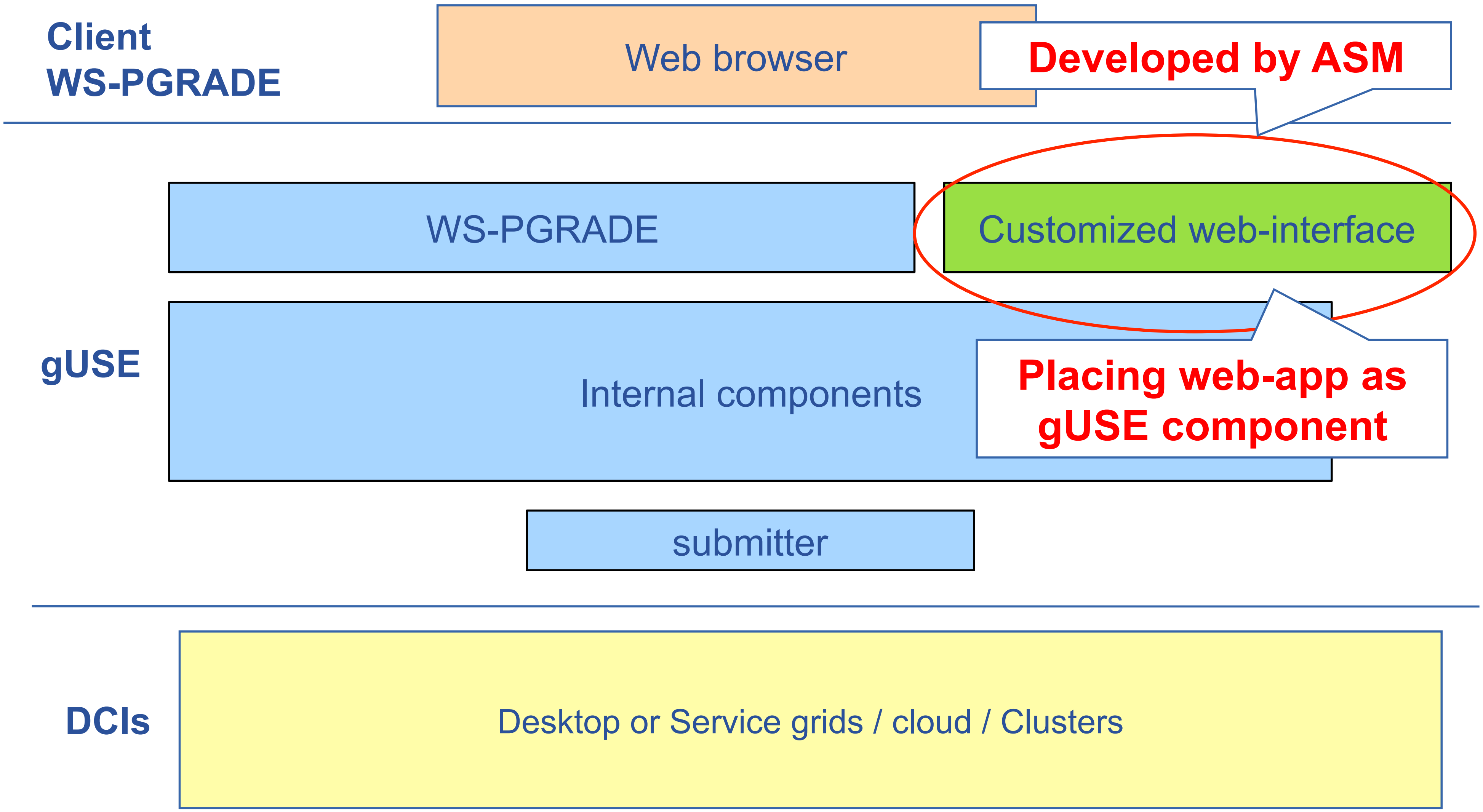
Statistical Seismology demo

https://www.sci-bus.eu/demos7

For further details please contact:

