



GEMLCA / P-GRADE: A workflow-oriented portal and application hosting environment

A short introduction

Tamas Kiss

University Of Westminster

kisst@wmin.ac.uk



www.portal.p-grade.hu
www.cpc.wmin.ac.uk/gemlca



E-scientists' concerns



- How to concentrate own **my own research** if the tool I would like to use is in continuous change?
- How can I learn and understand **the usage of the Grid**?
- How can I **develop Grid applications**?
- How can I **execute grid applications**?
- How to **tackle performance issues**?
- How to **use several Grids at the same time**?
- How to **migrate my application** from one grid to another?
- How can I utilise **legacy applications**?
- How can I **collaborate with fellow researchers**?

The GEMMLCA / P-GRADE Portal gives you the answers!



P-Grade Portal in a nutshell



- **General purpose, workflow-oriented computational Grid portal.** Supports the development and execution of workflow-based Grid applications – **a tool for Grid orchestration**
- Based on **GridSphere-2**
 - Easy to expand with new portlets (e.g. application-specific portlets)
 - Easy to tailor to end-user needs
- Developed by SZTAKI
- **Grid services** supported by the portal:

Service	EGEE grids (LCG/gLite)	Globus 2 grids
Job execution	Computing Element	GRAM
File storage	Storage Element	GridFTP server
Certificate management	MyProxy	
Information system	BDII	MDS-2, MDS-4
Brokering	Workload Management System	(GTbroker)
Job monitoring	Mercury	
Workflow & job visualization	PROVE	
Legacy Code Management	GEMMLCA	

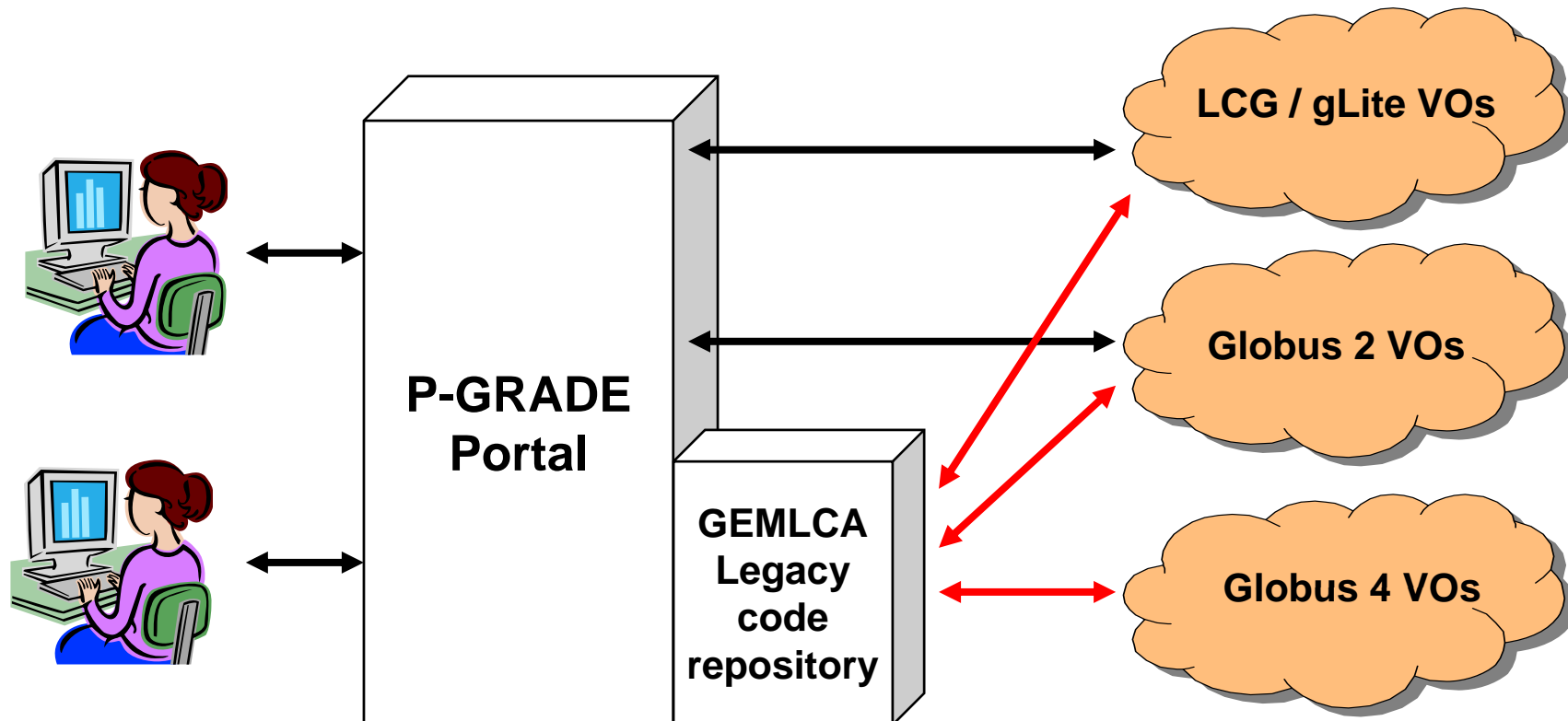
Solves Grid interoperability problem at the level of workflows



GEMMLCA / P-GRADE Portal in a nutshell



- **P-GRADE Portal extended with GEMMLCA back-end**
 - **Sharing jobs and legacy codes as workflow components**
 - **GEMMLCA is a grid service implemented by UoW**
- **A step towards collaborative e-Science**
- **Support for Globus 4 grids (besides GT2 and EGEE)**

