Exposing Legacy Applications through a Generic Grid Service Interface

P-GRADE portal

Grid Execution Management for Legacy Code Applications

Tamas Kiss
kisst@wmin.ac.uk

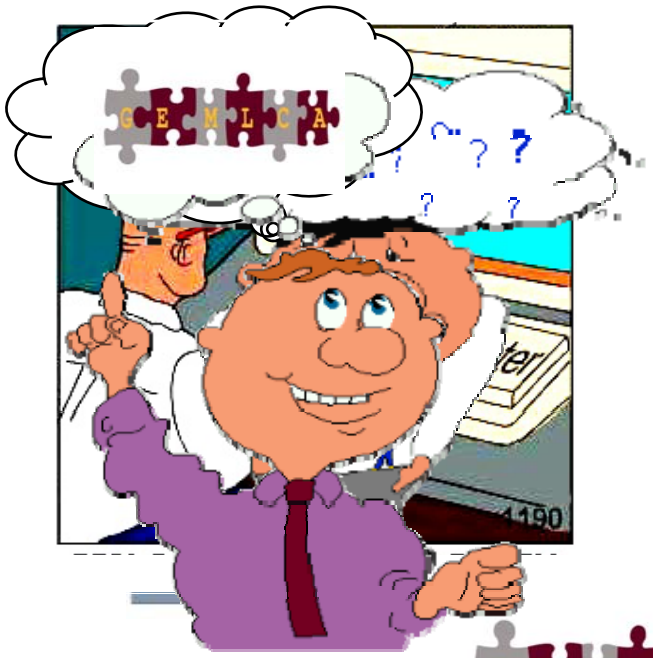
Joint EGGE/SEE-Grid Summer School
Budapest, Hungary, 28 June 2007

Legacy Applications

- Code from the past, maintained because it works
- Often supports business critical functions
- Not Grid enabled

What to do with legacy codes when utilising the Grid?

- Bin them and implement Grid enabled applications
- Reengineer them
- Port them onto the Grid with minimum user effort



Grid Execution Management for Legacy Code Applications

University of Westminster, London

GEMLCA – Grid Execution Management for Legacy Code Architecture

Objectives

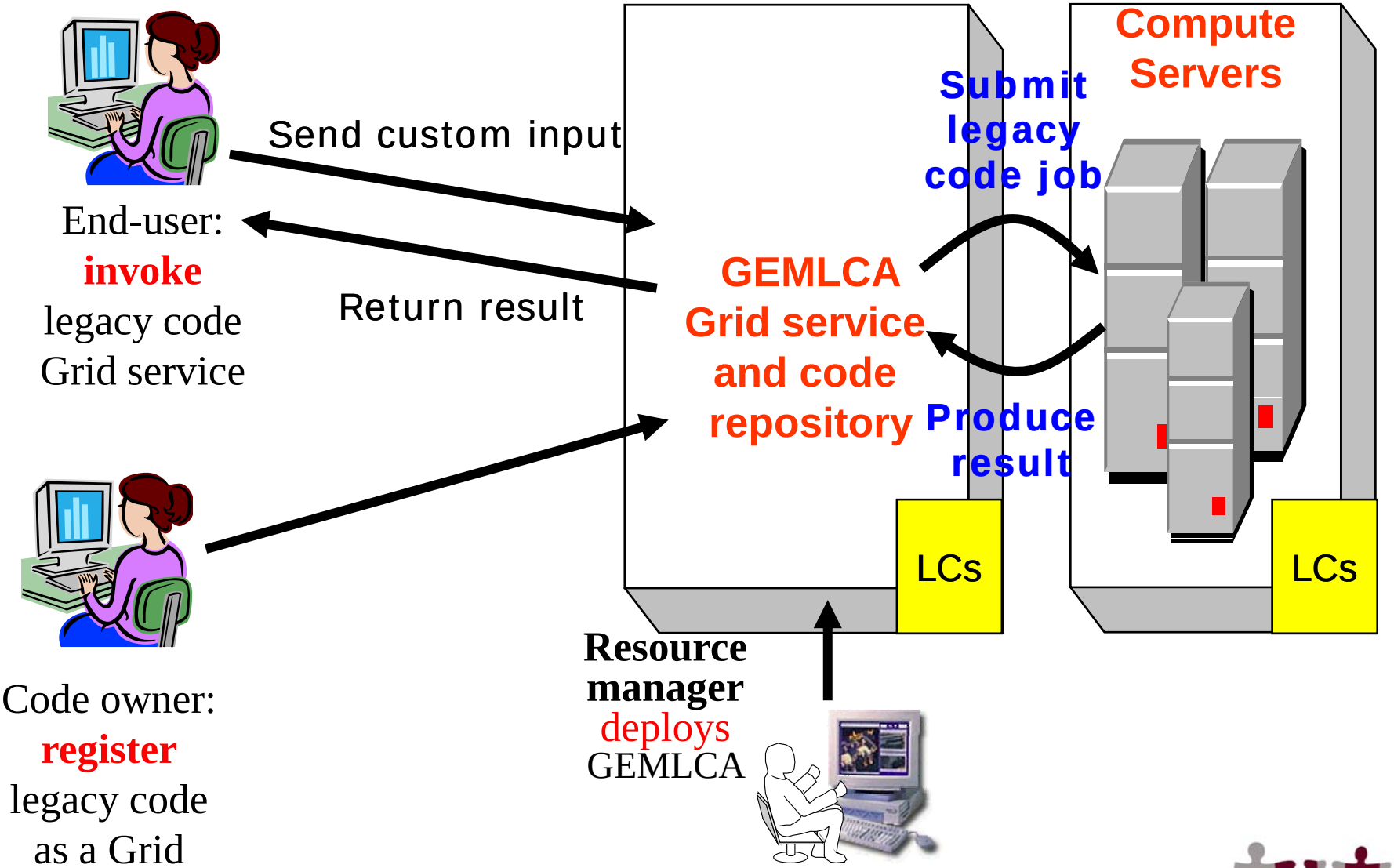
- To deploy legacy code applications as Grid services without reengineering the original code and minimal user effort

GEMLCA

- To create complex Grid workflows where components are legacy code applications
- To make these functions available from a Grid Portal

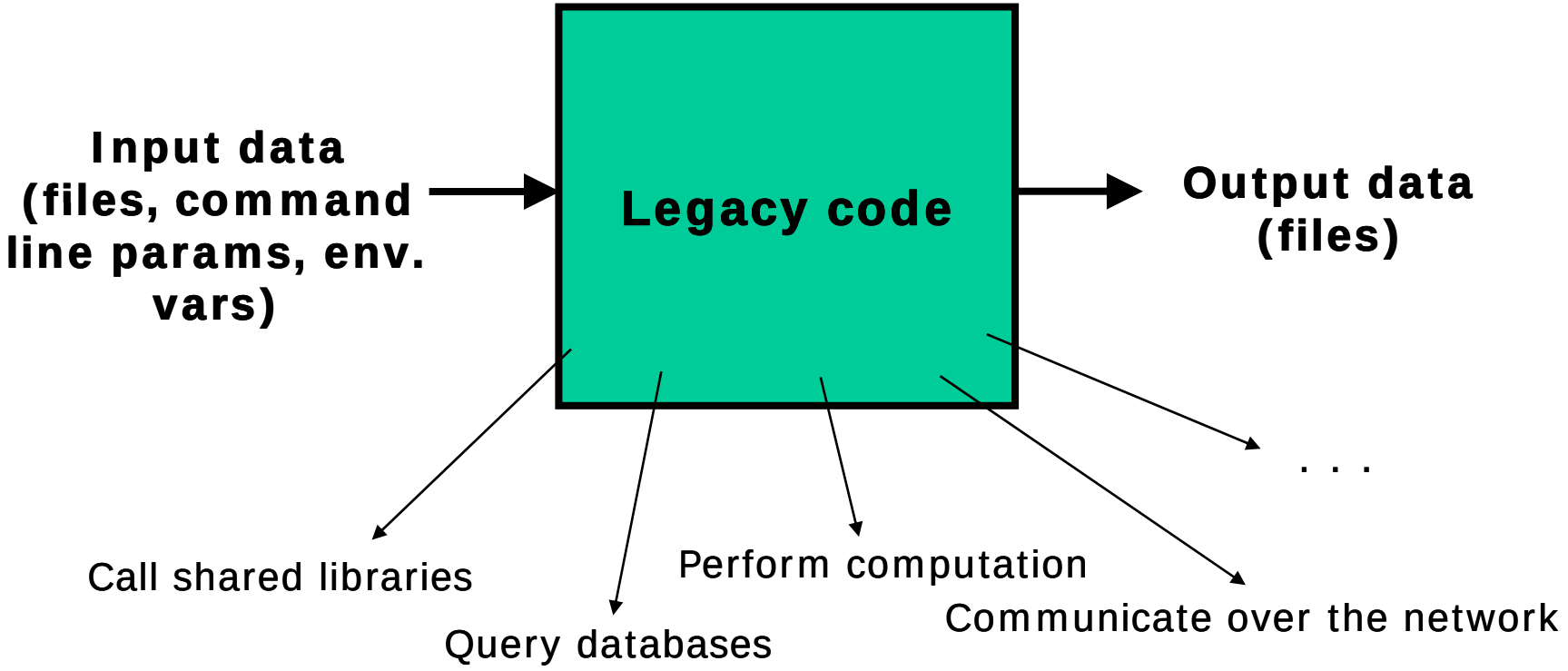
GEMLCA PGPortal Integration

GEMLCA Concept



The GEMLCA-view of a legacy code

- Any code that corresponds to the following model can be exposed as Grid service by GEMLCA:

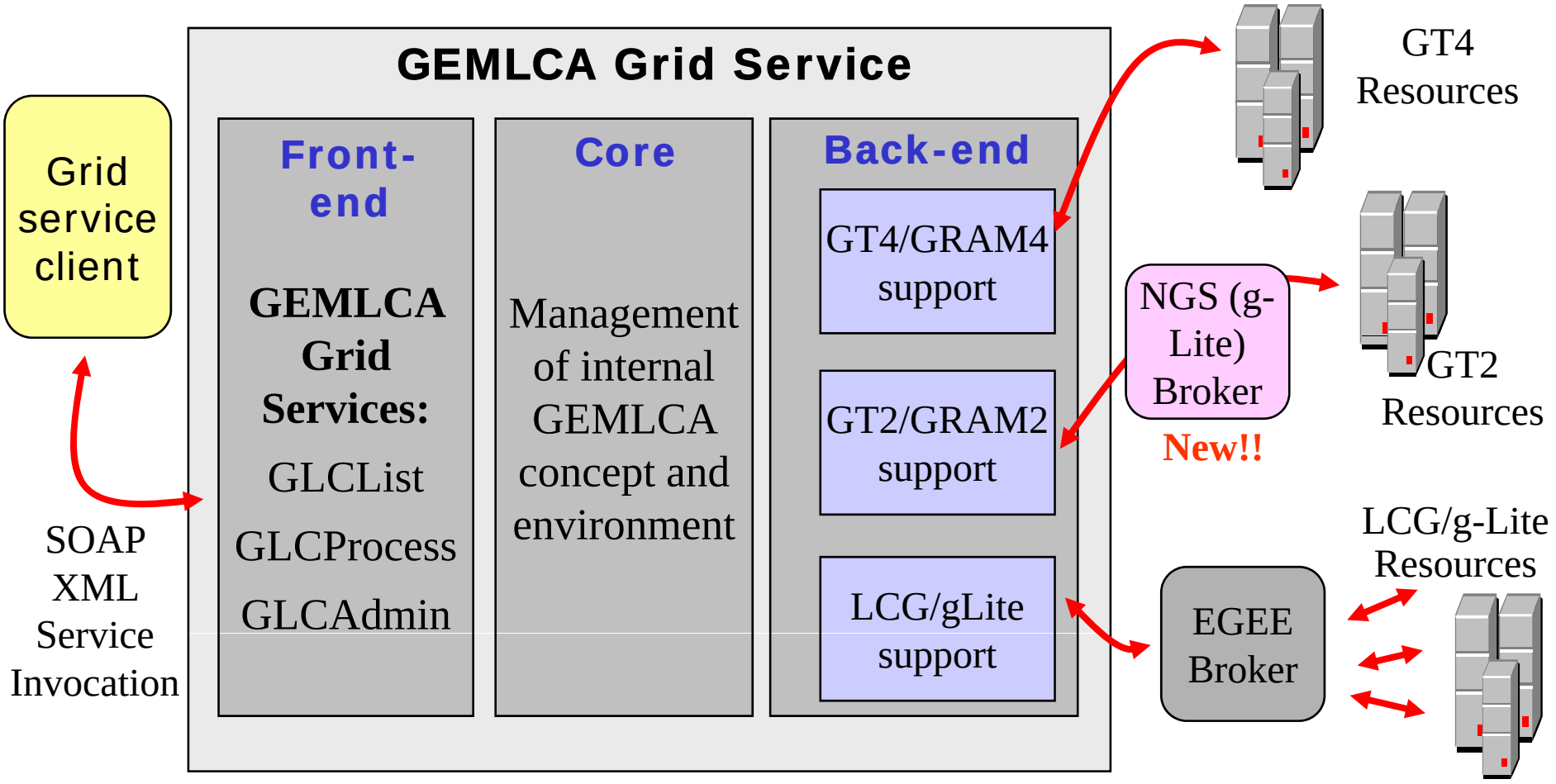


Implementing the concept

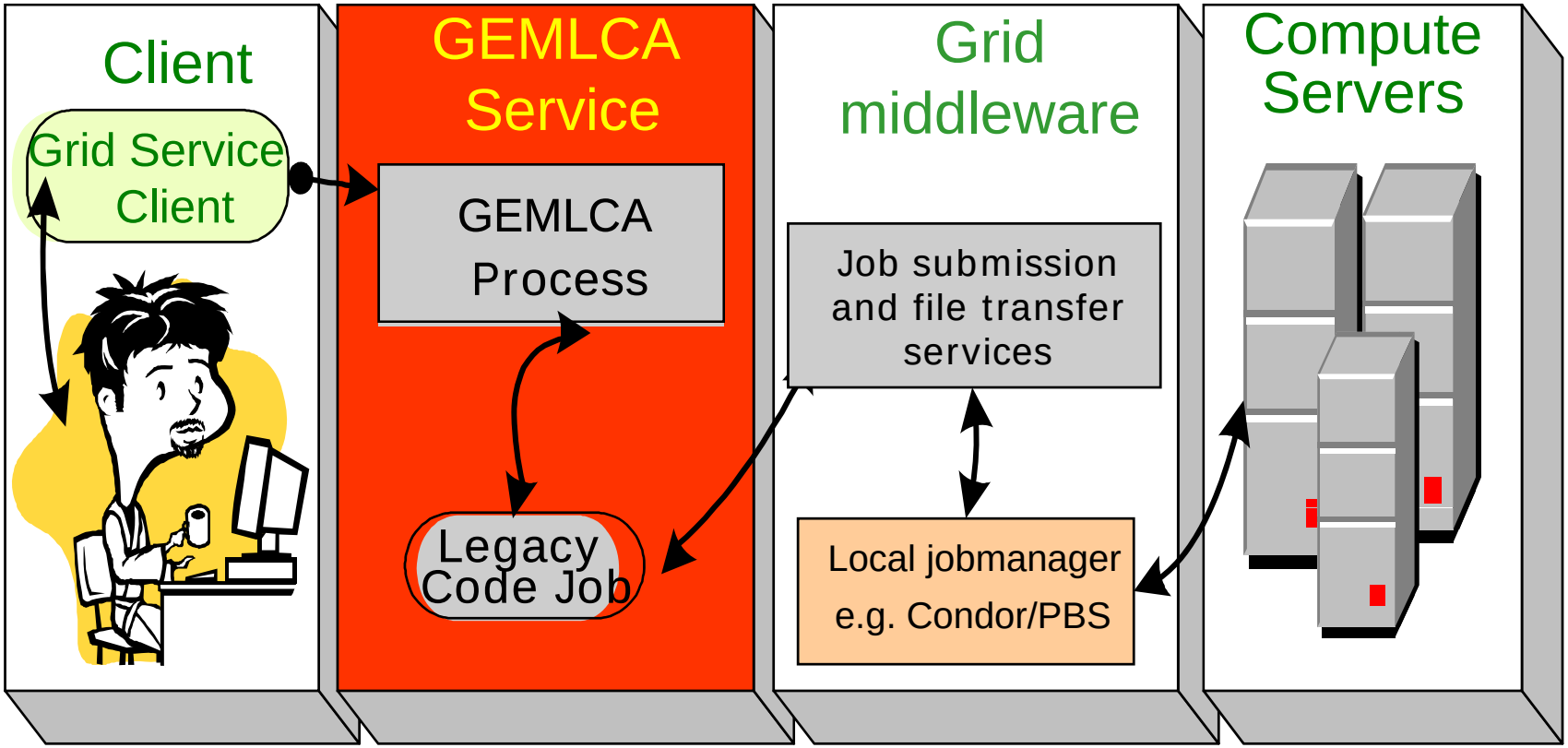
- The GEMLCA service can be implemented with any grid/service-oriented technology E.g:
  - Globus (3 or) 4 → **currently available implementations**
  - Jini
  - Web services
  - …
- GEMLCA service could invoke legacy codes in many different ways. Current implementation:
  - **Submit the legacy code as a batch job to a local job manager (e.g. Condor or PBS) through a Grid middleware layer (e.g. GT2/3/4, LCG/g-Lite)**

Implementing the GEMLCA concept

The centralised and decoupled GEMLCA architecture



How GEMLCA works?



What's the point?

- **Heterogeneous codes can be hidden behind the same interface**
  - Different programs can be invoked in the same way using SOAP/XML communication only
- **Extend non grid-aware programs with security infrastructure** (access enabled through a Grid service)
- **Create and browse repositories of legacy applications**
  - Share your codes with your colleagues or partner institutes
  - Expose business logic to your employees or customers
- **Build customized GEMLCA clients** (such as the GEMLCA P-GRADE Portal)
  - Compose complex processes by connecting multiple legacy code grid services together

The GEMLCA P-GRADE Portal

A Web-based GEMLCA client environment...