Cevat Sener - sener@ceng.metu.edu.tr

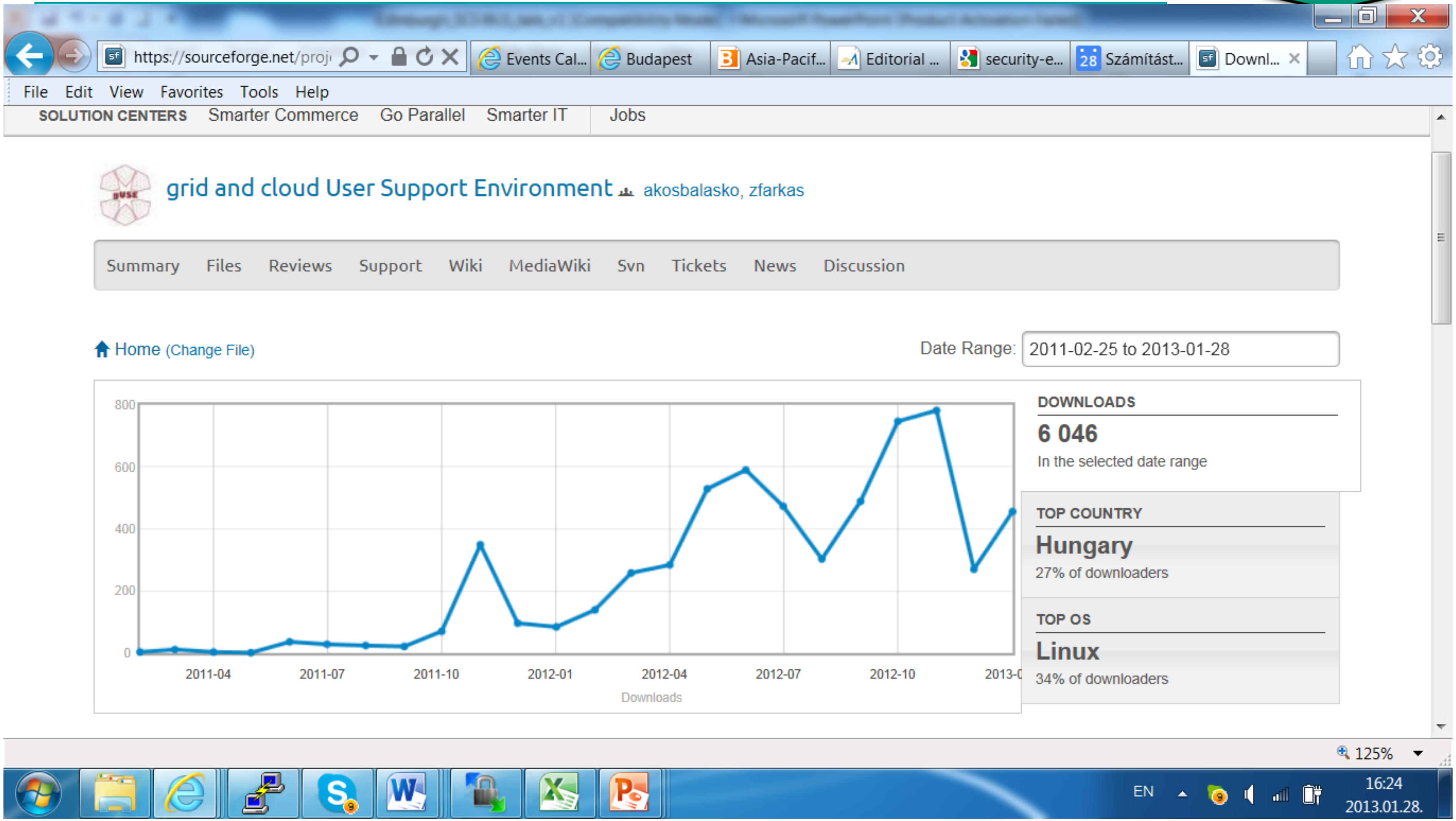
Customization by the Application Specific Module



Where to find further information?

- SCI-BUS web page:
  - http://www.sci-bus.eu/
- SHIWA web page:
  - http://www.shiwa-workflow.eu/
- gUSE/WS-PGRADE:
  - http://www.guse.hu/
- gUSE on sourceforge
  - http://sourceforge.net/projects/guse/
  - http://sourceforge.net/projects/guse/forums/forum/
  - http://sourceforge.net/projects/guse/develop

Sourceforge download statistics



grid and cloud User Support Environment akosbalasko, zfarkas

Date Range: 2011-02-25 to 2013-01-28

DOWNLOADS
6 046
In the selected date range

TOP COUNTRY
Hungary
27% of downloaders

TOP OS
Linux
34% of downloaders

Roadmap of further developing WS-PGRADE/gUSE

- Direct WS-PGRADE workflow upload/ download to/from SHIWA Repository (by March 2013)
- Robot certificate (by April 2013)
- SCI-BUS portlet repository (by May 2013)
- Data Bridge service (by June 2013)
- Introducing interactive workflow nodes (by July 2013)

Conclusions

- If you want to develop a SG instance, use a SG framework instead of developing it from scratch
- Why to select WS-PGRADE/gUSE and join the SCI-BUS communit ?

  1.Robustn
    - Alre

  2.Sustaina
    - The ation plan gua

  3.Function
    - Rich functionalities that are growing according to the SCI-BUS and sourceforge community needs

  4.How easy to adapt for the needs of the new user community?
    - Already large number of gateways customized from gUSE/WS-PGRADE

  5.You can influence the progress of WS-PGRADE/gUSE

Join SCI-BUS as associated member