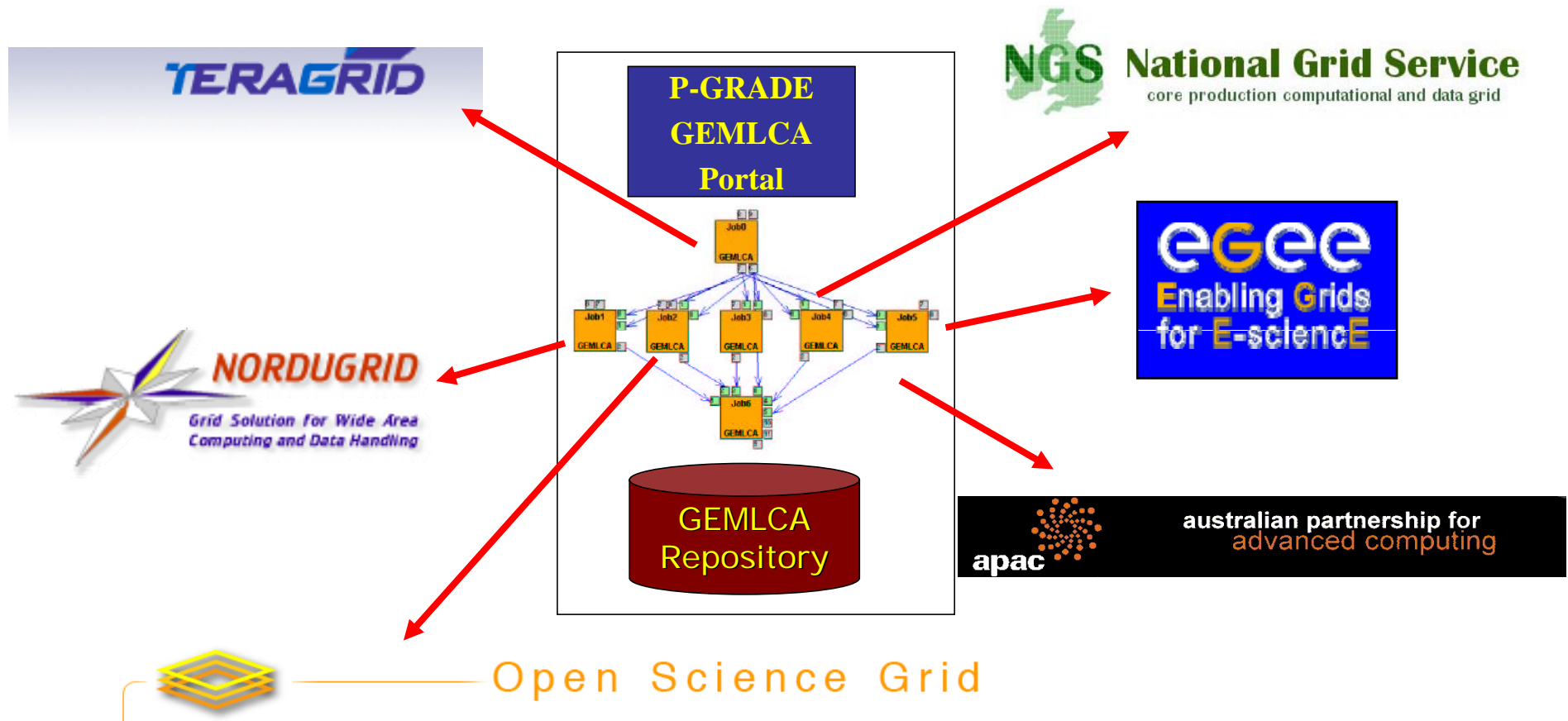




Workflow-level Grid interoperability: The GIN Resource Testing portal

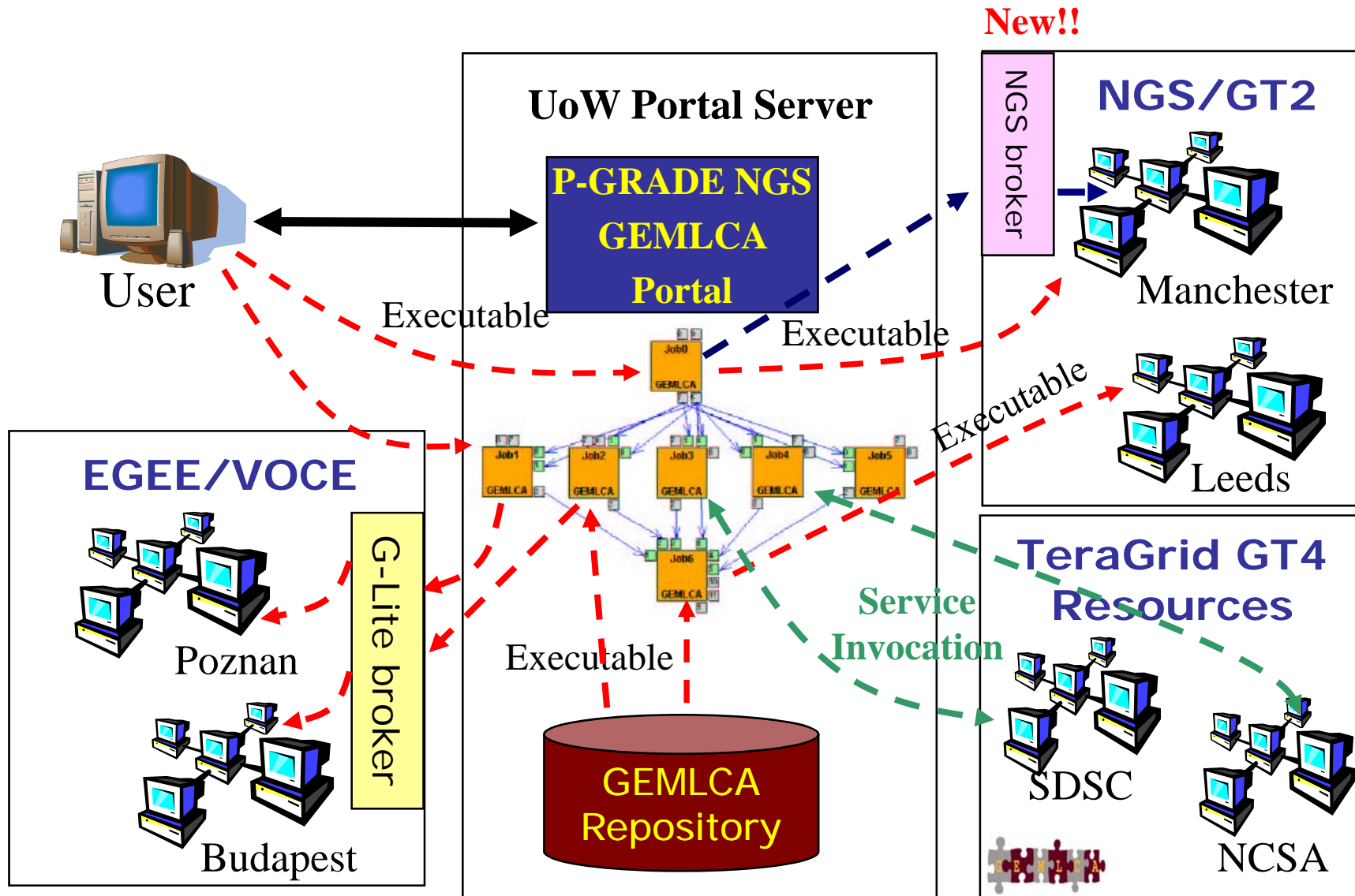


OGF effort to demonstrate workflow level grid interoperability between major production Grids and to monitor OGF GIN VO resources





Workflow level interoperability of Grid systems



