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 GEMLCA / P-GRADE Portal gives you the answers!

P-GRADE Portal in a nutshell

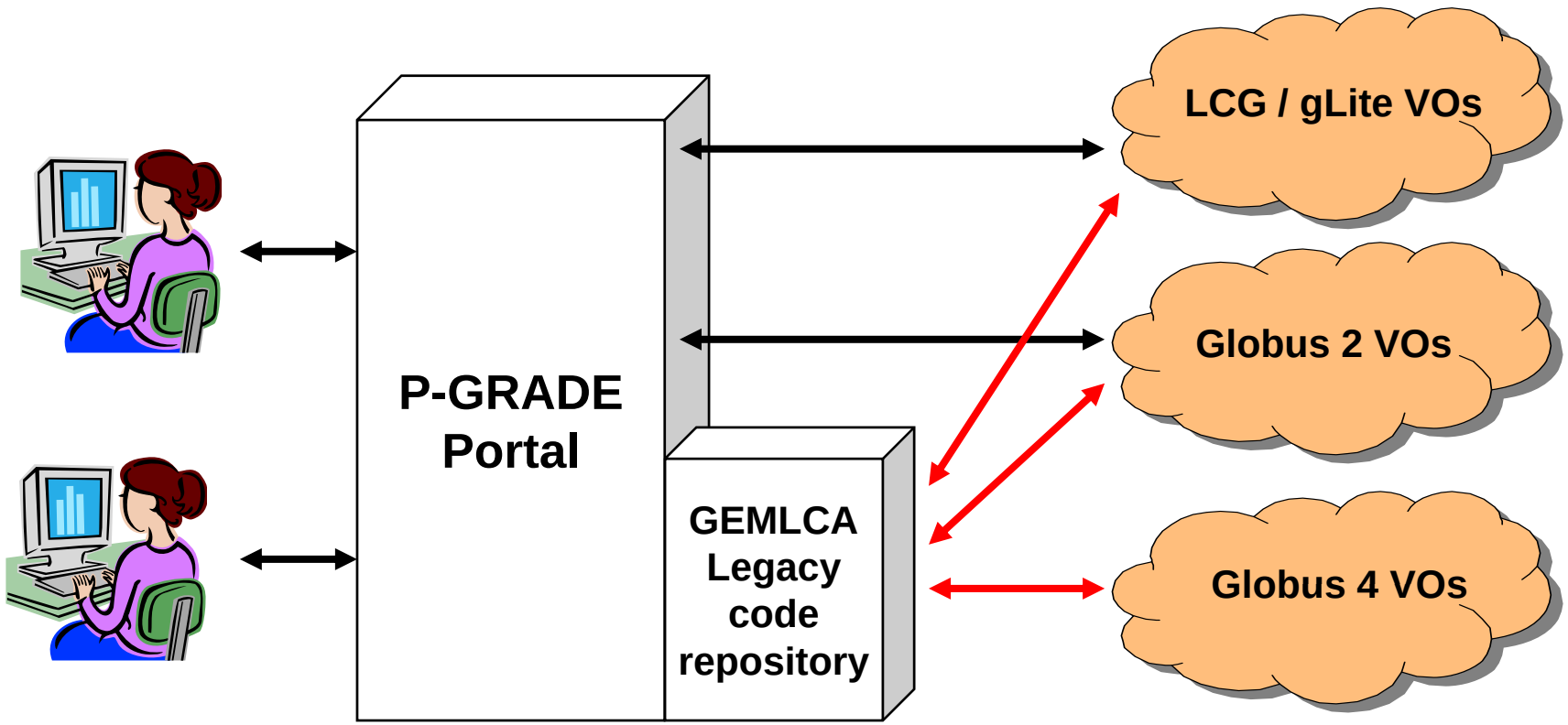
- **General purpose**, **workflow-oriented computational Grid portal**. Supports the development and execution of workflow-based Grid applications – **a tool for Grid orchestration**
- Based on **GridSphere-2**
  - Easy to expand with new portlets (e.g. application-specific portlets)
  - Easy to tailor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