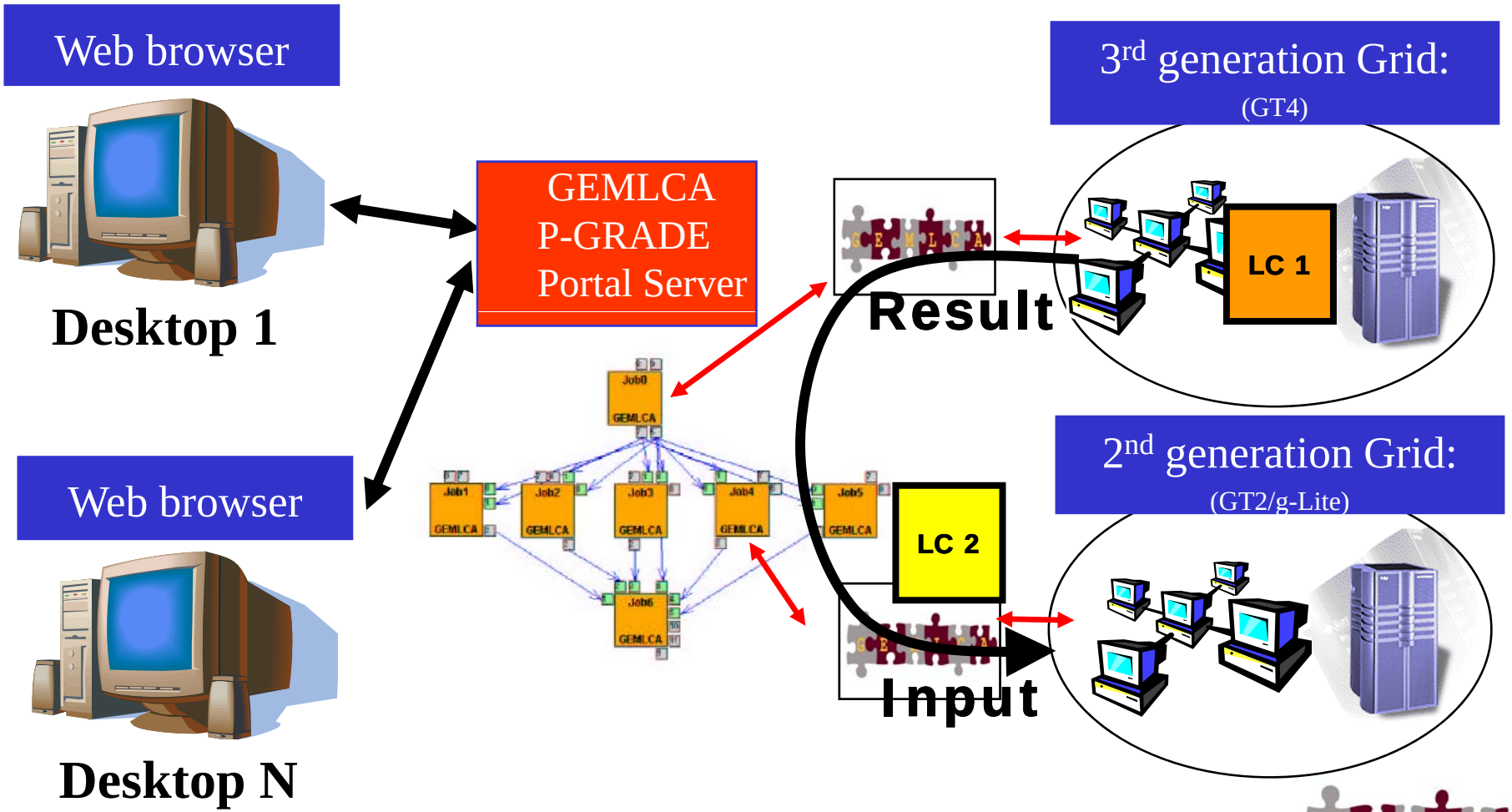
P-GRADE portal

University of Westminster, London
MTA SZTAKI, Budapest

The aims of the GEMLCA P-GRADE Portal

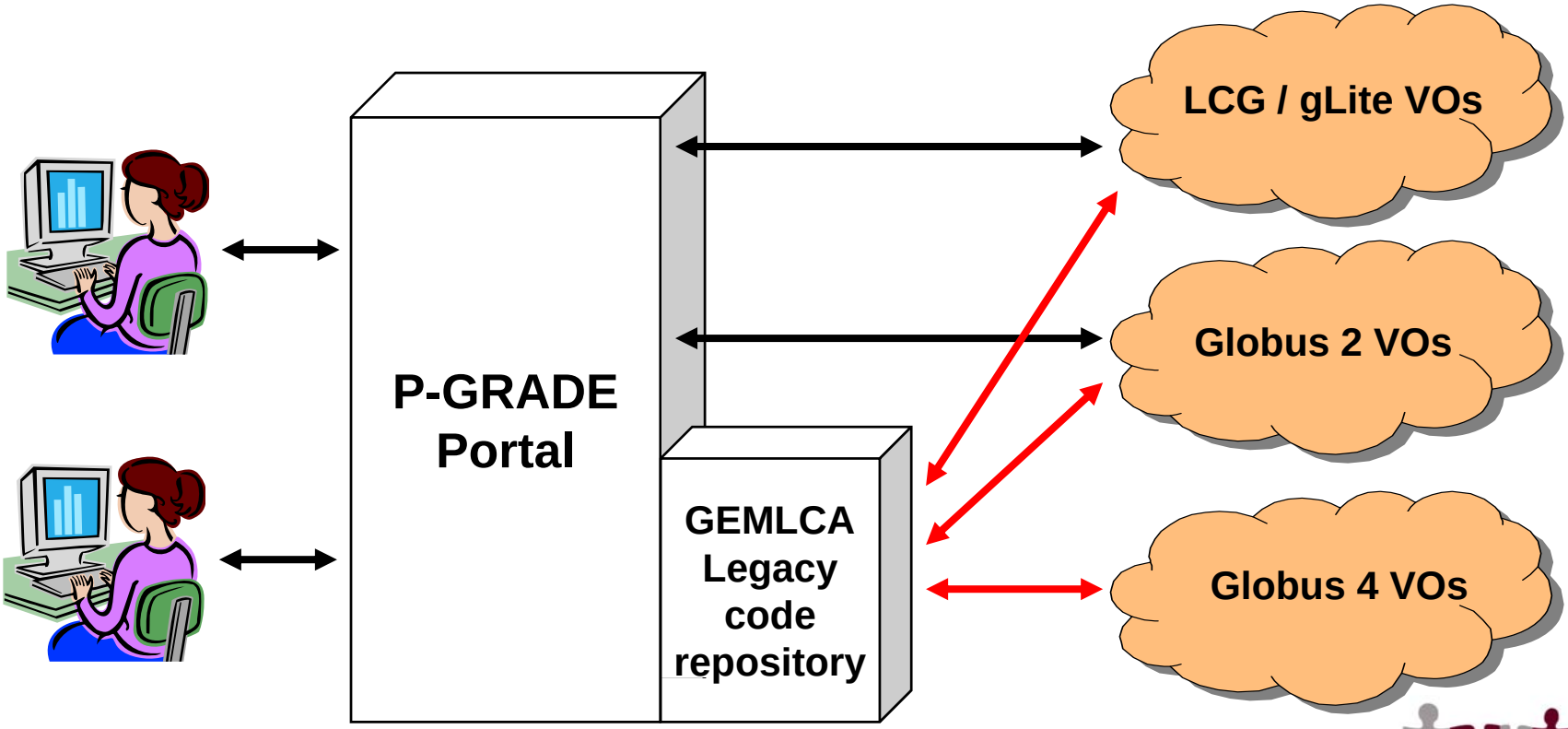
- To provide graphical clients to GEMLCA with a portal-based solution
- To enable the integration of legacy code grid services into workflows

GEMLCA in the P-GRADE Portal



How GEMLCA extends the P-GRADE Portal

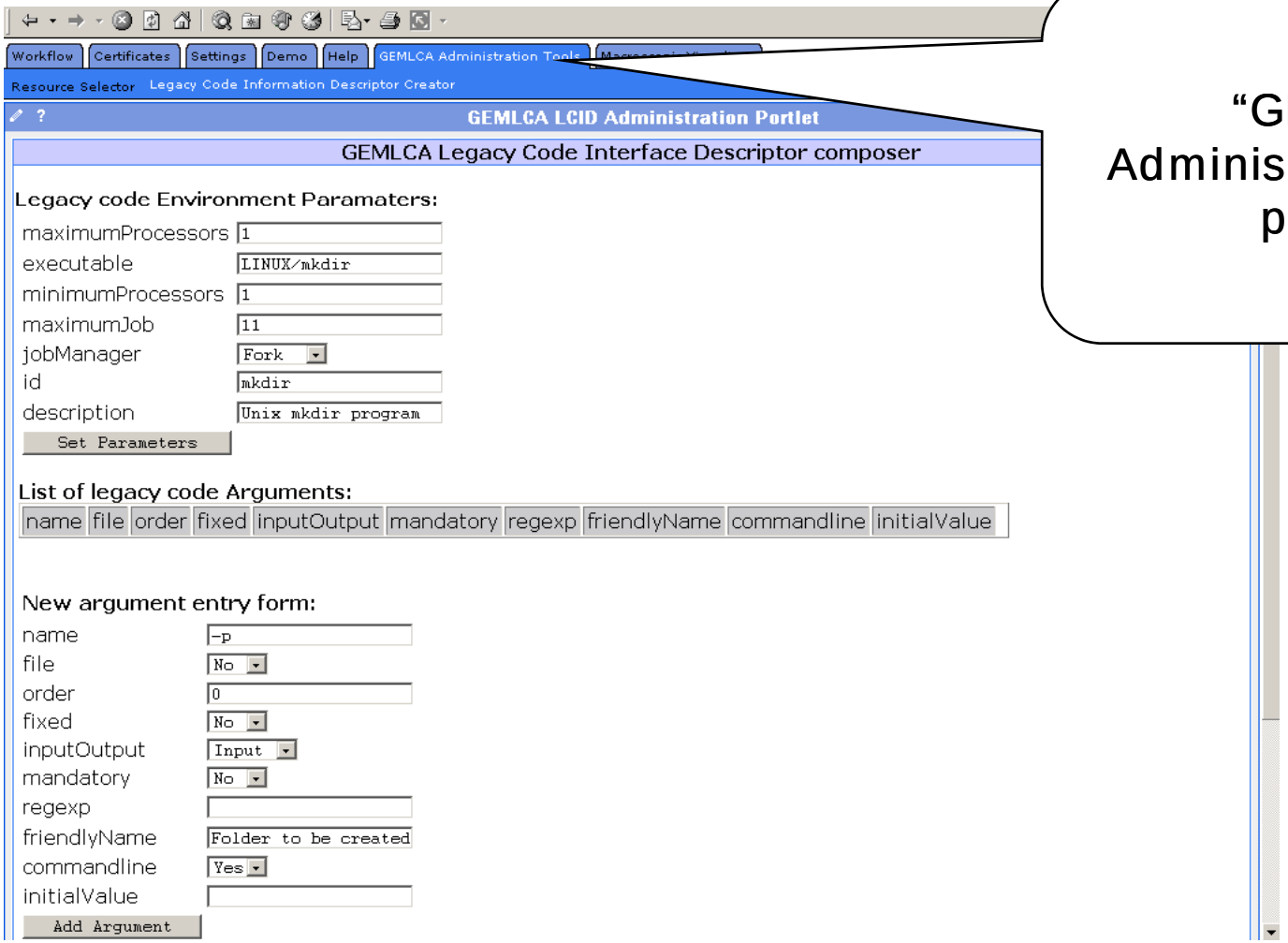
- P-GRADE Portal extended with GEMLCA back-end
  - Sharing jobs and legacy codes as workflow components
- A step towards collaborative e-Science
- Support for Globus 4 grids (besides GT2 and EGEE)



The GEMLCA-specific version of the P-GRADE Portal is different from the original P-GRADE Portal!