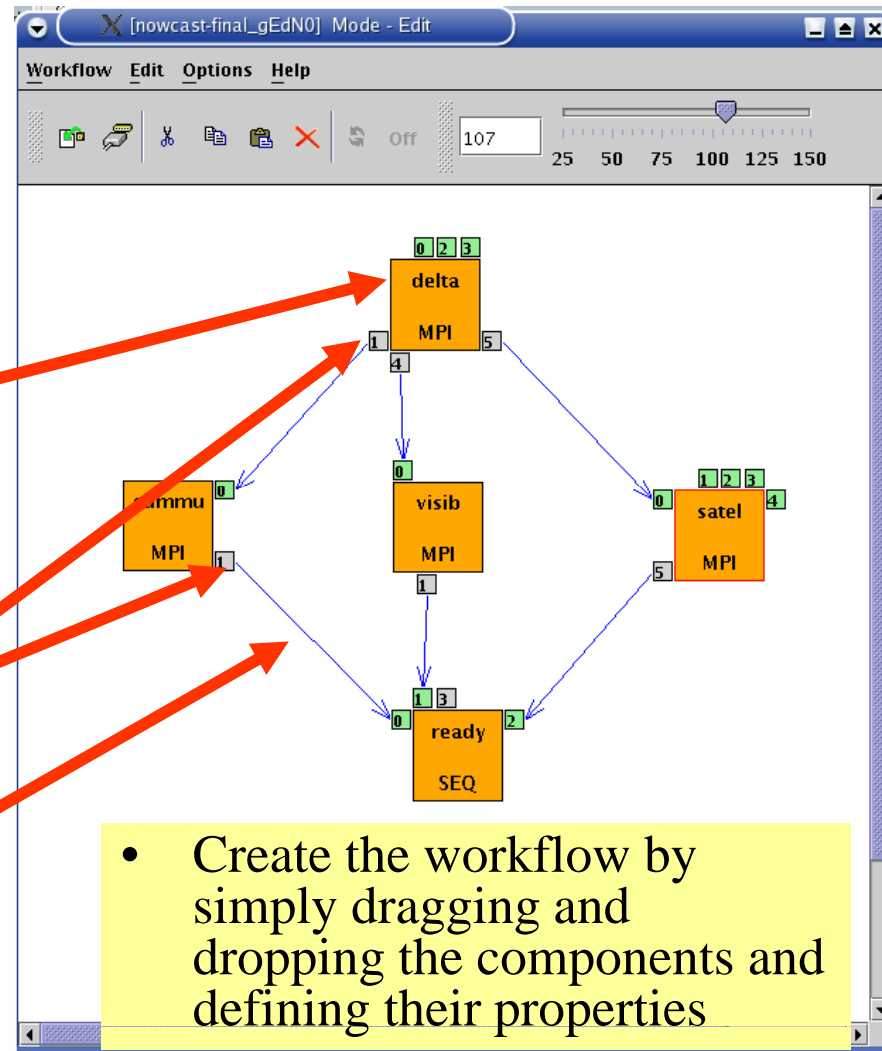


What is a GEMLCA / P-GRADE Portal workflow?



- A directed acyclic graph where:

- Nodes represent jobs - either sequential or parallel programs
- Ports represent input/output files the jobs expect/produce
- Arcs represent file transfer between the jobs