Solves Grid interoperability problem at the level of workflows

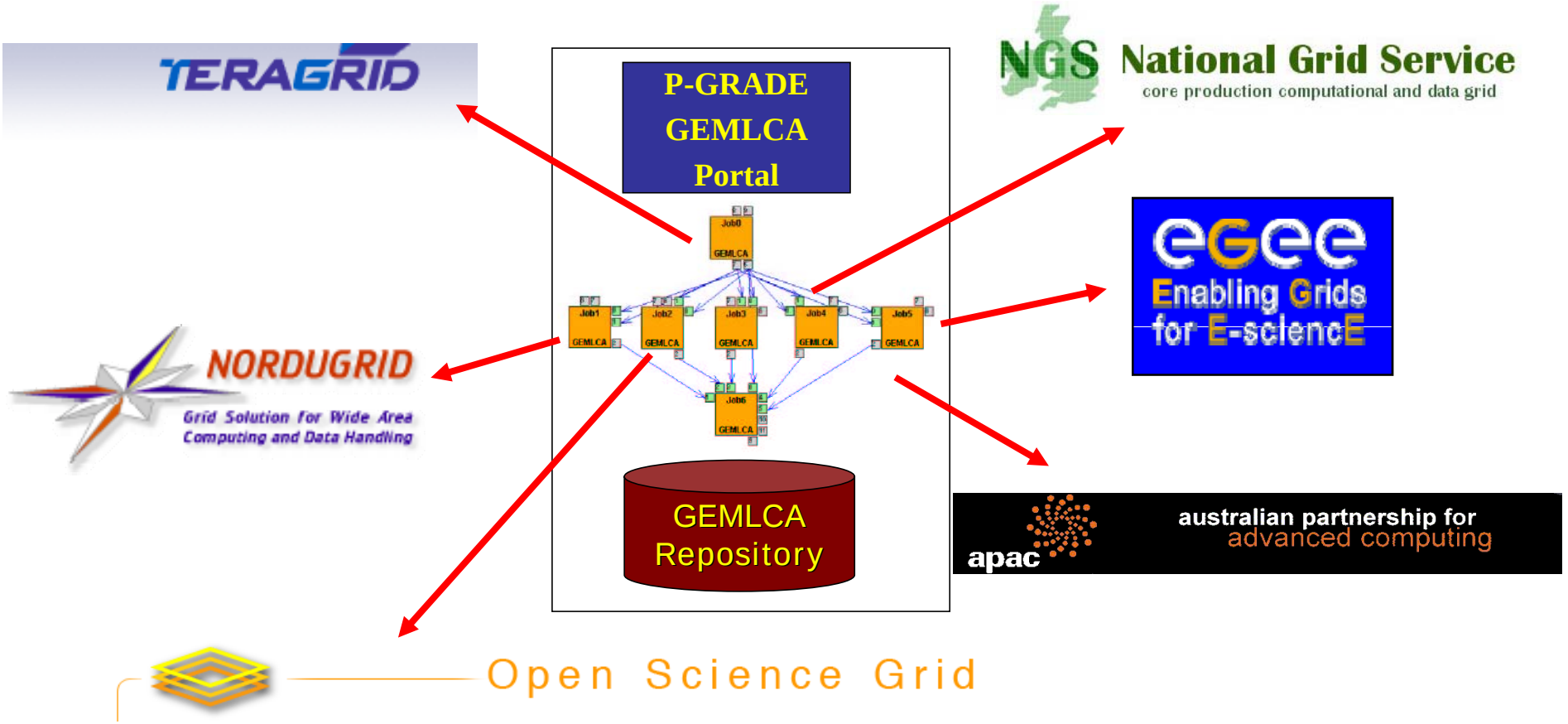
GEMLCA / P-GRADE Portal in a nutshell

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

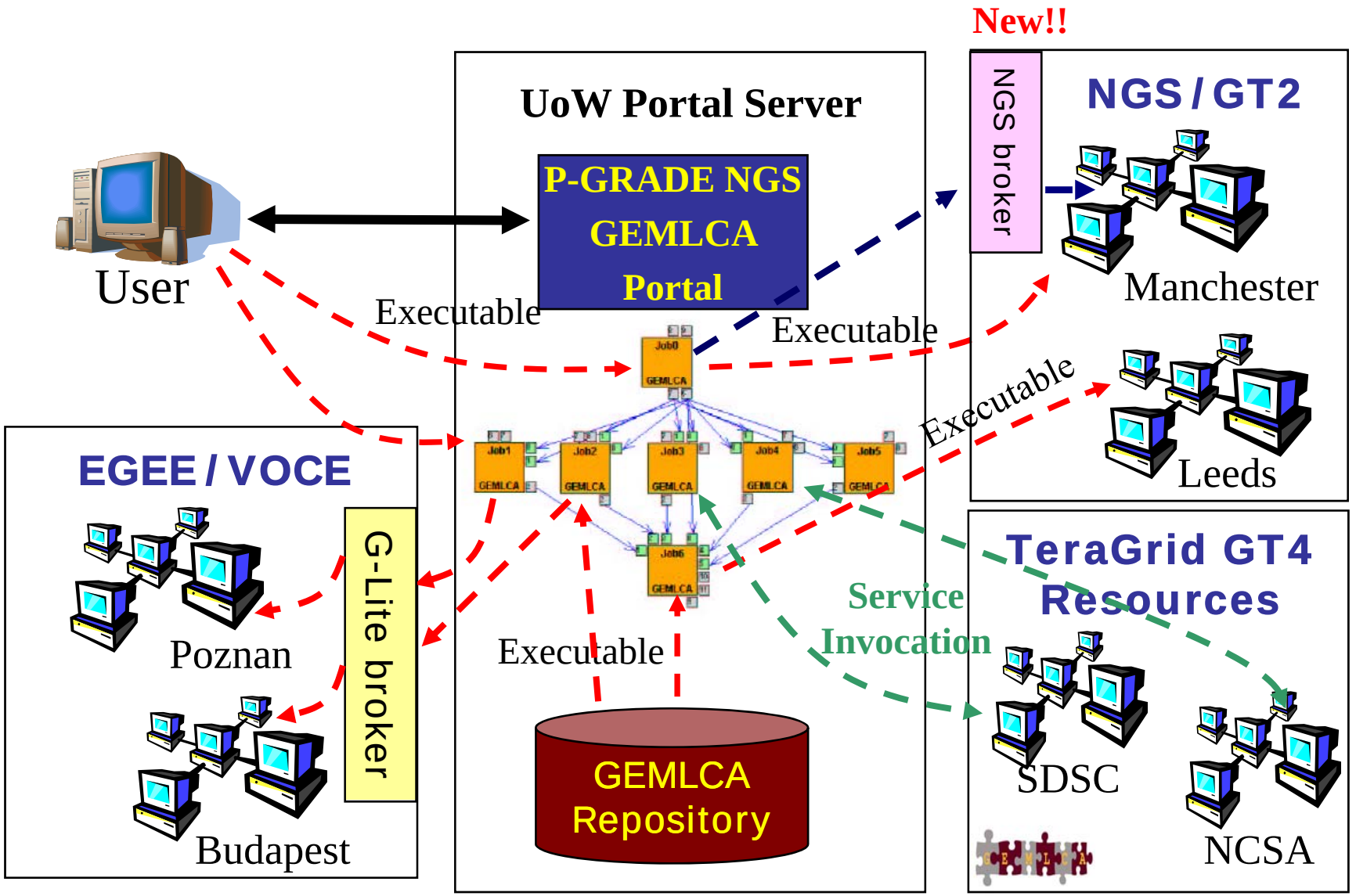


Workflow-level Grid interoperability: The GIN Resource Testing portal

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

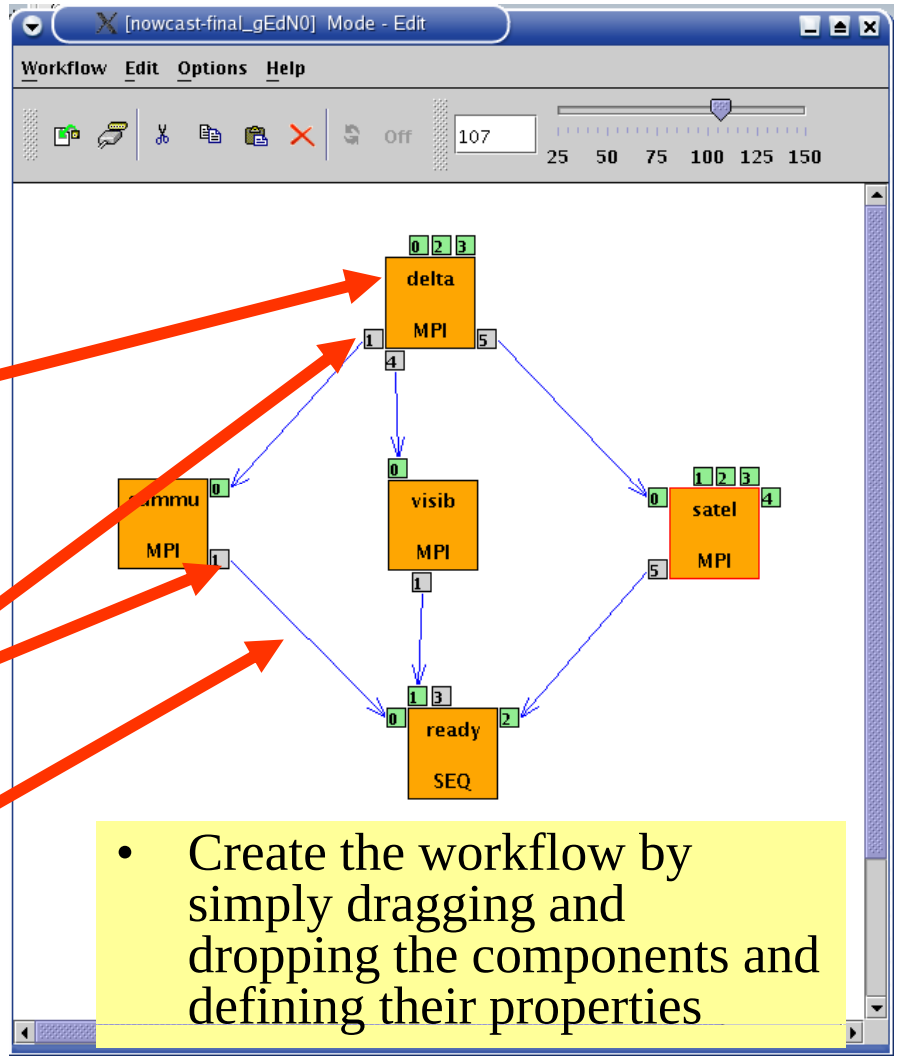


Workflow level interoperability of Grid systems