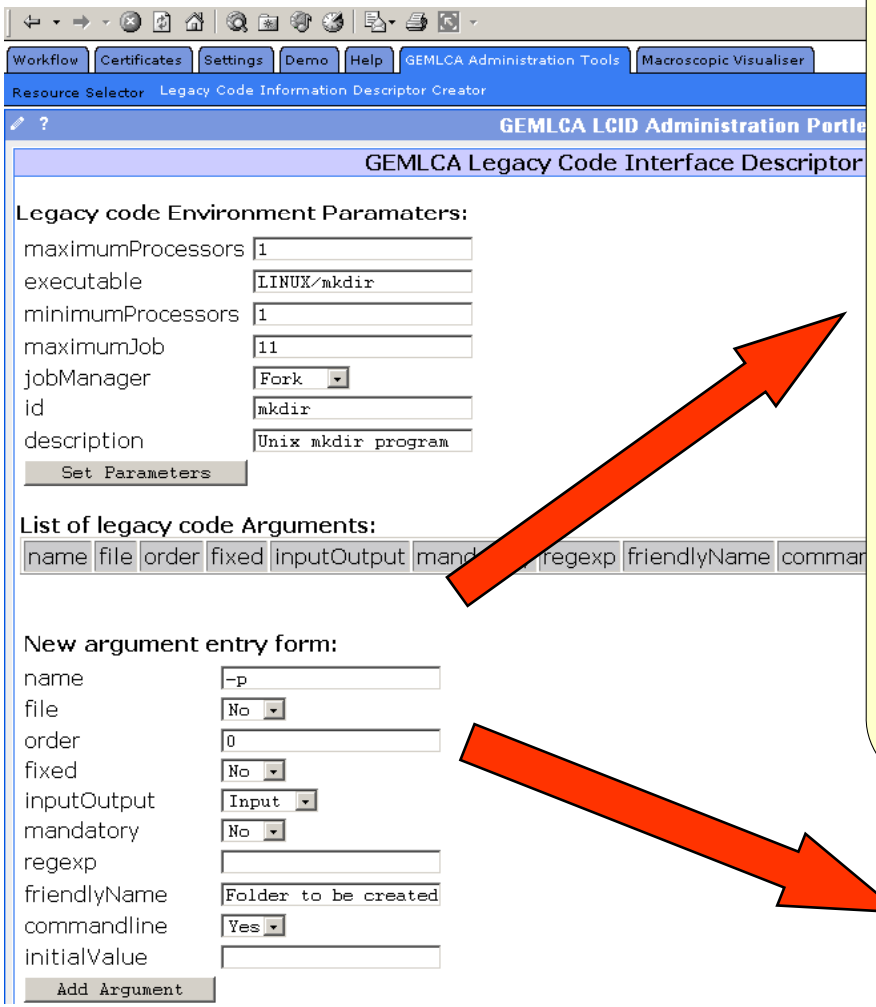
- It contains a web page to register legacy codes as grid services
- It contains a GEMLCA-specific workflow editor
  - Workflow components can be “legacy code grid services” (not only batch jobs)
- It contains a GEMLCA-specific workflow manager subsystem
  - It can invoke GEMLCA services (not only submitting jobs)

Legacy code registration page



"GEMLCA Administration Tool" portlet

Legacy code registration page



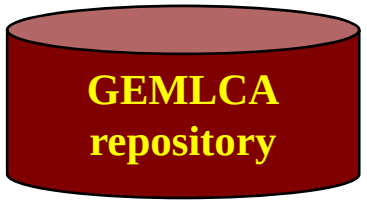
Mkdir Legacy Code exposed as a Grid Service

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

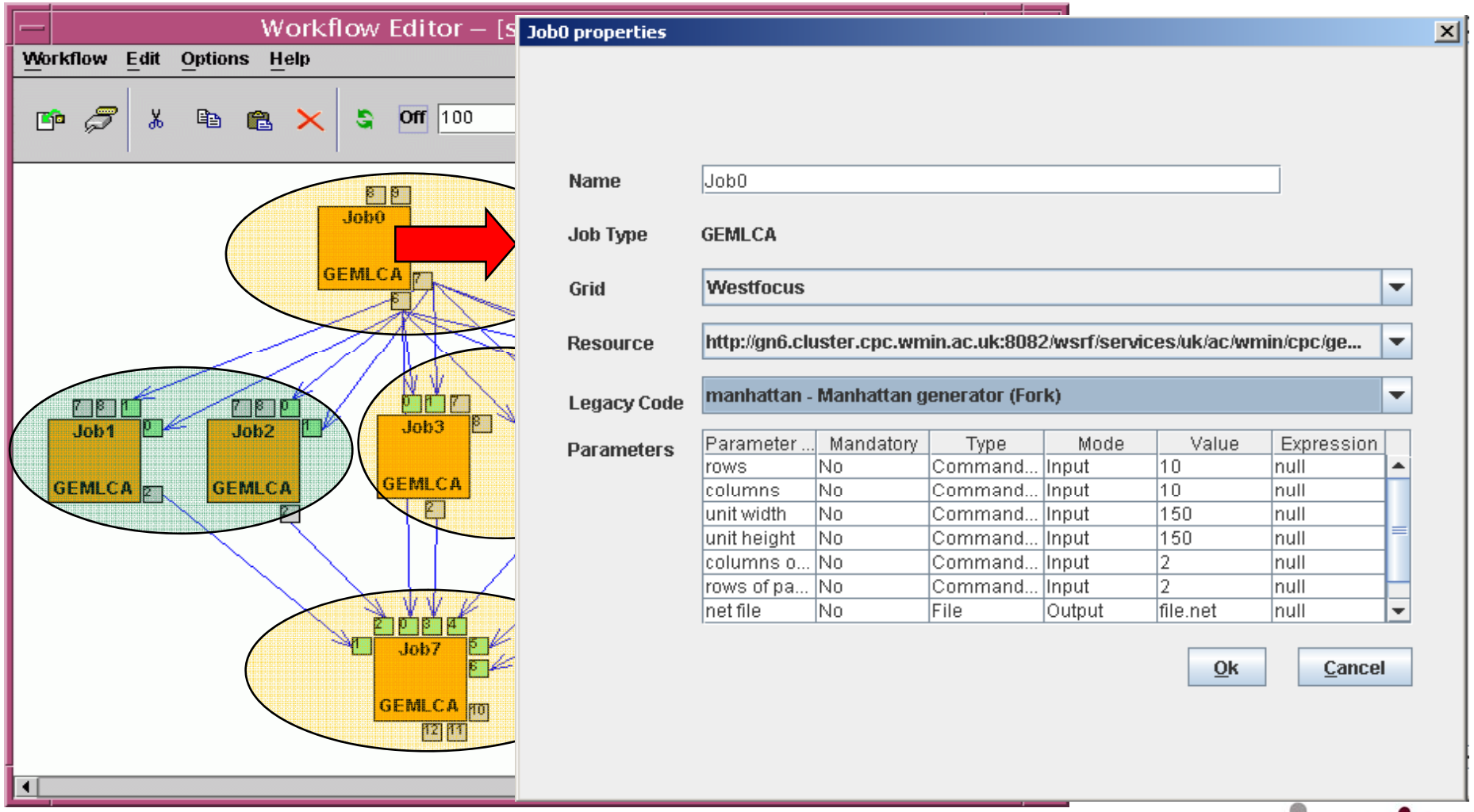
Content : i) mkdir binary or link ii) config.xml