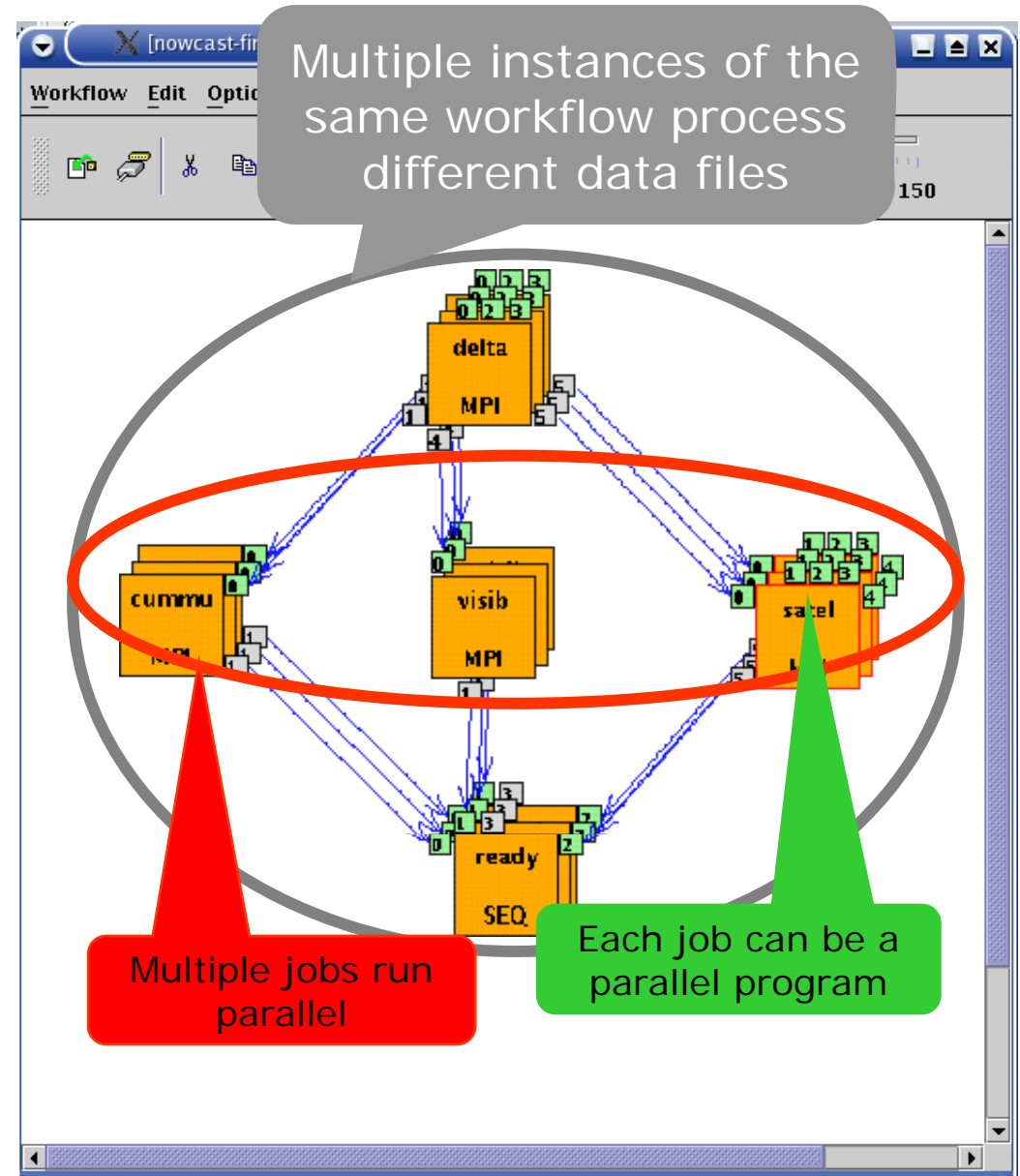




Three levels of parallelism within a workflow

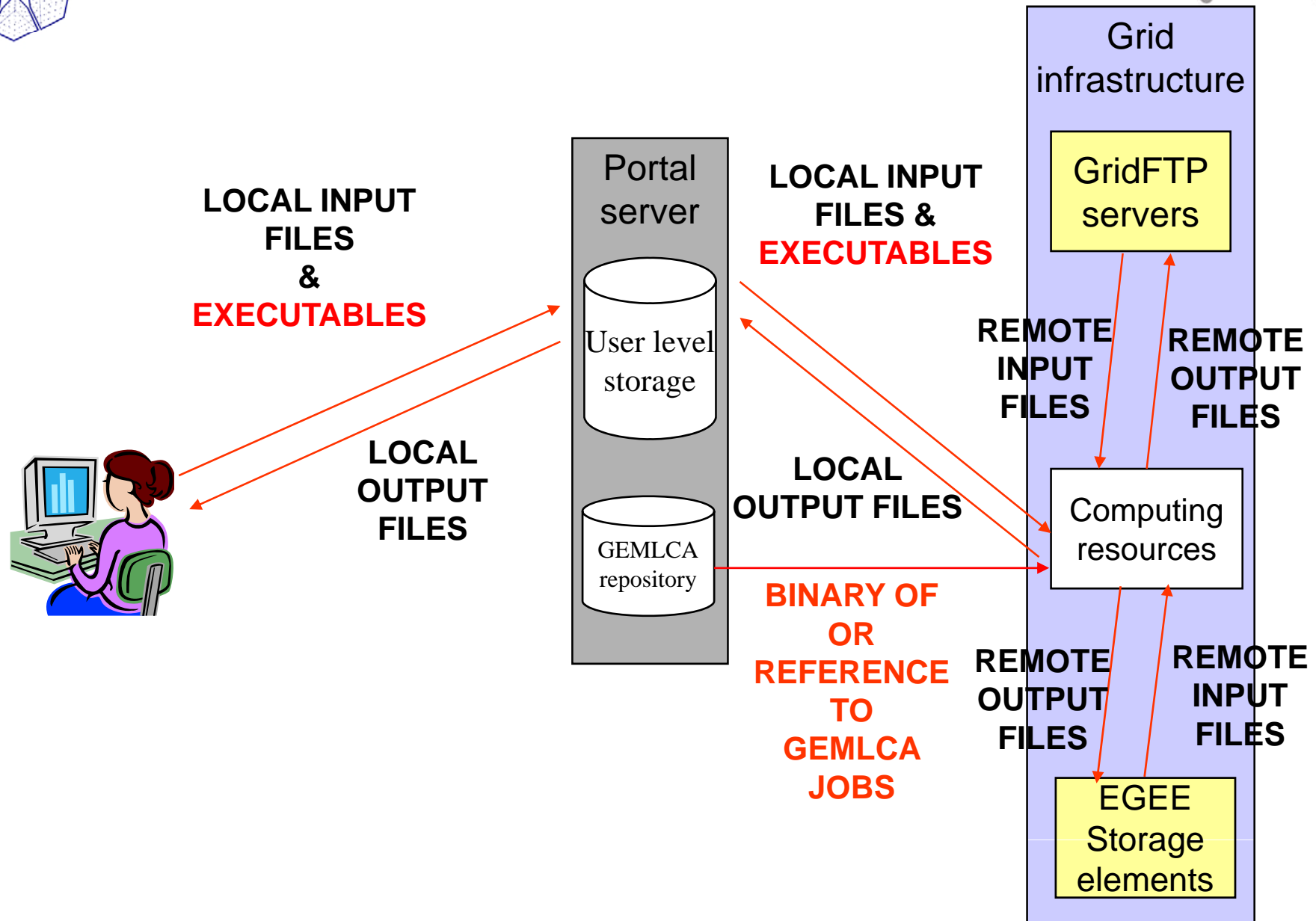


- The workflow concept of the GEMMLCA/ P-GRADE Portal enables the efficient parallelization of complex problems
- Semantics of the workflow enables three levels of parallelism:
 - Parallel execution inside a workflow node (SIMD/MIMD/MISD)
 - Parallel execution among workflow nodes (SIMD/MIMD/MISD)
 - Parameter study execution of the workflow (SIMD)