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



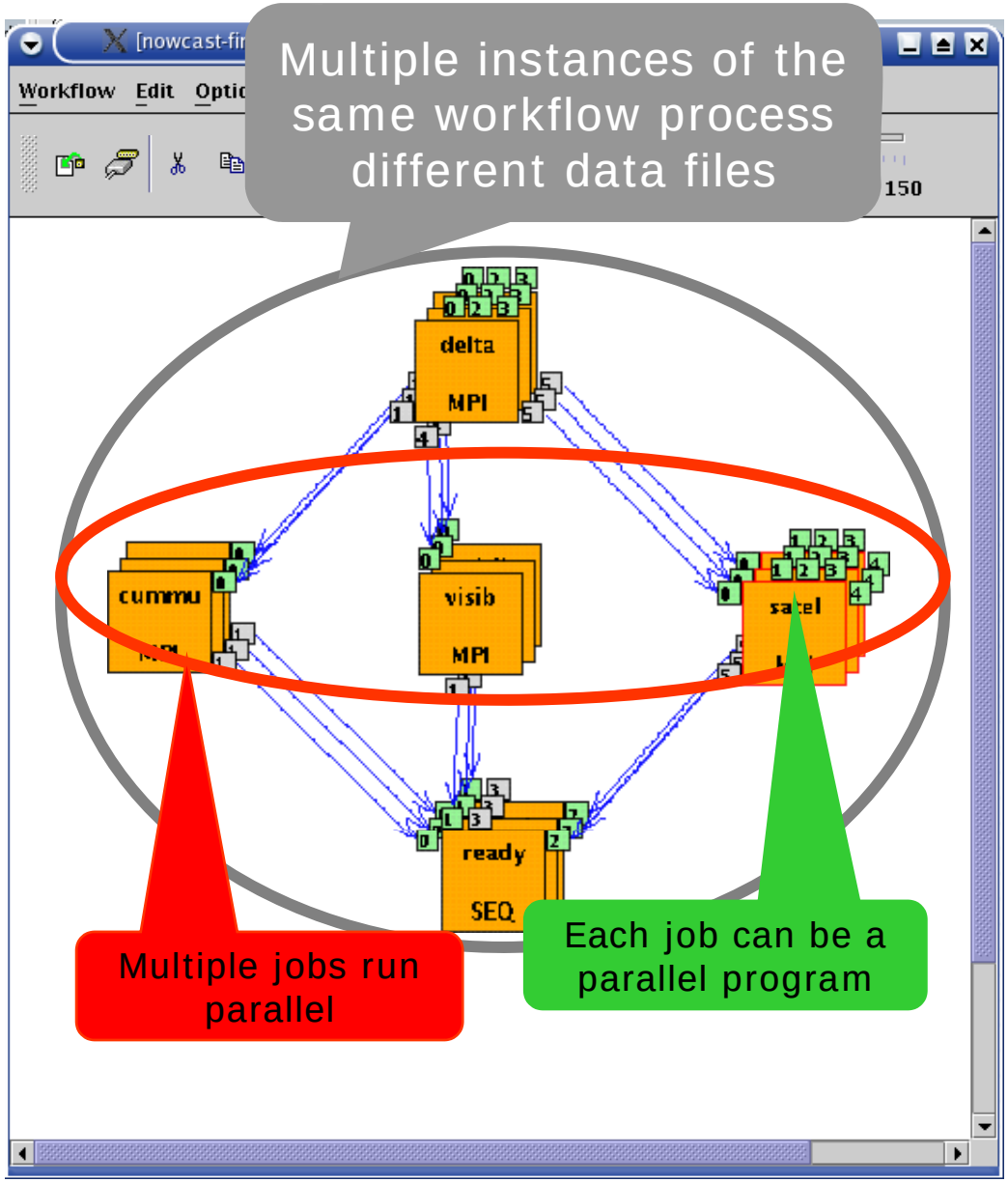
Three levels of parallelism within a workflow

- The workflow concept of the GEMLCA/ P-GRADE Portal enables the efficient parallelization of complex problems
- Semantics of the workflow enables three levels of parallelism:

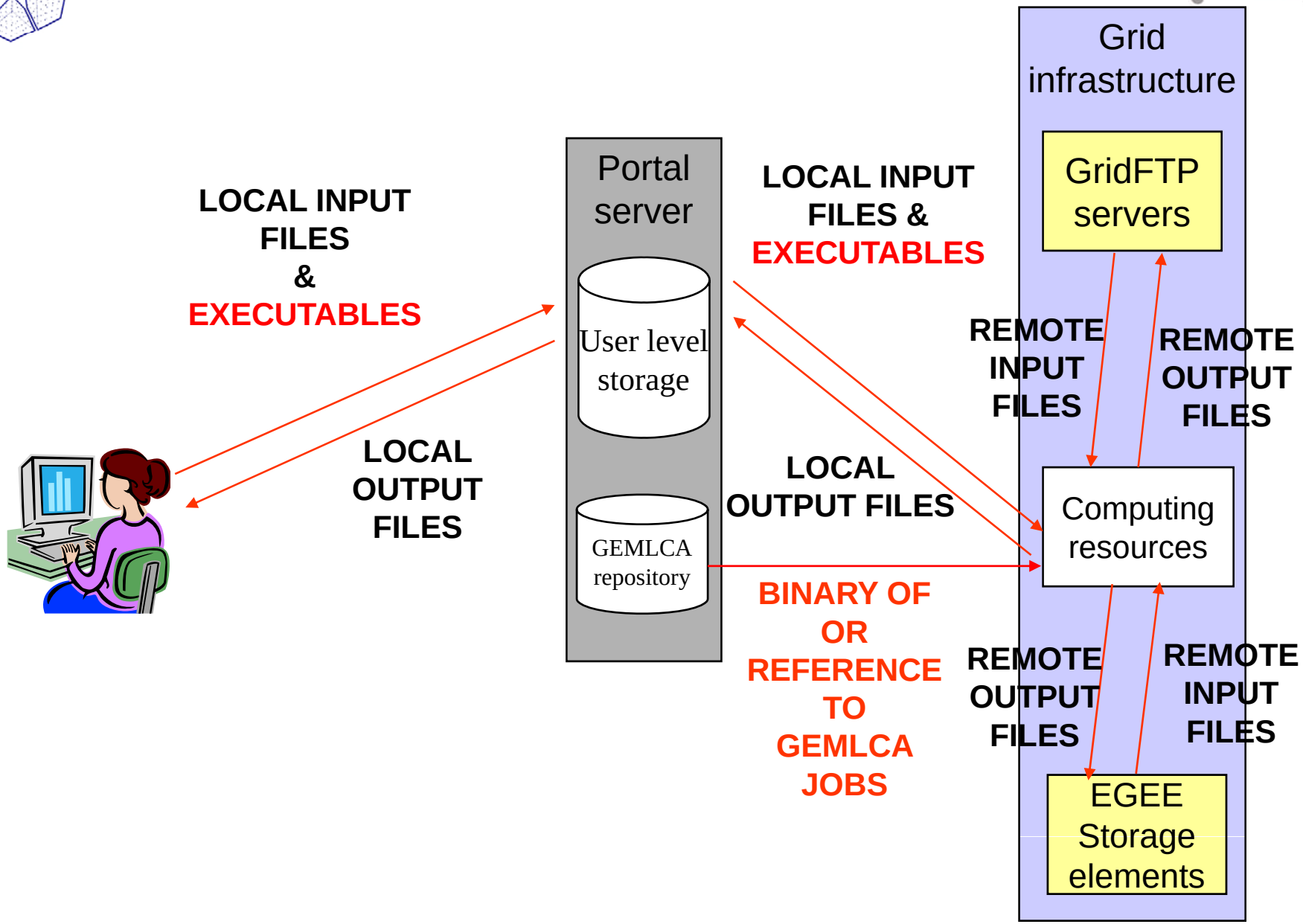
  – **Parallel execution inside a workflow node (SIMD/MIMD/MISD)**