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>
```

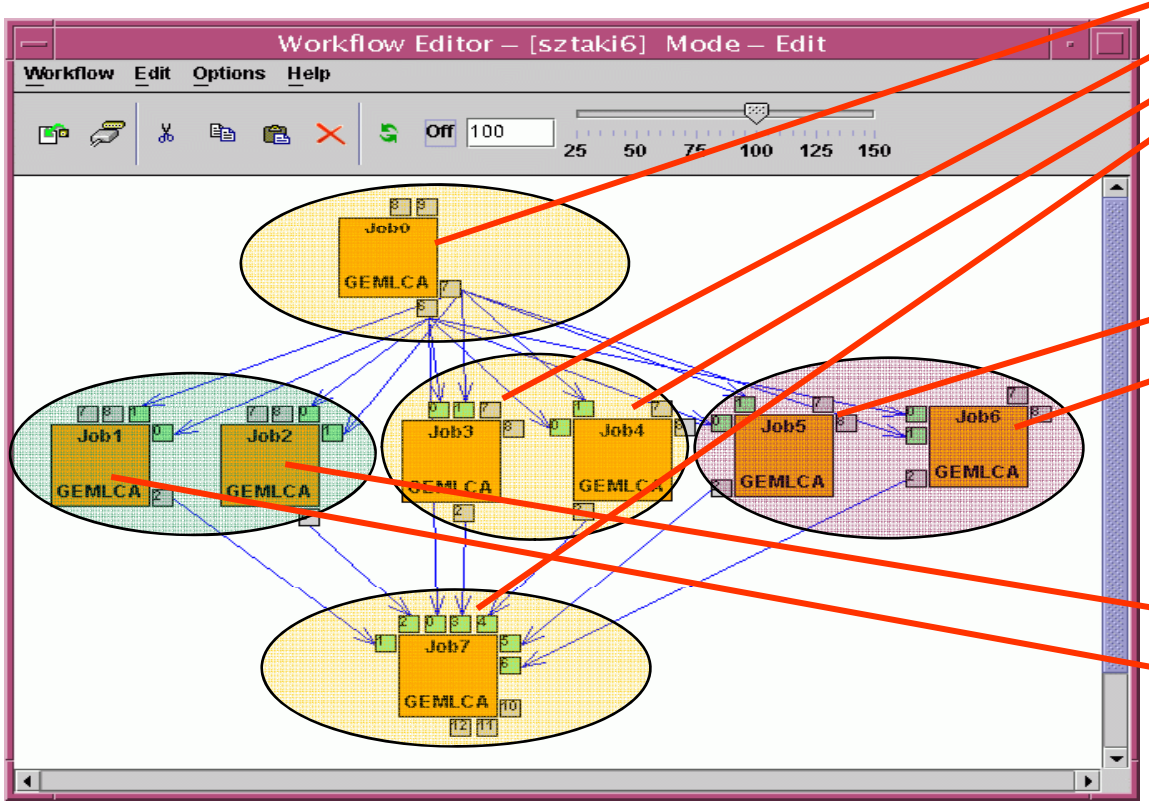


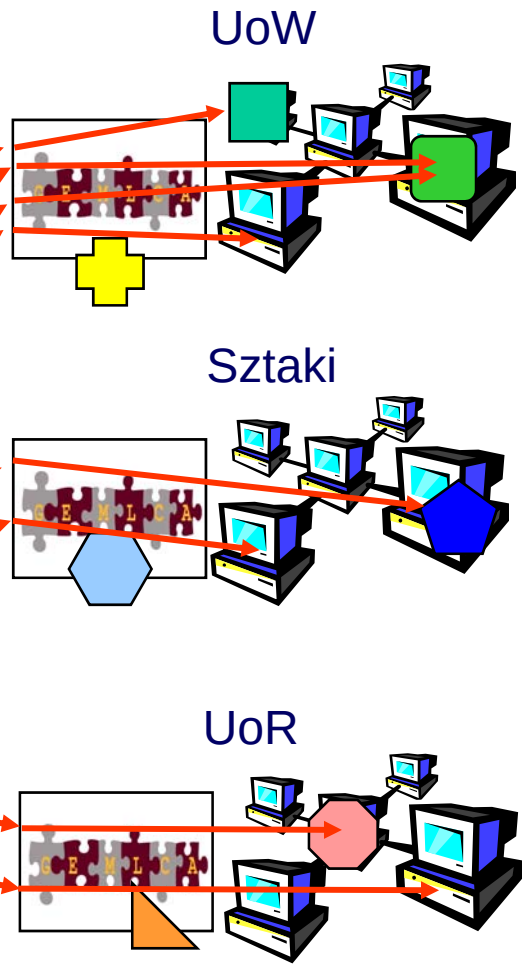
GEMLCA Specific Workflow editor



GEMLCA workflow editor in a nutshell

Workflow Creation





Batch components vs. GEMLCA components in P-GRADE Portal workflows

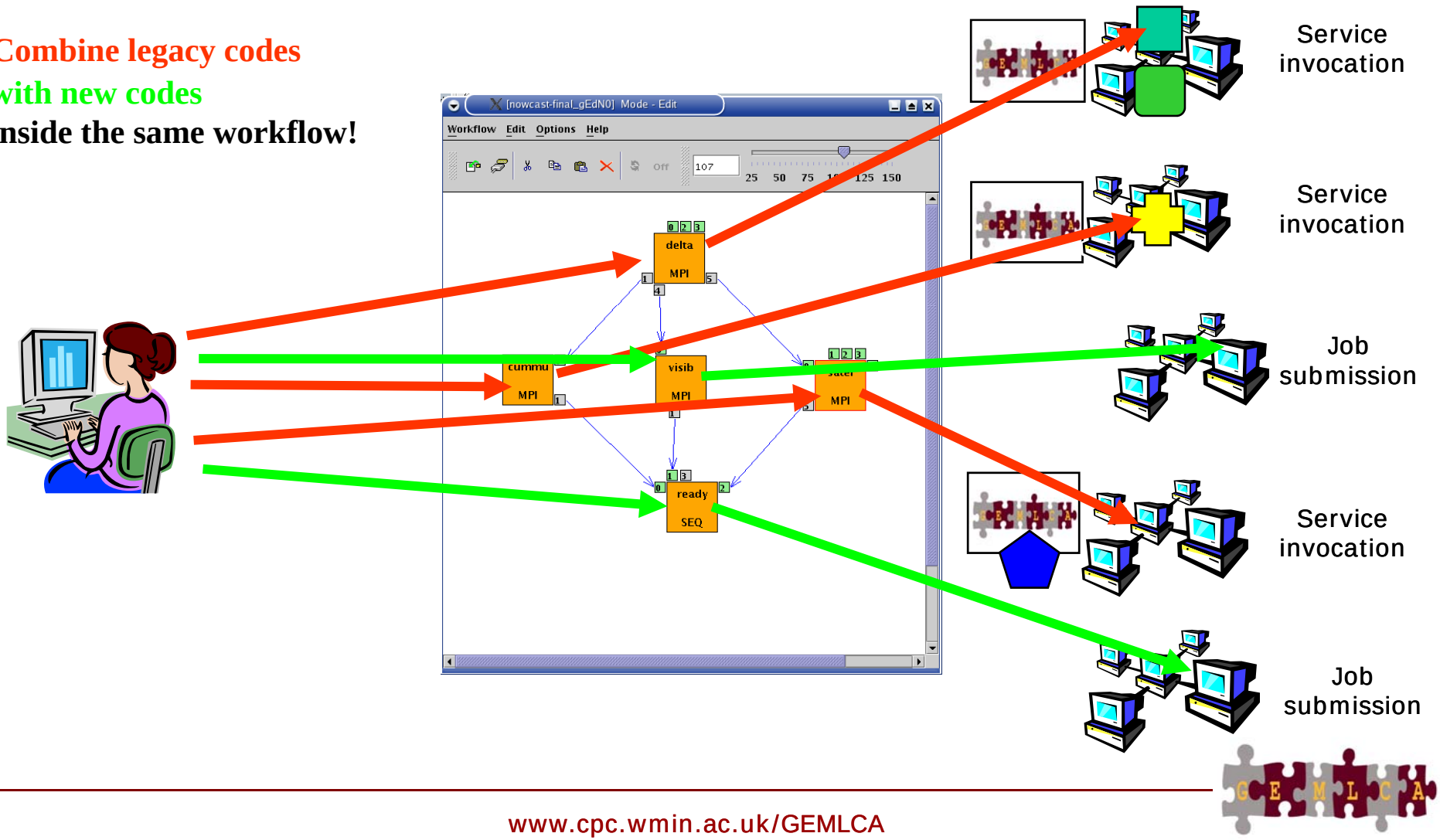
Batch component GEMLCA component

- Workflow components must be defined in different ways
- Input files represented by ports
- Output files represented by ports

Ports guarantee compatibility → batch and GEMLCA components can mutually produce data to each other!

Combining legacy and non-legacy components

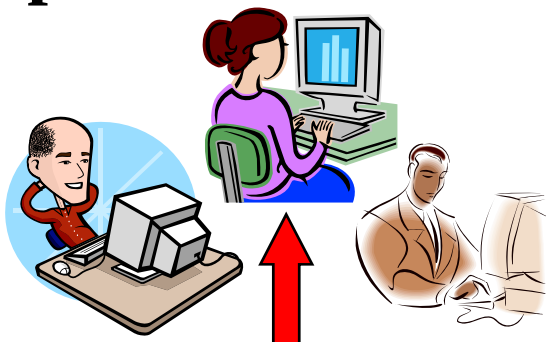
Combine legacy codes
with new codes
inside the same workflow!



The GEMLCA P-GRADE portal on production Grids

NGS National Grid Service
core production computational and data grid

Leeds
Oxford
GT2
UoW
GT2/GT4
Manchester
Rutherford



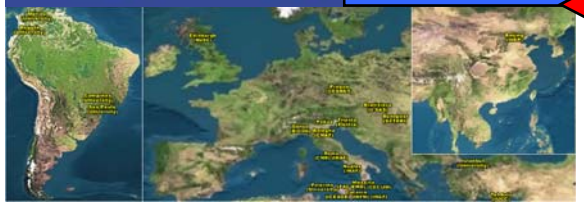
User-friendly