Workflow level file transfer





Browsing computing resources by the information system portlet



PGrade Portal - Microsoft Internet Explorer

http://hgportal.hpcc.sztaki.hu:7080/gridsphere/gridsphere?action=doChangeVO&cid=15

Workflow Certificates Settings Information System Help

MDS Monitor LOG Monitor

Monitor

Select Grid: SEE-GRID View

Select VO: seegrid View

Grid: SEE-GRID VO: seegrid

Site Name	Computing Element						Storage Element		
	CPU			Job			Space		
	Total	Free	Usage	Running	Waiting	Load	Total	Available	Usage
AEGIS01-PHY-SCL	112	80	29%	7	0	0%	226.793 GB	216.34 GB	5%
AEGIS02-RCUB	20	20	0%	0	0	0%	398.466 GB	396.58 GB	0%
BG01-IPP	54	18	67%	4	0	0%	609.554 GB	473.543 GB	22%
	20	16	20%	1	0	0%	131.775 GB	79.957 GB	39%
	3	3	0%	0	0	0%	566.608 GB	566.376 GB	0%
	48	32	33%	2	5	71%	554.647 GB	475.767 GB	14%
	60	12	80%	4	0	0%	78.317 GB	6.271 GB	92%
	28	28	0%	0	0	0%	69.709 GB	69.075 GB	1%
	54	24	56%	5	36	88%	849.666 GB	828.387 GB	3%
-01	24	24	0%	0	0	0%	862.807 GB	848.676 GB	2%
	4	4	0%	0	0	0%	4.566 GB	2.871 GB	37%
	35	28	20%	1	0	0%	1.335 TB	1.335 TB	0%

**Graphical
interface for
GIIS and
BDII servers**



GMT – GEMLCA Monitoring Toolkit



- to test resource availability
- implementation is based on MDS4
- probes are implemented as scripts and their outputs are displayed in a monitoring portlet
- Runs on the NGS and GIN portals

GridSphere Portal - Microsoft Internet Explorer

RELEASE 2.4

University of Westminster

GridSphere | **GT4/GEMLCA Monitor** | GIN VO Information

GT4/GEMLCA Monitor

ServiceGroup Overview

This page provides a brief overview of Web Services and/or WS-Resources that are members of a WS-ServiceGroup.

This WS-ServiceGroup has 54 direct entries, 54 in whole hierarchy.

Resource Type	ID	Information
gmtgemcalistcodes	161.74.12.24	GTT Probe "gmtgemcalistcodes" for https://161.74.12.24:9000/wsrf/services/grid-compute.cpc.wmin.ac.uk
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3104/wsrf/services/tg-grid.uc.teragrid.org
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3107/wsrf/services/tg-login1.sdsc.teragrid.org
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3101/wsrf/services/gm6.cluster.cpc.wmin.ac.uk
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3106/wsrf/services/maverick.tacc.utexas.edu
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3103/wsrf/services/grid-hg.ncsa.teragrid.org
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3100/wsrf/services/node40.cluster.cpc.wmin.ac.uk
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3105/wsrf/services/th1.uits.iupui.edu
gmtgridftptest	antaeus.hpcc.ttu.edu	GTT Probe "gmtgridftptest" for gsiftp://antaeus.hpcc.ttu.edu:2811
gmtprewsgramtest	antaeus.hpcc.ttu.edu	GTT Probe "gmtprewsgramtest" for gram://antaeus.hpcc.ttu.edu:2119/jobmanager-fork
gmtgridftptest	ariane.doc.ic.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://ariane.doc.ic.ac.uk:55101
gmtwsgramtest	ariane.doc.ic.ac.uk	GTT Probe "gmtwsgramtest" for https://ariane.doc.ic.ac.uk:55100/wsrf/services/ManagedJobFactoryService
gmtgridftptest	ariane.doc.ic.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://ariane.doc.ic.ac.uk:55001
gmtprewsgramtest	ariane.doc.ic.ac.uk	GTT Probe "gmtprewsgramtest" for gram://ariane.doc.ic.ac.uk:55000/jobmanager-fork
gmtprewsgramtest	atlas.iu.edu	GTT Probe "gmtprewsgramtest" for gram://atlas.iu.edu:2119/jobmanager-pbs
gmtgridftptest	atlas.iu.edu	GTT Probe "gmtgridftptest" for gsiftp://atlas.iu.edu:2811
gmtgridftptest	cmsdsk00.hep.ph.ic.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://cmsdsk00.hep.ph.ic.ac.uk:2811
gmtgridftptest	fermigrd1.fnal.gov	GTT Probe "gmtgridftptest" for gsiftp://fermigrd1.fnal.gov:2811
gmtprewsgramtest	fermigrd1.fnal.gov	GTT Probe "gmtprewsgramtest" for gram://fermigrd1.fnal.gov:2119/jobmanager-condor
gmtgridftptest	grid-compute.cpc.wmin.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://grid-compute.cpc.wmin.ac.uk:2811
gmtprewsgramtest	grid-compute.cpc.wmin.ac.uk	GTT Probe "gmtprewsgramtest" for gram://grid-compute.cpc.wmin.ac.uk:2119/jobmanager-condor



Certificate Manager

Multi-grid → Multi-proxy



Multiple proxies can be available on the portal server at the same time!