  – **Parallel execution among workflow nodes (SIMD/MIMD/MISD)**

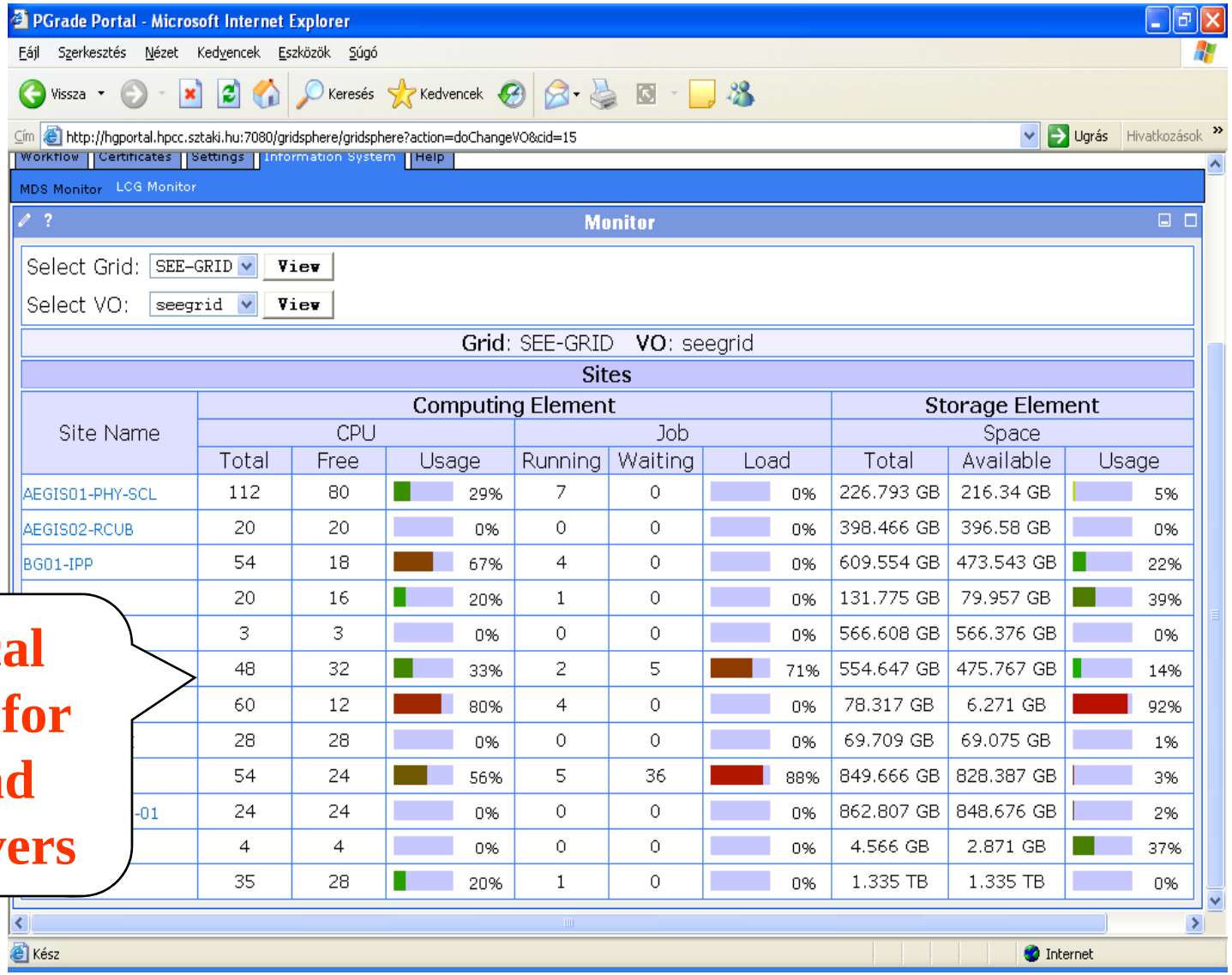
  – **Parameter study execution of the workflow (SIMD)**



Workflow level file transfer



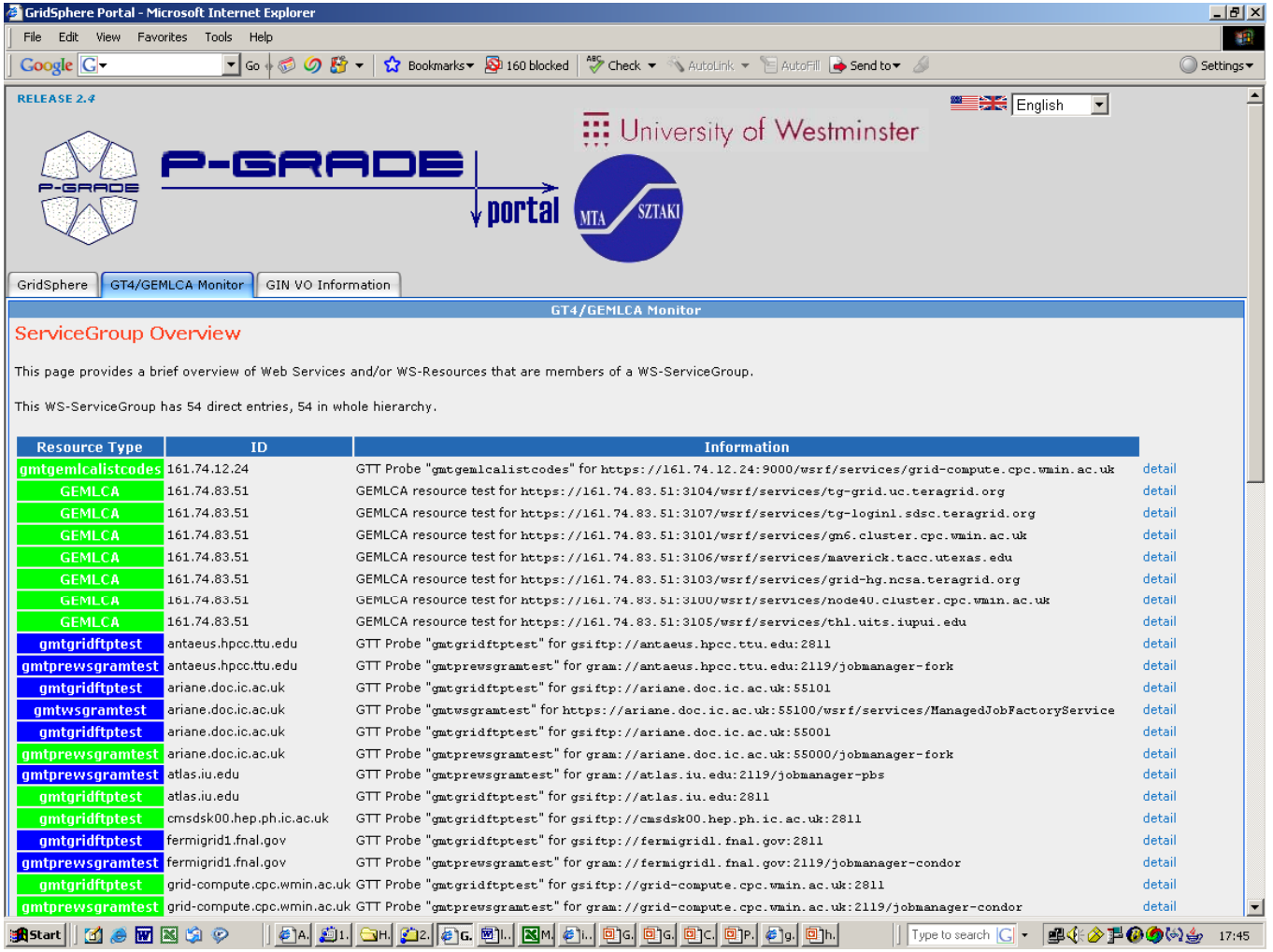
Browsing computing resources by the information system portlet