Web interface

P-GRADE portal

Portal server (at UoW)

eGee Enabling Grids for E-sciencE

g-Lite resources



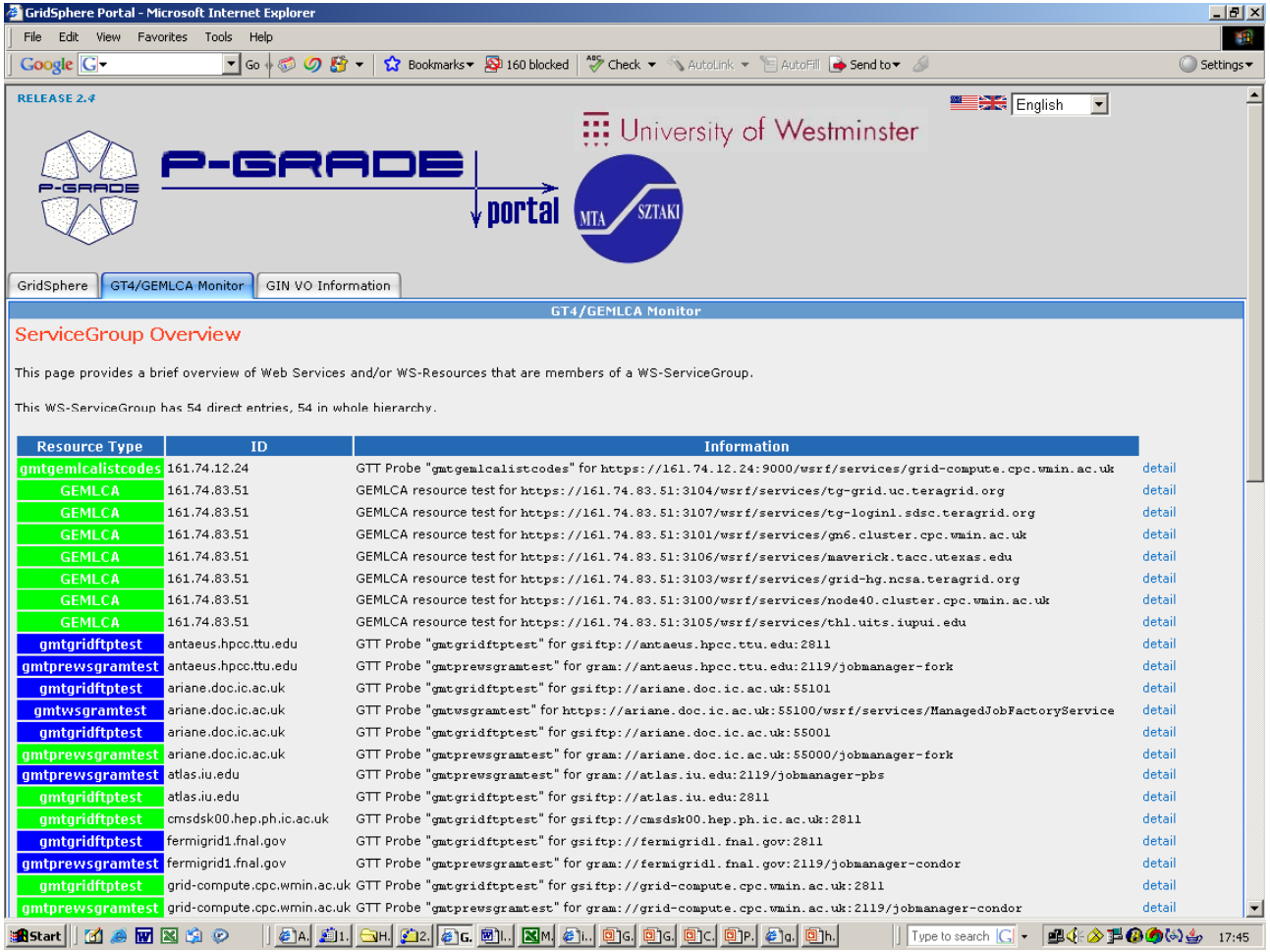
GEMLCA Resources

Run as third party services
Decoupled from portal and Grid middleware

GMT – GEMLCA Monitoring Toolkit

Resources need to be constantly monitored in production environments

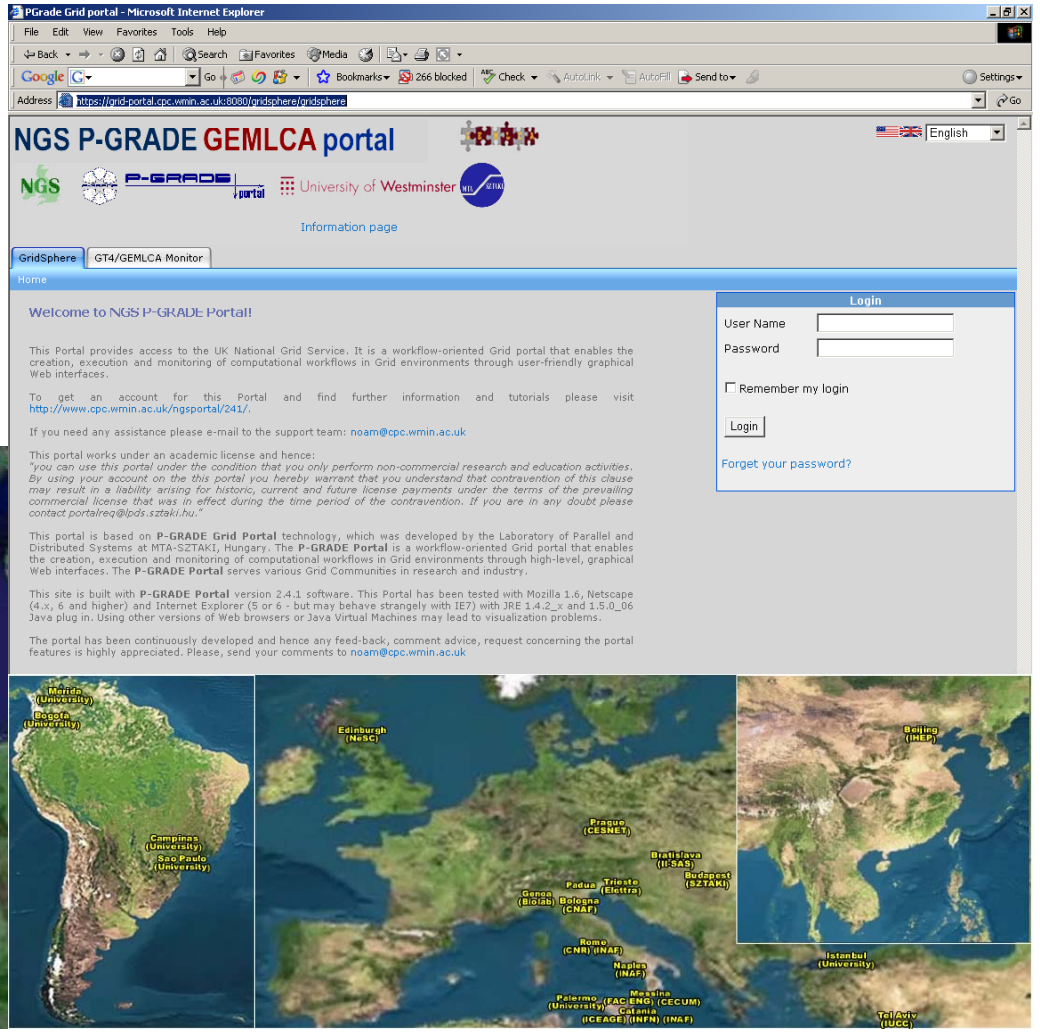
- to test resource availability
- implementation is based on MDS4
- probes are implemented as scripts and their outputs are displayed in a monitoring portlet
- Runs on the NGS and GIN portals



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

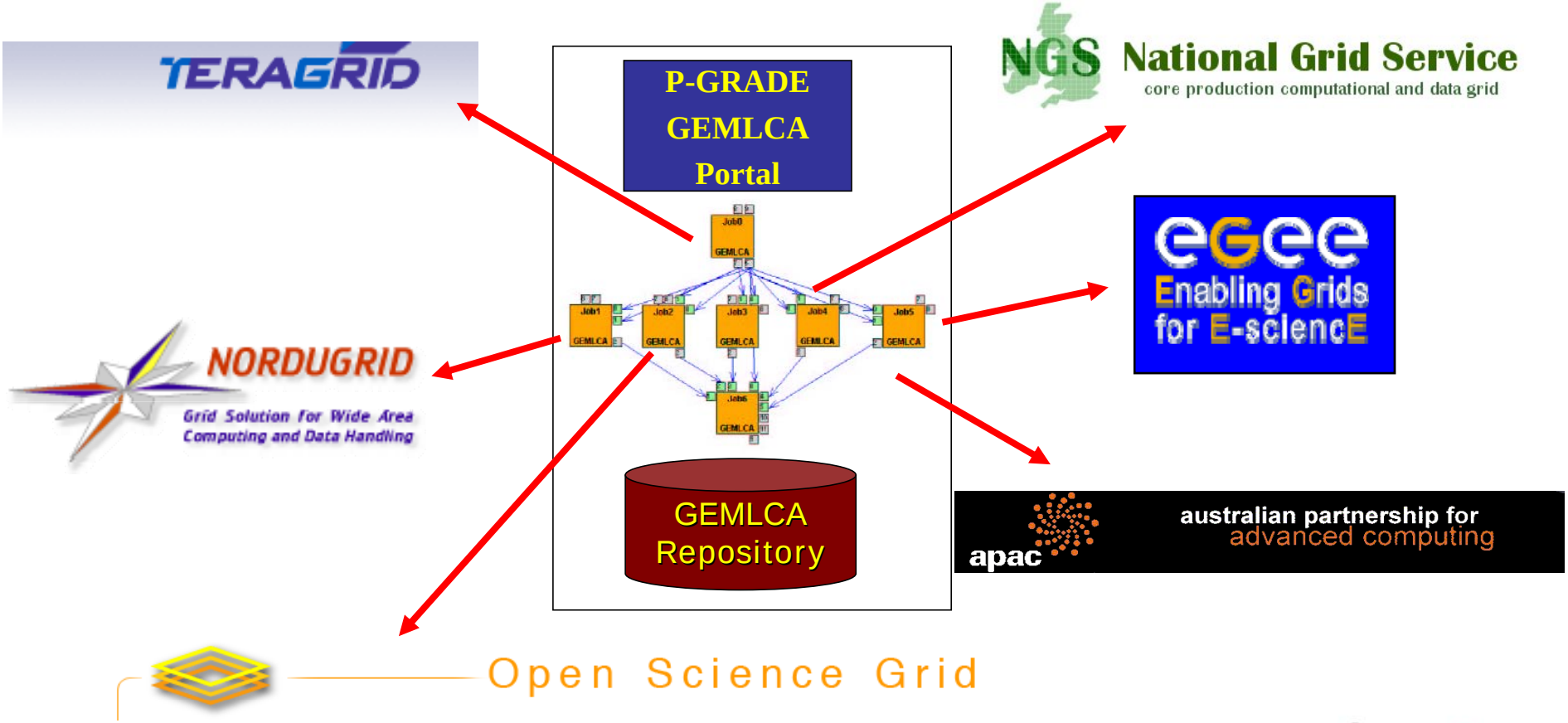
NGS P-GRADE GEMLCA Portal

- **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

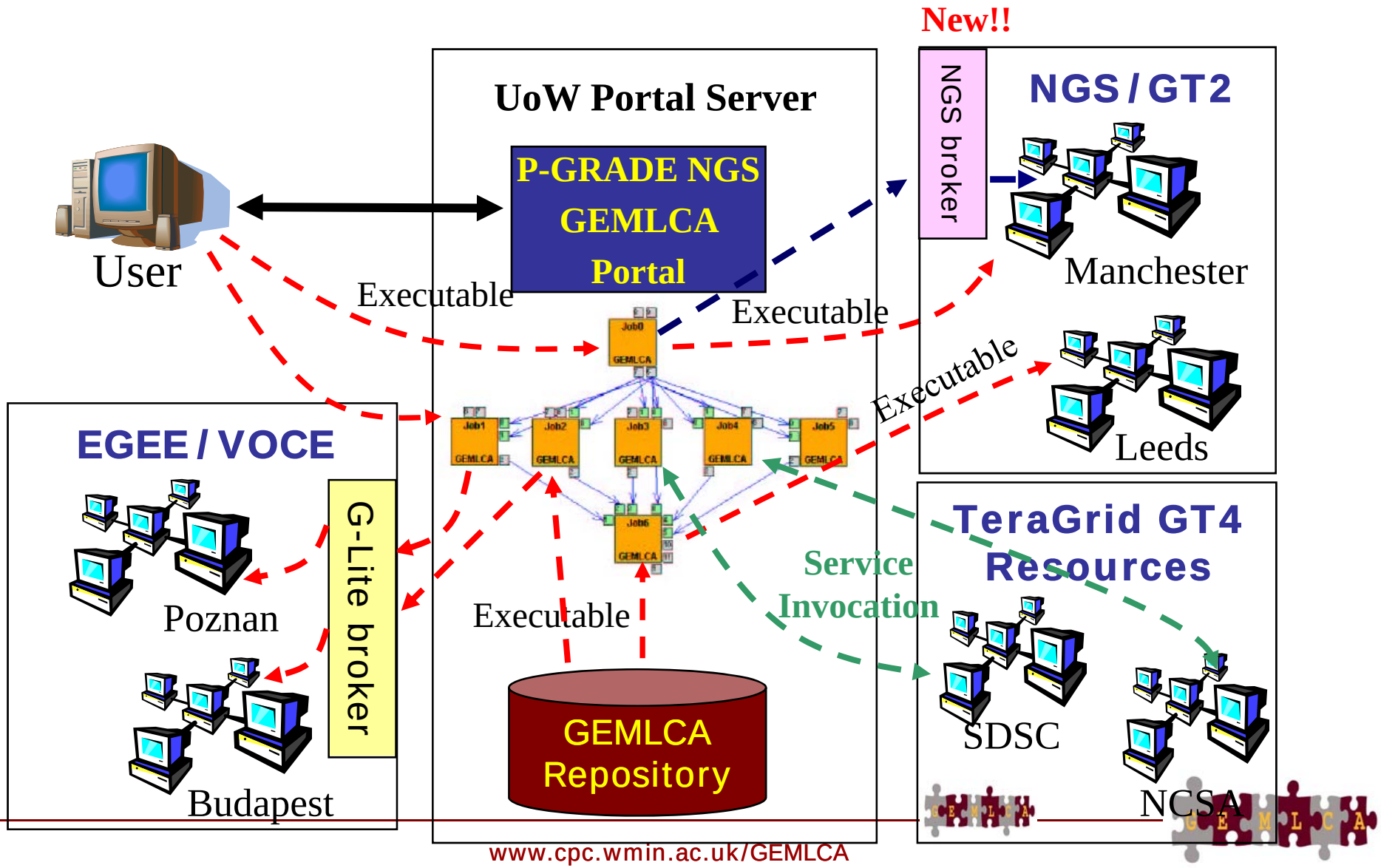


The GIN Resource Testing portal

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



Workflow level interoperability of Grid systems



Traffic simulation on multiple Grids