PGrade Grid portal - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://grid-portal.cpc.wmin.ac.uk:8080/gridsphere/gridsphere?cid=91

Cavendish IT Services Finance Helpdesks Human Resources Report Fault Support services Westminster

NGS P-GRADE GEMILCA portal

NGS P-GRADE portal University of Westminster

Information page

Welcome Workflow **Certificates** Settings Information System GEMILCA Administration Tool Help

Certificates

Certificate Manager

Issuer	Set for Grids	Time left	[Actions]
C=UK,O=eScience,OU=Westminster,L=ComputerScience,CN=tamas kiss,CN=proxy	NGS dteam_GLITE_BROKER	75:16:26	Details Set for Grid Delete
C=HU,O=NIIF CA,OU=GRID,OU=NIIF,CN=Gergei Sipos,E=sipos@sztaki.hu,CN=proxy	gilda_LCG_2_BROKER	75:17:59	Details Set for Grid Delete

[Refresh](#)

[Download](#) (Download certificate from MyProxy server.) [Upload](#) (Upload authentication data to MyProxy server.)

Message: [Press a button.]

9 January 2008

NGS

EGEE GILDA



Workflow Execution

(observation by the workflow portlet)



PGrade Portal - Microsoft Internet Explorer

Workflow Manager

Refresh Back

Workflow	Job	Gridname	Hostname	Status	[Logs]	[Output]	[Visualization]	
LM_9_DEMO_TOTAL				running	-	N/A	Visualize	All Abort
	INIT	SEE-GRID	ce01.grid.acad.bg	finished	- -		-	
	LM_P	SEE-GRID	n40.hpcc.sztaki.hu	init	- -		-	
	LM_P.2	SEE-GRID	n40.hpcc.sztaki.hu	init	- -		-	
	LM_S	SEE-GRID	grid-ce.ii.edu.mk	running	- -		-	
	LM_S.2	SEE-GRID	grid1.irb.hr	finished	Out	-	-	
	LM_S.3	SEE-GRID	grid1.netmode.ece.ntua.gr	running	Out	-	-	
	LM_S.4	SEE-GRID	grid1.irb.hr	finished	Out	-	-	
	LM_S.5	SEE-GRID	testbed001.grid.ici.ro	running	Out	-	-	
	LM_S.6	SEE-GRID	chemgrid3.chemres.hu	finished	Out	-	-	
	TIFF	HUNGRID	grid109.kfki.hu	init	- -		-	

Message: Job list refreshed.

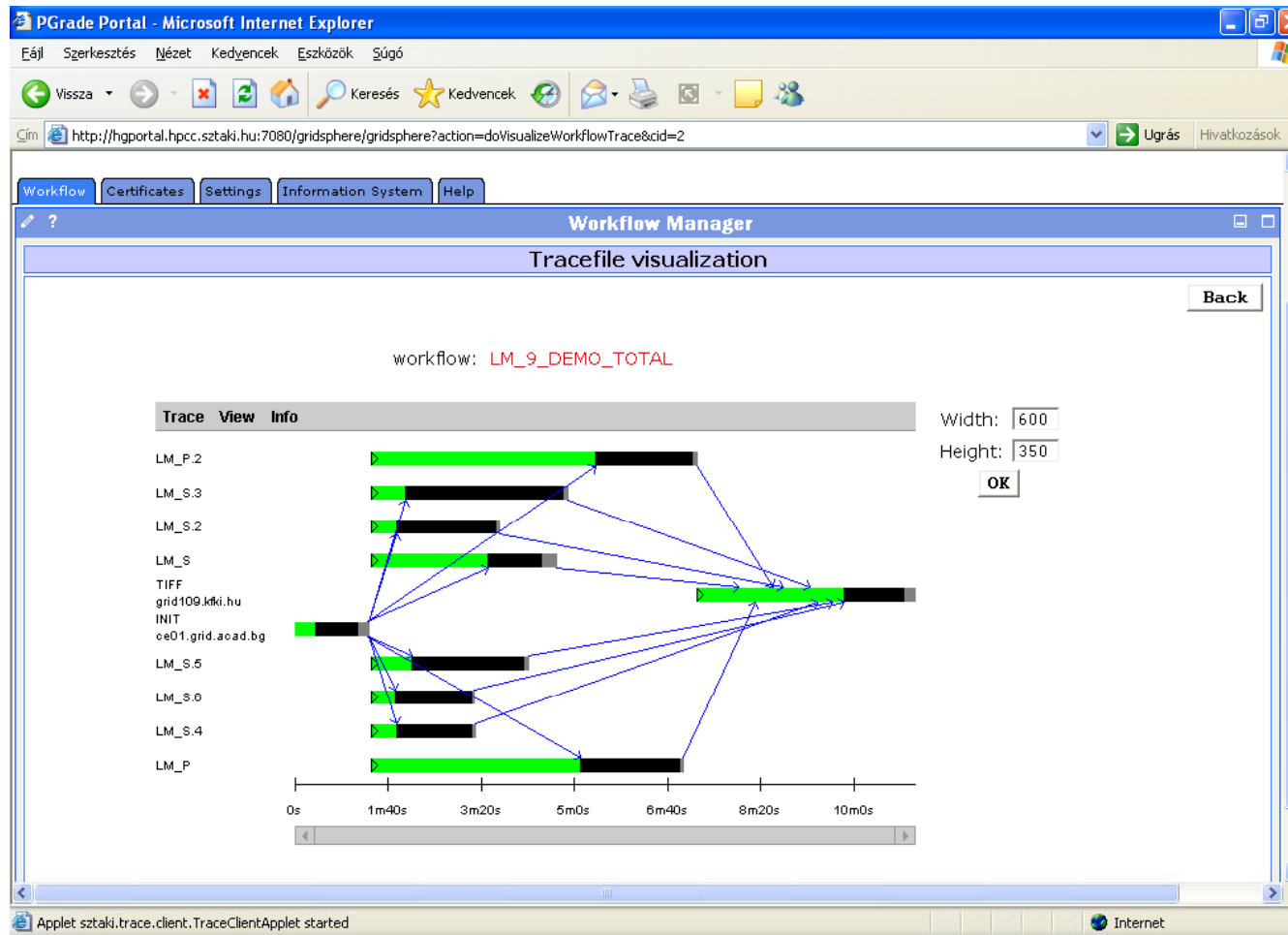
White/Red/Green color means the job is initial/running/finished state