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

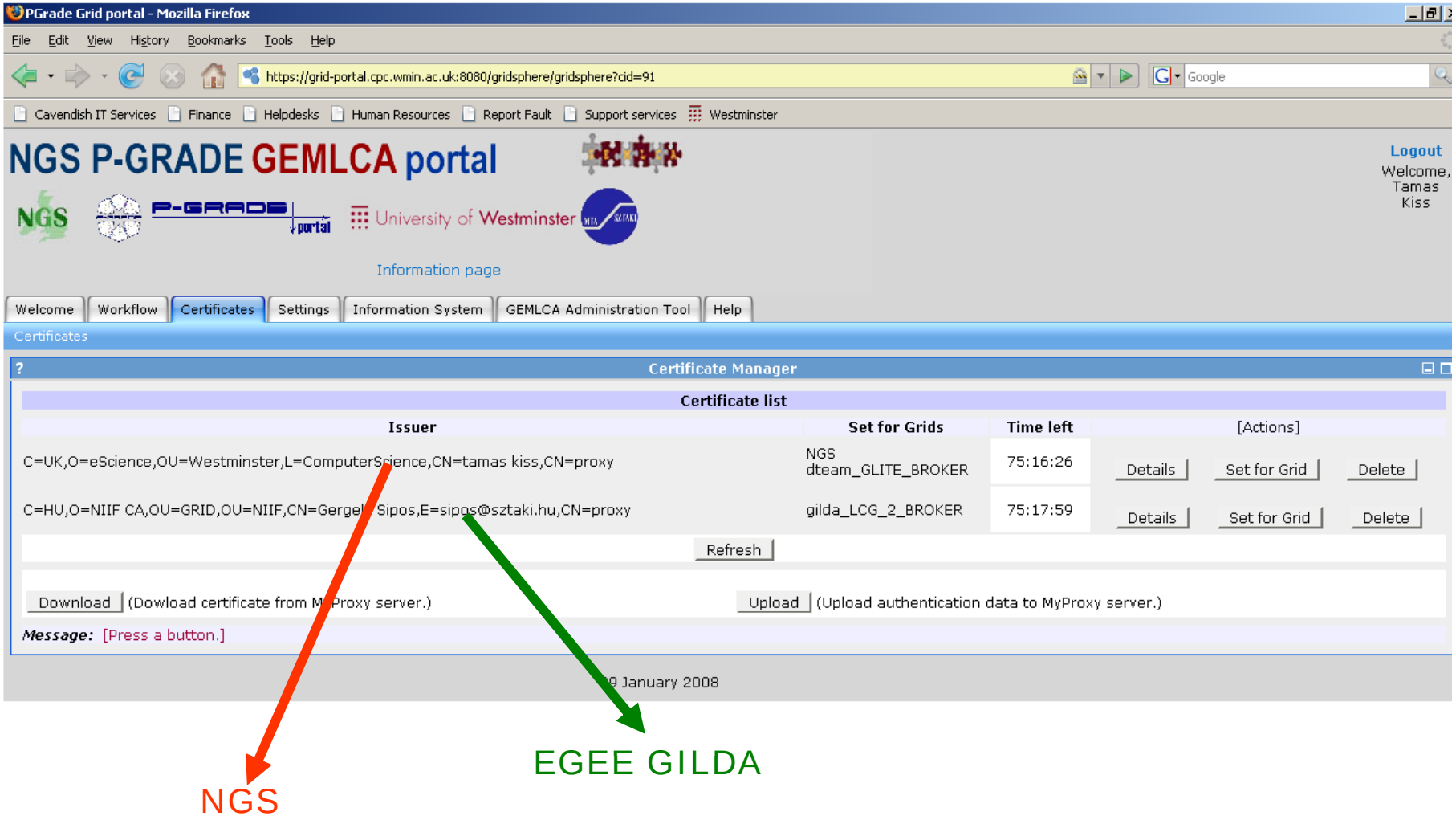


Resource Type	ID	Information	
gmtgemlcalistcodes	161.74.12.24	GTT Probe "gmtgemlcalistcodes" for https://161.74.12.24:9000/wsrf/services/grid-compute.cpc.wmin.ac.uk	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3104/wsrf/services/tg-grid.uc.teragrid.org	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3107/wsrf/services/tg-loginl.sdsc.teragrid.org	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3101/wsrf/services/gm6.cluster.cpc.wmin.ac.uk	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3106/wsrf/services/maverick.tacc.utexas.edu	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3103/wsrf/services/grid-hg.ncsa.teragrid.org	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3100/wsrf/services/node40.cluster.cpc.wmin.ac.uk	detail
GEMLCA	161.74.83.51	GEMLCA resource test for https://161.74.83.51:3105/wsrf/services/thl.uits.iupui.edu	detail
gmtgridftptest	antaeus.hpcc.ttu.edu	GTT Probe "gmtgridftptest" for gsiftp://antaeus.hpcc.ttu.edu:2811	detail
gmtprewsgramtest	antaeus.hpcc.ttu.edu	GTT Probe "gmtprewsgramtest" for gram://antaeus.hpcc.ttu.edu:2119/jobmanager-fork	detail
gmtgridftptest	ariane.doc.ic.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://ariane.doc.ic.ac.uk:55101	detail
gmtwsgramtest	ariane.doc.ic.ac.uk	GTT Probe "gmtwsgramtest" for https://ariane.doc.ic.ac.uk:55100/wsrf/services/ManagedJobFactoryService	detail
gmtgridftptest	ariane.doc.ic.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://ariane.doc.ic.ac.uk:55001	detail
gmtprewsgramtest	ariane.doc.ic.ac.uk	GTT Probe "gmtprewsgramtest" for gram://ariane.doc.ic.ac.uk:55000/jobmanager-fork	detail
gmtprewsgramtest	atlas.iu.edu	GTT Probe "gmtprewsgramtest" for gram://atlas.iu.edu:2119/jobmanager-pbs	detail
gmtgridftptest	atlas.iu.edu	GTT Probe "gmtgridftptest" for gsiftp://atlas.iu.edu:2811	detail
gmtgridftptest	cmsdsk00.hep.ph.ic.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://cmsdsk00.hep.ph.ic.ac.uk:2811	detail
gmtgridftptest	fermigrid1.fnal.gov	GTT Probe "gmtgridftptest" for gsiftp://fermigrid1.fnal.gov:2811	detail
gmtprewsgramtest	fermigrid1.fnal.gov	GTT Probe "gmtprewsgramtest" for gram://fermigrid1.fnal.gov:2119/jobmanager-condor	detail
gmtgridftptest	grid-compute.cpc.wmin.ac.uk	GTT Probe "gmtgridftptest" for gsiftp://grid-compute.cpc.wmin.ac.uk:2811	detail
gmtprewsgramtest	grid-compute.cpc.wmin.ac.uk	GTT Probe "gmtprewsgramtest" for gram://grid-compute.cpc.wmin.ac.uk:2119/jobmanager-condor	detail