GT2 Job submission to NGS (Oxford)

LCG Job submission to EGEE/GIN (IC)

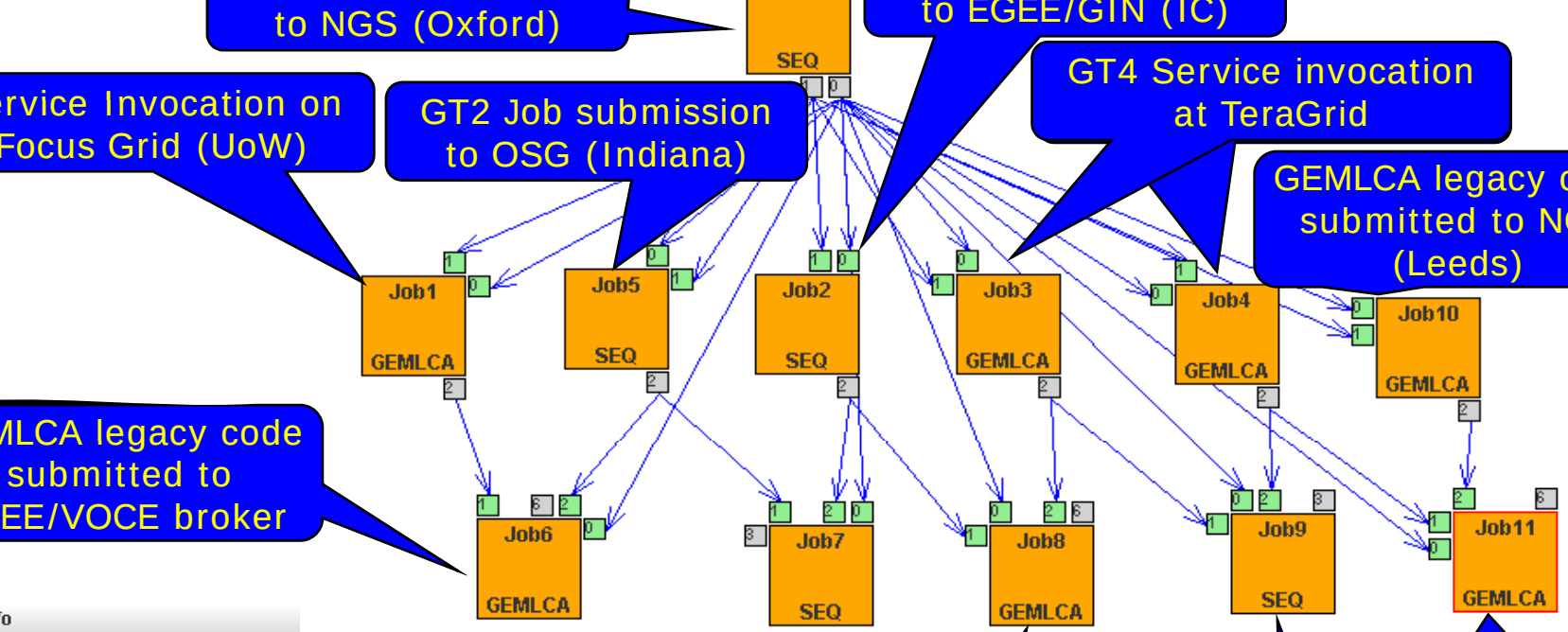
GT4 Service Invocation on WestFocus Grid (UoW)

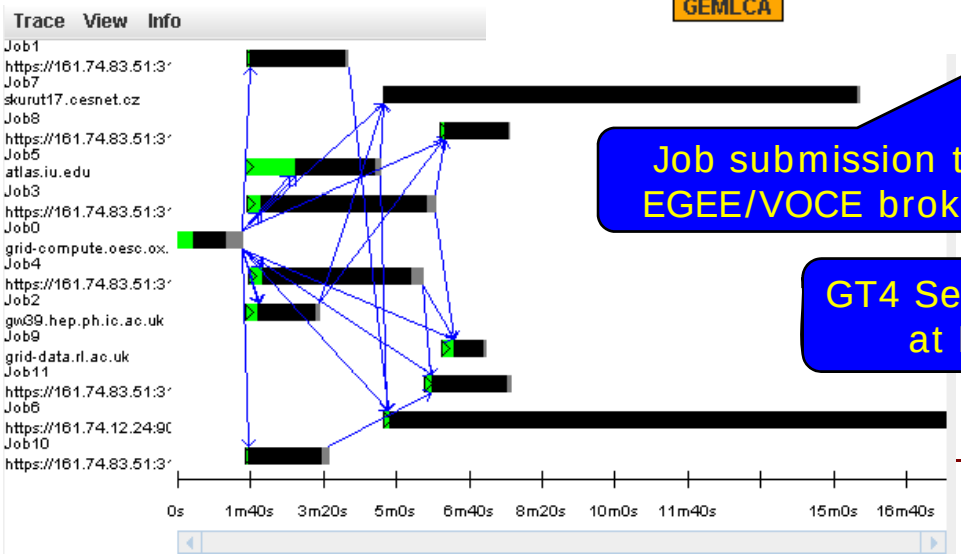
GT2 Job submission to OSG (Indiana)

GT4 Service invocation at TeraGrid

GEMLCA legacy code submitted to NGS (Leeds)

GEMLCA legacy code submitted to EGEE/VOCE broker





Job submission to EGEE/VOCE broker

GT4 Service invocation at NGS (UoW)

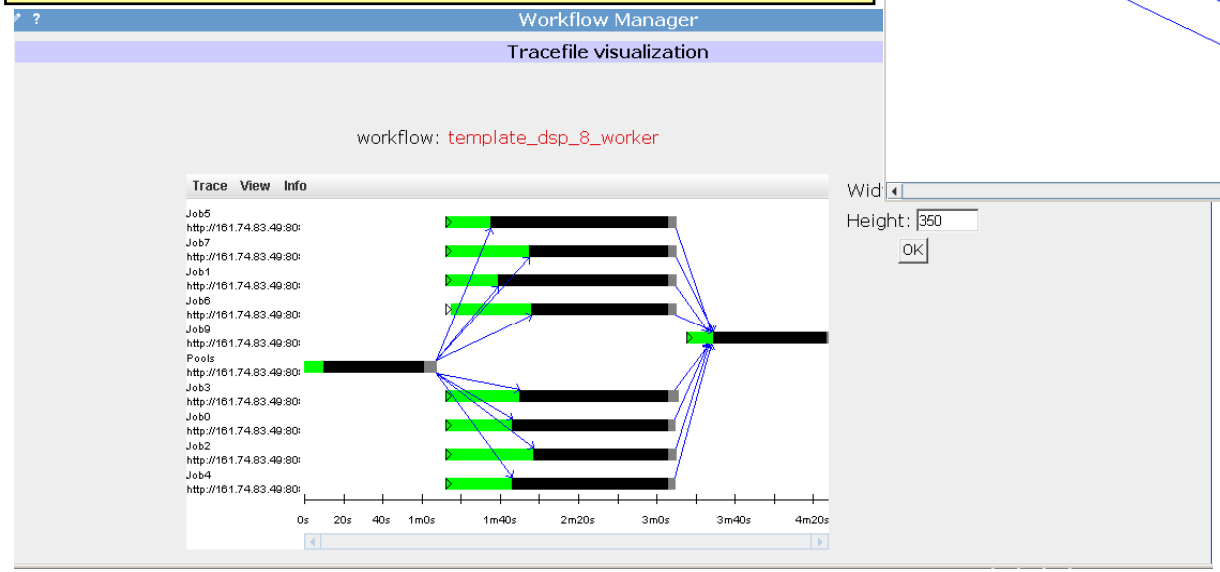
GT4 Service Invocation on WestFocus Grid (UoW)

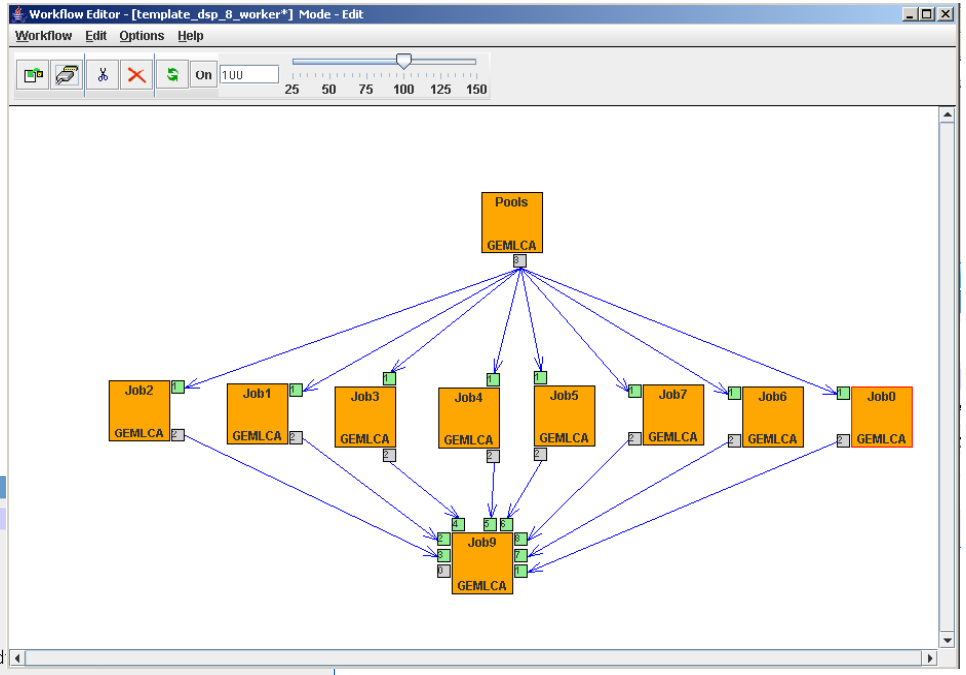
GT2 Job submission to NGS (Rutherford)

Application examples

DSP-Designing Optimal Periodic Nonuniform Sampling Sequences

T Factor	Sequential	GEMLCA
18	~19min	~8min
20	~3h 33min	35min
22	~41h 53min	~7h 23min

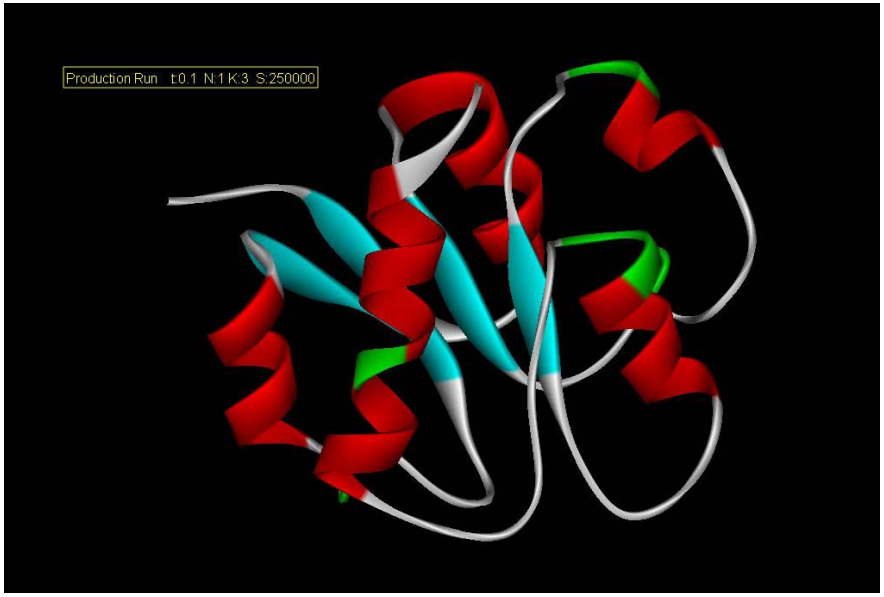


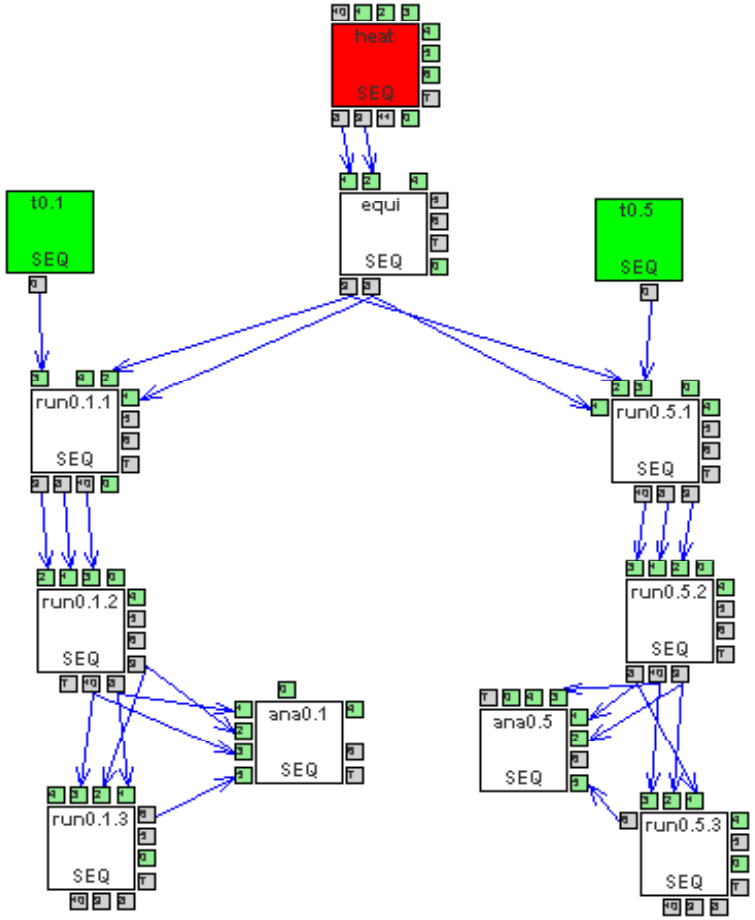


Application examples

Molecular Dynamics Study of Water Penetration in Staphylococcal Nuclease using CHARMm