On-Line Monitoring both at the workflow and job levels *(workflow portlet)*



- The portal monitors and visualizes workflow progress



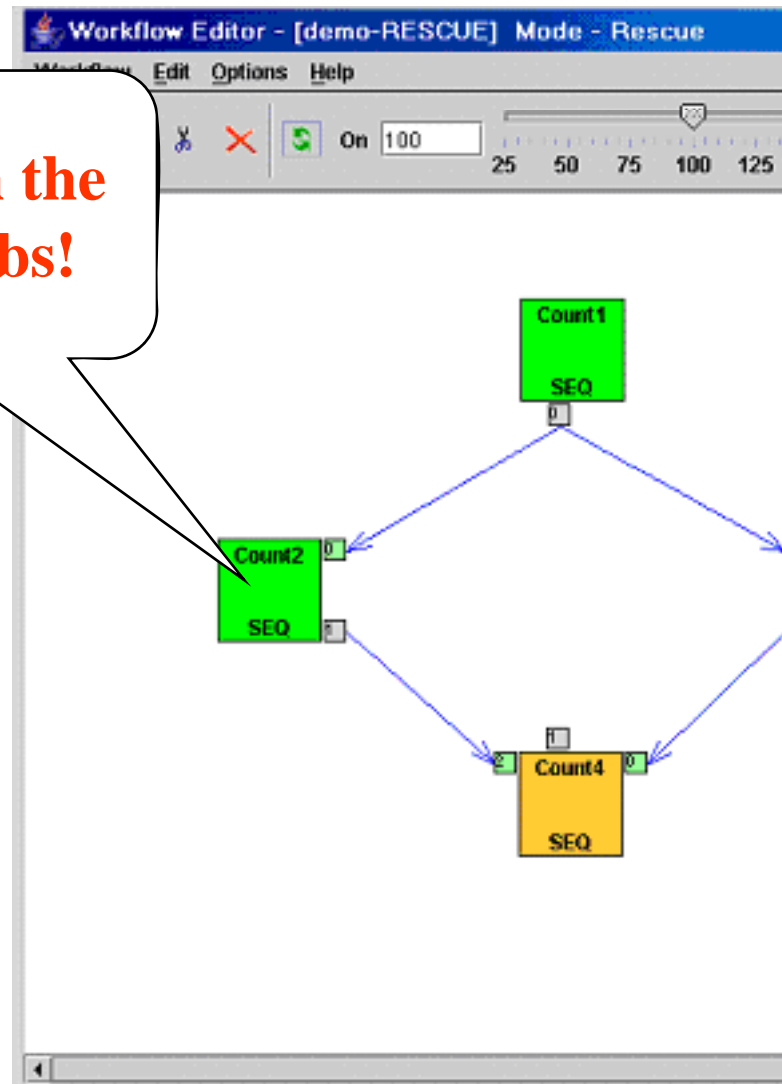


Rescuing a failed workflow



Don't touch the finished jobs!

Map the failed job onto a different resource or download a new proxy for it



The execution can continue from the point of failure



Downloading the results...



The screenshot shows the P-Grade Portal web interface in a Mozilla browser window. The page title is "PGrade Portal - Mozilla" and the URL is "http://fn2.hpcc.sztaki.hu:9080/gridsphere/gridsphere?action=doGotoPage&cid=2". The page features the P-Grade logo and a "portal" link. A "Logout" link is visible in the top right corner. Below the logo, there is a navigation menu with "Workflow", "Credentials", "Settings", "Demo", and "Help" buttons. The main content area is titled "Workflow Manager" and contains a "Refresh" button and a "Back" button. A table titled "Job list" displays the following data:

Workflow	Job	Hostname	Status	[Logs]	[Output]	[Visualization]	[Action]
nowcast-final-g_SGE			finished		<input checked="" type="checkbox"/>	Visualize All	Subm Attach Delete
	cummu	n0.hpcc.sztaki.hu	finished	--		Visualize	
	delta	n0.hpcc.sztaki.hu	finished	--		Visualize	
	ready	n0.hpcc.sztaki.hu	finished	--		Visualize	
	satel	n0.hpcc.sztaki.hu	finished	--		Visualize	
	visib	n0.hpcc.sztaki.hu	finished	--		Visualize	

Below the table, a message states: "Message: Job list refreshed." A dialog box titled "Opening nowcast_final_g.zip" is overlaid on the bottom right of the browser window. The dialog contains the following text: "The file 'nowcast_final_g.zip' is of type application/x-zip-compressed, and Mozilla does not know how to handle this file type. This file is located at: e:\pri\mc04". Below this text, it asks "What should Mozilla do with this file?" and provides four radio button options: "Open it with the default application", "Open it with" (with a text input field and a "Choose..." button), "Save it to disk" (which is selected), and "Always perform this action when handling files of this type". At the bottom of the dialog are "OK" and "Cancel" buttons.



Putting a successfully finished job into the GEMLCA repository



Workflow Certificates Settings Demo Help GEMLCA Administration Tools Macroscopic Visualiser

Resource Selector Legacy Code Information Descriptor Creator

GEMLCA LCID Administration Portl

GEMLCA Legacy Code Interface Descriptor

Legacy code Environment Paramaters:

maximumProcessors

executable

minimumProcessors

maximumJob

jobManager

id

description

List of legacy code Arguments:

name	file	order	fixed	inputOutput	mandatory	regexp	friendlyName	commandline
-p	No	0	No	Input	No		Folder to be created	Yes

New argument entry form:

name

file

order

fixed

inputOutput

mandatory

regexp

friendlyName

commandline

initialValue

Mkdir Legacy Code exposed as a Grid Service

Folder : ../.gemlca/legacycodes/mkdir

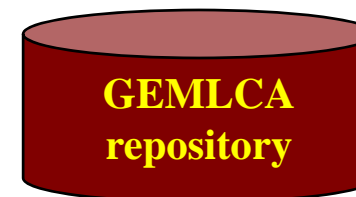
Content : i) mkdir binary or link ii) config.xml