Certificate Manager

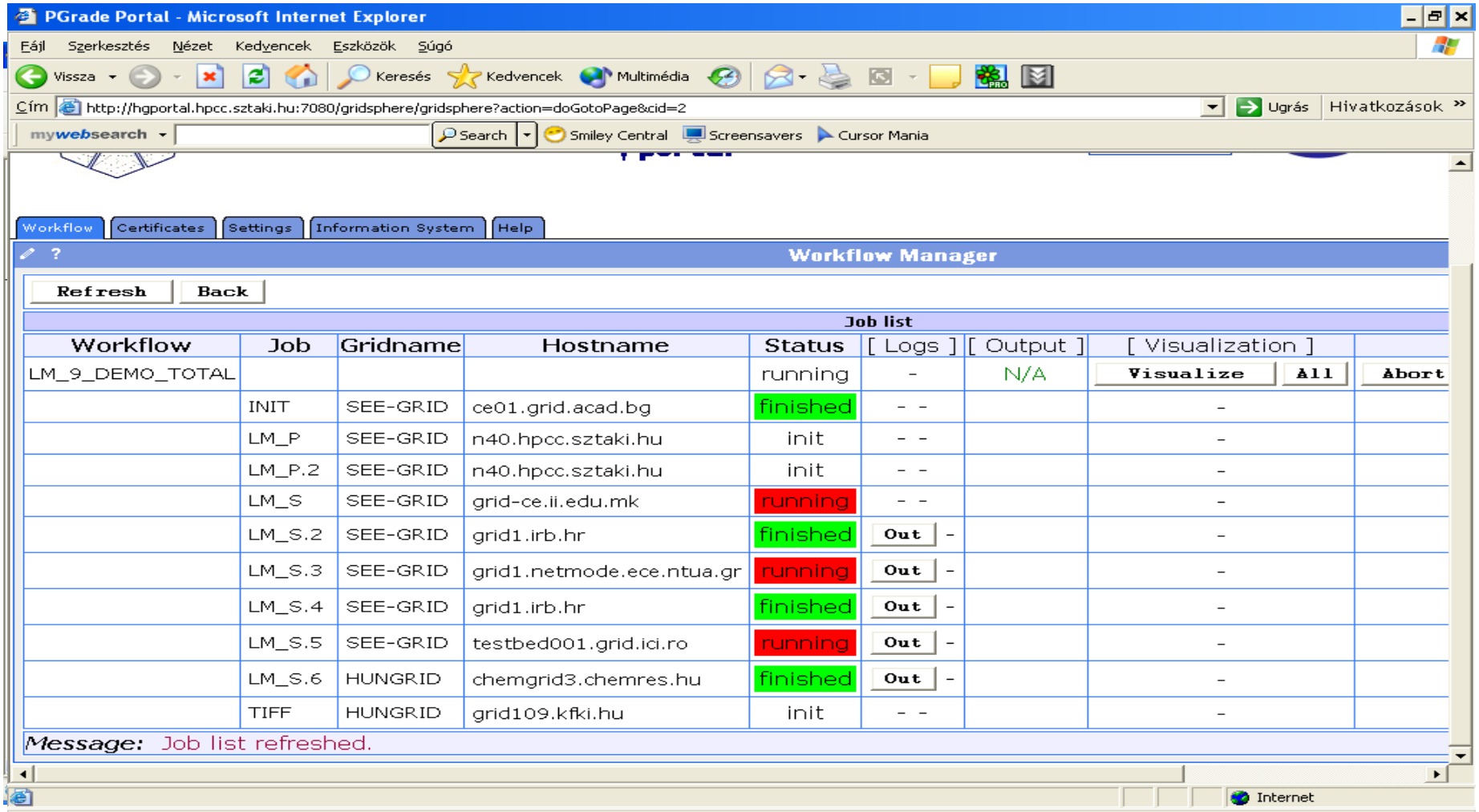
Multi-grid → Multi-proxy

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



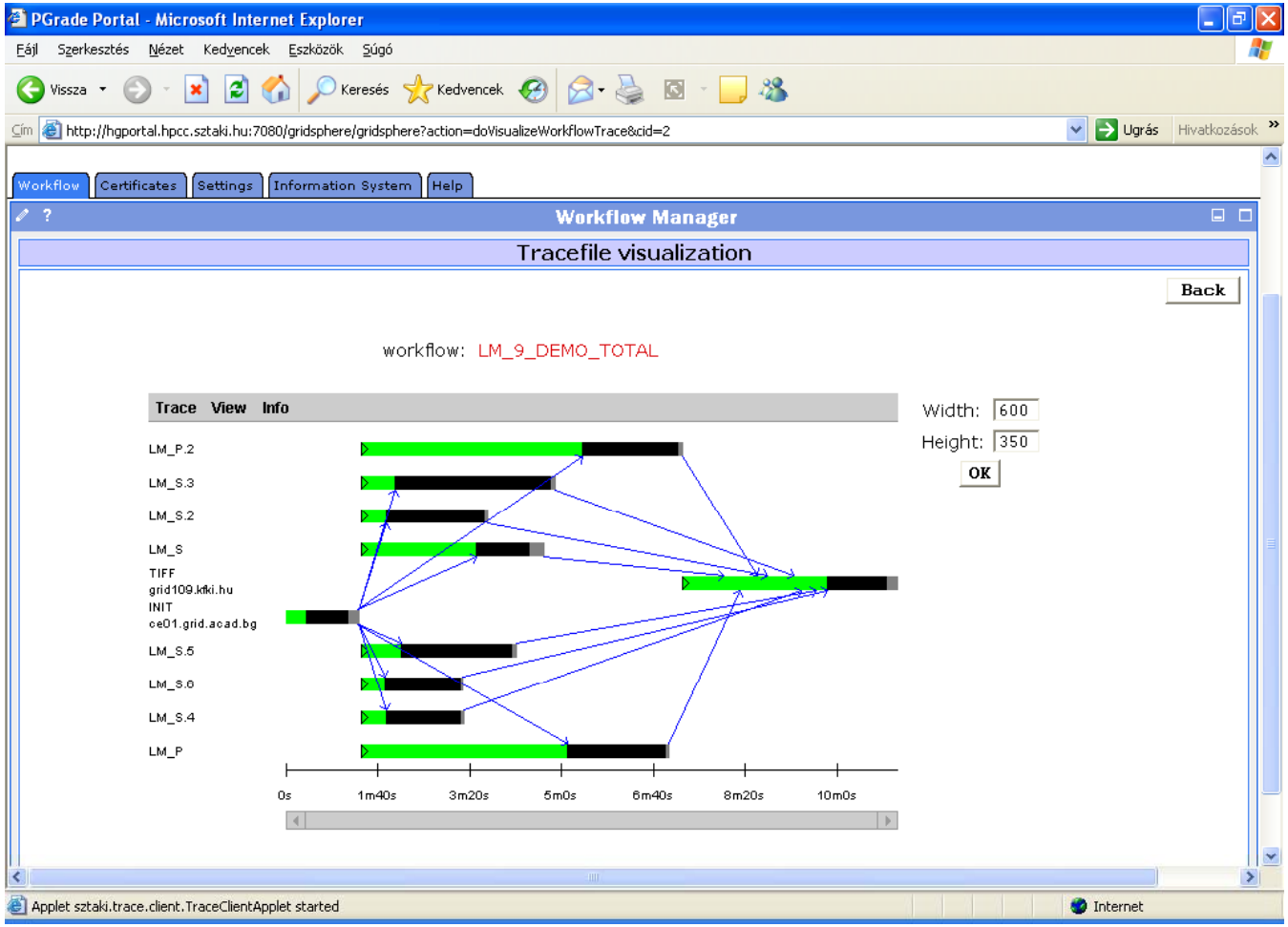
Workflow Execution

(observation by the workflow portlet)