- Analysis of several production runs with different parameters following a common heating and equilibrium phase

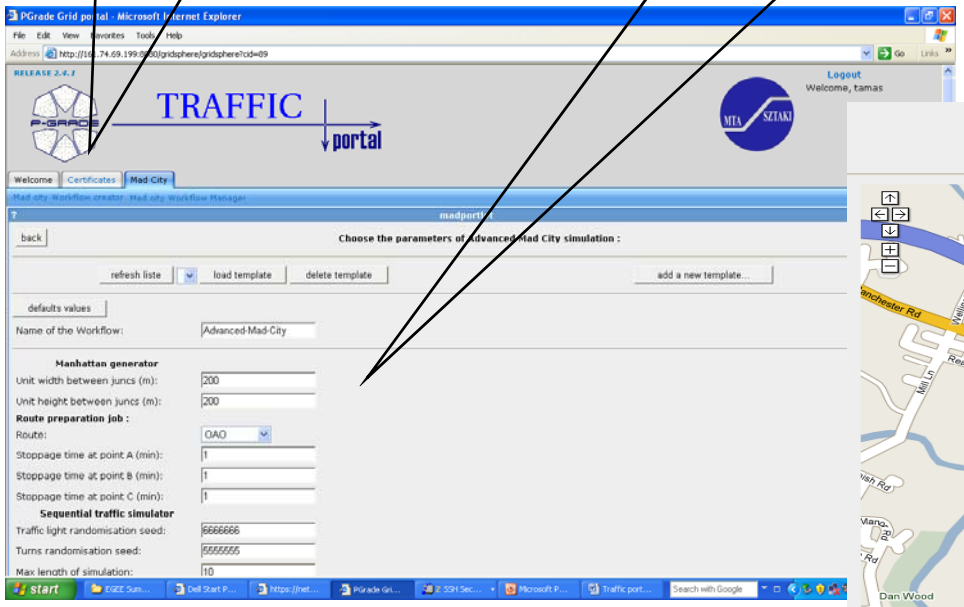


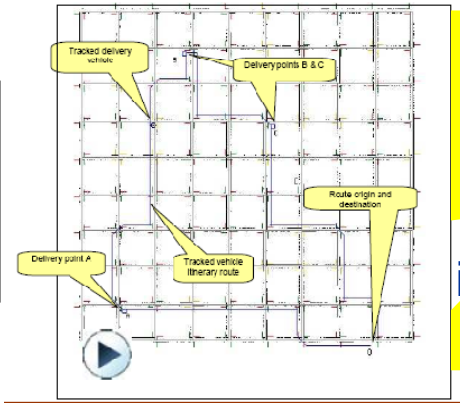


Application Specific Portals - The Traffic Portal

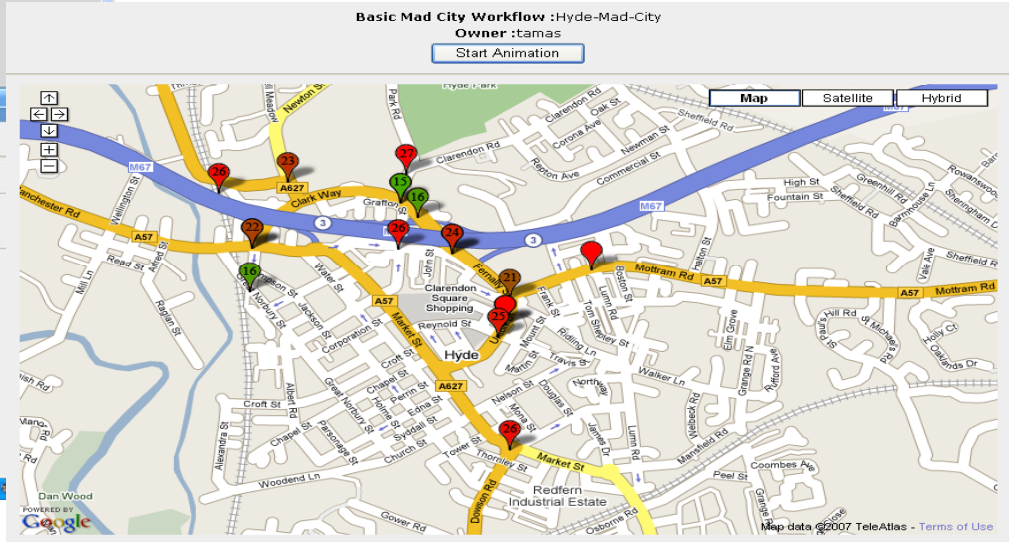
Only essential/related portlets displayed

Workflow creation and structure is hidden from the user





Application specific visualisation is embedded



Specifically tailored for one specific application - transport and logistics studies with a traffic simulator

Conclusions

- GEMLCA enables the deployment of legacy code applications as Grid services without any real user effort.
- GEMLCA is integrated with the P-GRADE portal to offer user-friendly development and execution environment.
- The integrated GEMLCA P-GRADE solution is available for the UK NGS as a service! **www.cpc.wmin.ac.uk/ngsportal**

Thank you for your attention!

http://www.cpc.wmin.ac.uk/gemlca

gemlca-discuss@cpc.wmin.ac.uk