Legacy Code Interface Description File: config.xml

```

<?xml version="1.0"?>
<!DOCTYPE GLCEnvironment "gemlcaconfig.dtd">
<GLCEnvironment
  id="mkdir" executable="LINUX/mkdir" jobManager="Fork"
  maximumJob="11" minimumProcessors="1"
  maximumProcessors="1" universe="PVM"
>
<Description>Unix mkdir program</Description>
<GLCParameters>
  <Parameter name="-p" friendlyName="Folder to be created"
    fixed="No" inputOutput="Input" order="0"
    mandatory="No" fileCommandline="Commandline">
    <initialValue> </initialValue>
  </Parameter>
</GLCParameters>
</GLCEnvironment>

```

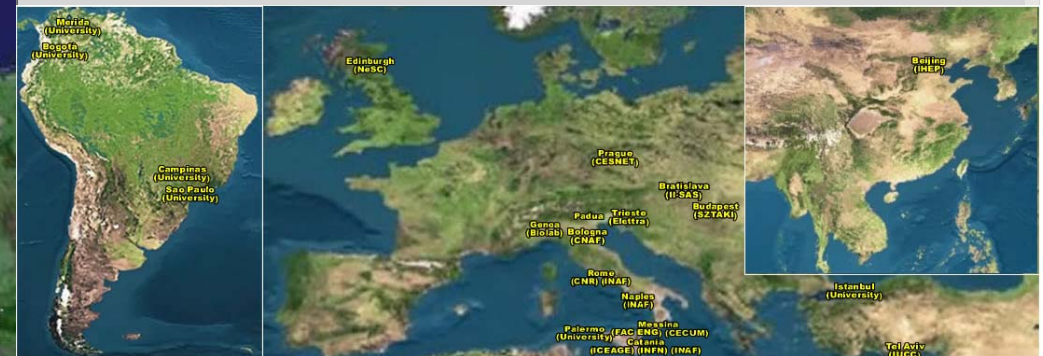
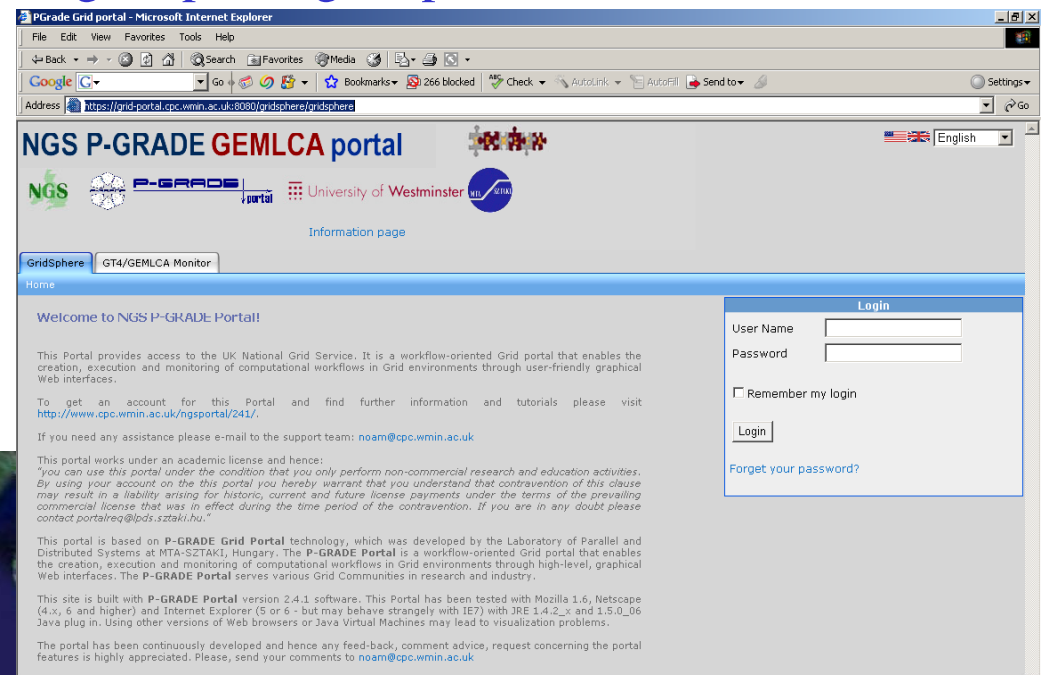




GEMLCA / P-GRADE on the UK NGS: NGS P-GRADE GEMLCA Portal



- portal website:
<https://gngs-portal.cpc.wmin.ac.uk:8080/gridsphere/gridsphere>
- Interface for NGS GT2 sites
- Interface for GT4 Westminster site
- Interface for EGEE GILDA sites
- Connected to the NGS and the GILDA Resource Brokers





Other P-GRADE Portal installations



- P-GRADE Portal service is available for
 - SEE-GRID infrastructure
 - Central European VO of EGEE
 - GILDA: Training VO of EGEE
 - US Open Science Grid, TeraGrid
 - Economy-Grid, Swiss BioGrid, Bio and Biomed EGEE VOs, BioInfoGrid, BalticGrid
 - **OGF GIN (also connected to NGS)**





Both P-GRADE portal and GEMMLCA are open source



- **P-GRADE Portal**
 - <http://sourceforge.net/projects/pgportal/>

- **GEMMLCA**
 - <http://dev.globus.org/wiki/Incubator/GEMMLCA>
 - A Globus incubator project



Coming Soon to the NGS *P-Grade* portal!



- SRB support
 - SRB resources are integrated at **workflow** level
 - Input/output ports can represent SRB data sources
- OGSA-DAI support
 - A set of OGSA-DAI browser and manipulation portlets
- Parameter study support
 - If the user has a workflow he can run it with many different parameters
 - Workflow = code to execute
 - Input files = parameters



***Thank you for your
attention!***

***Hands-on session with the P-
GRADE/GEMMLCA portal will now
follow.***