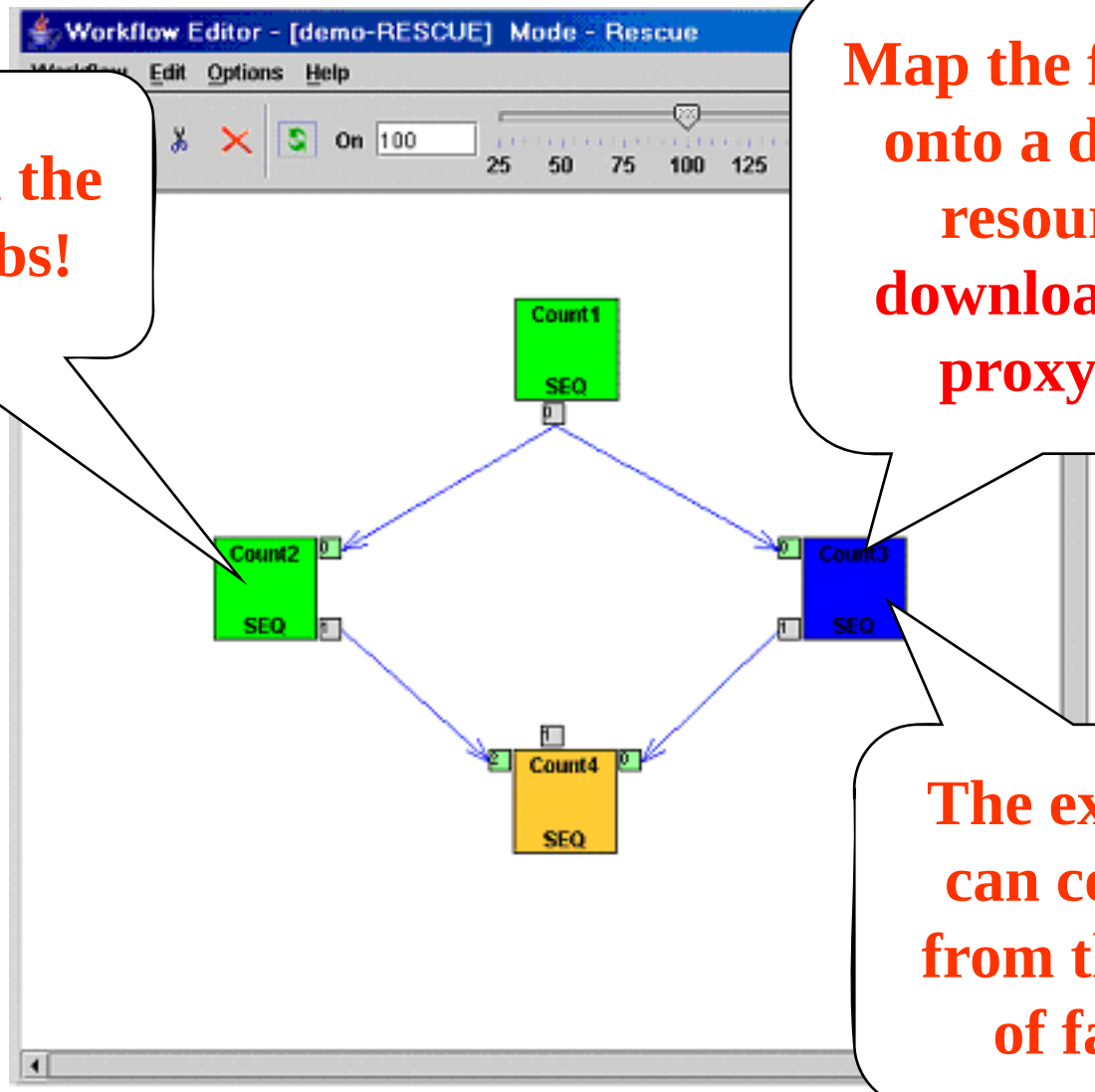
White/Red/Green color means the job is initial/running/finished state

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



- The portal monitors and visualizes workflow progress

Rescuing a failed workflow

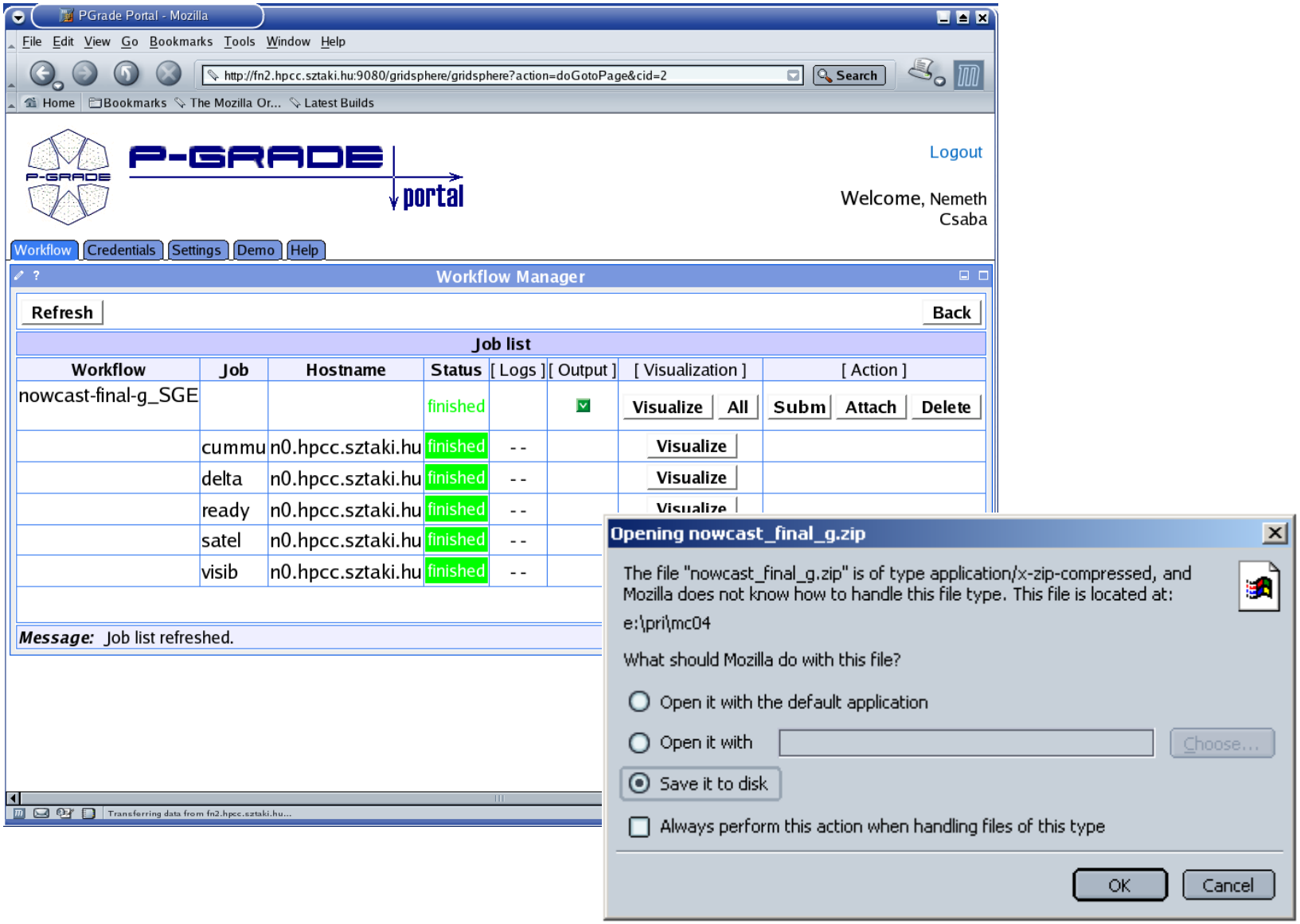


Don't touch the finished jobs!

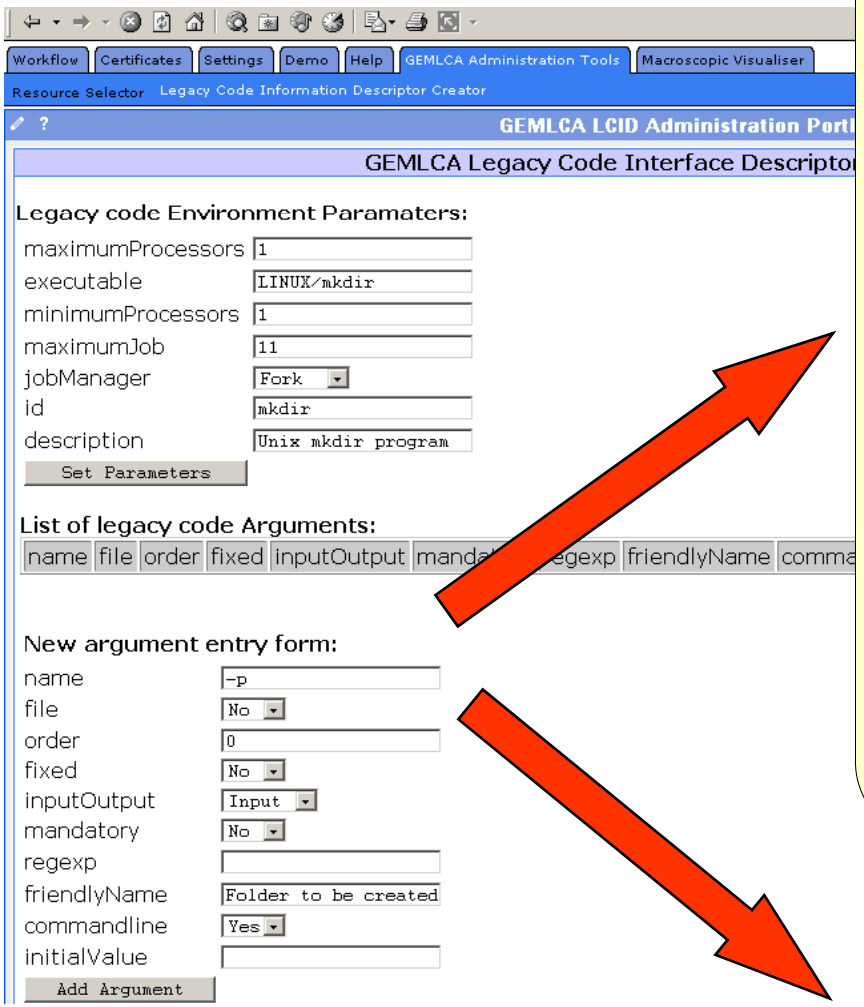
Map the failed job onto a different resource or download a new proxy for it

The execution can continue from the point of failure

Downloading the results…



Putting a successfully finished job into the GEMLCA repository



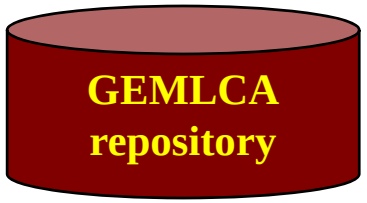
Mkdir Legacy Code exposed as a Grid Service

Folder : /../.gemlca/legacycodes/mkdir

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

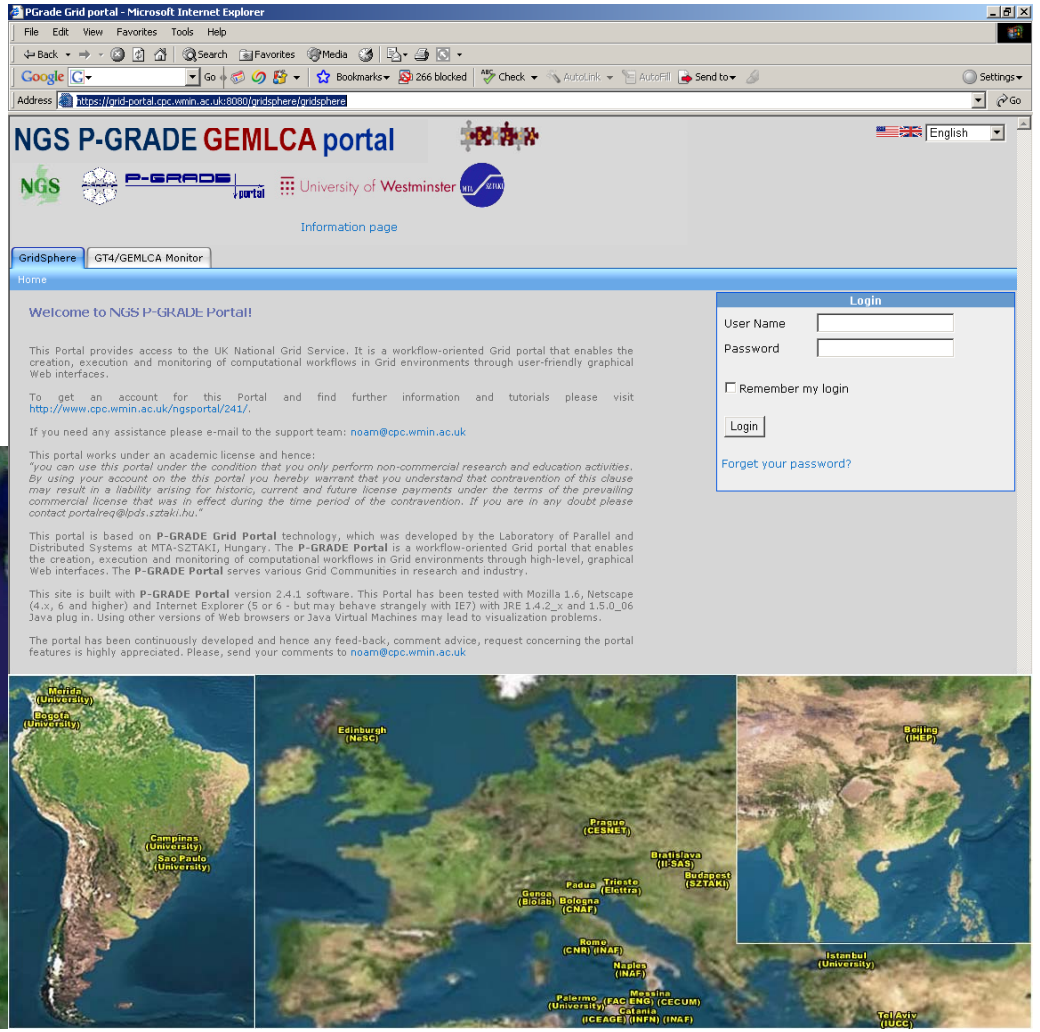
Legacy Code Interface Description File: config.xml

```
<?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 GEMLCA are open source

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

- **GEMLCA**
  - **http://dev.globus.org/wiki/Incubator/GEMLCA**
  - **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/GEMLCA portal will now follow.