UNIVERSITY OF WESTMINSTER

P-GRADE Portal:
An easy to use graphical interface for Globus and EGEE Grids

Motivation to the P-GRADE Portal

- **Fast evolution of Grid middleware technologies and tools:**
  - **GT2, OGSA, GT3 (OGSI), GT4 (WSRF), LCG-2, gLite, …**
- **Many production Grids are now freely available for e-Scientists**
  - **EGEE** (LCG-2 → gLite), **UK NGS** (GT2), **US Open Science Grid** (GT2 → GT4), **NorduGrid** (ARC), …
- **The same set of services are available everywhere, but implemented in different ways**
  - Computation services, data services, security services, (brokers)

Let's provide a technology-neutral graphical interface for the most common Grid middleware services!

P-GRADE Portal in a nutshell

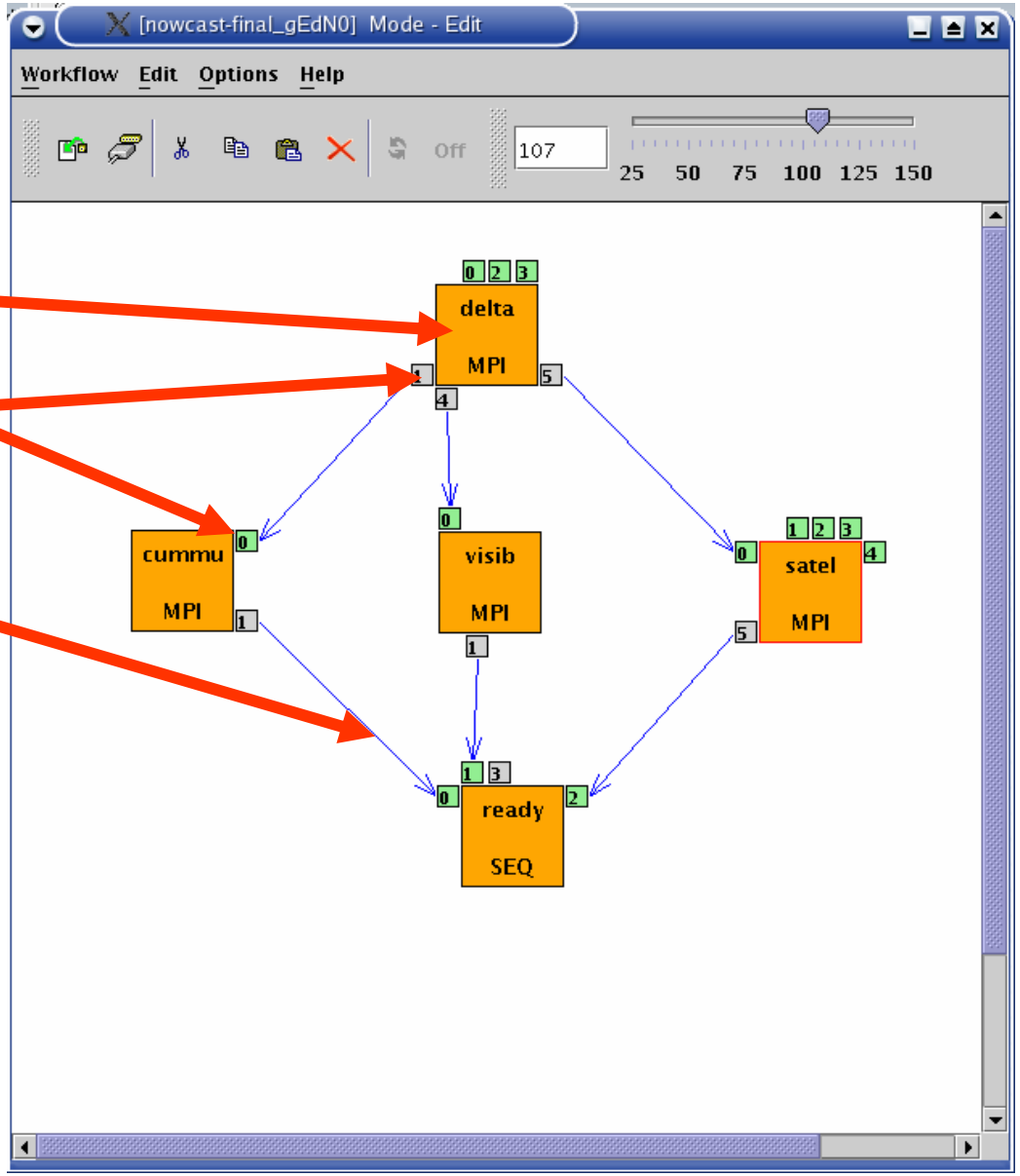
- **General purpose, workflow-oriented computational Grid portal.** Supports the development and execution of workflow-based Grid applications.
- **Based on standard portlet framework (Gridsphere)**
  - Easy to expand with new portlets (e.g. application-specific portlets)
  - Easy to tailor to community needs
- **Grid services** supported by the portal:

Service	EGEE grids	Globus grids
Job execution	Computing Element	GRAM
File storage	Storage Element	GridFTP server
Certificate management	MyProxy	
Information system	BDII	MDS-2
Brokering	Workload Management System	---
Job monitoring	Mercury	
Workflow & job visualization	PROVE	

The P-GRADE Portal hides middleware technologies and solves Grid interoperability problem at the workflow level

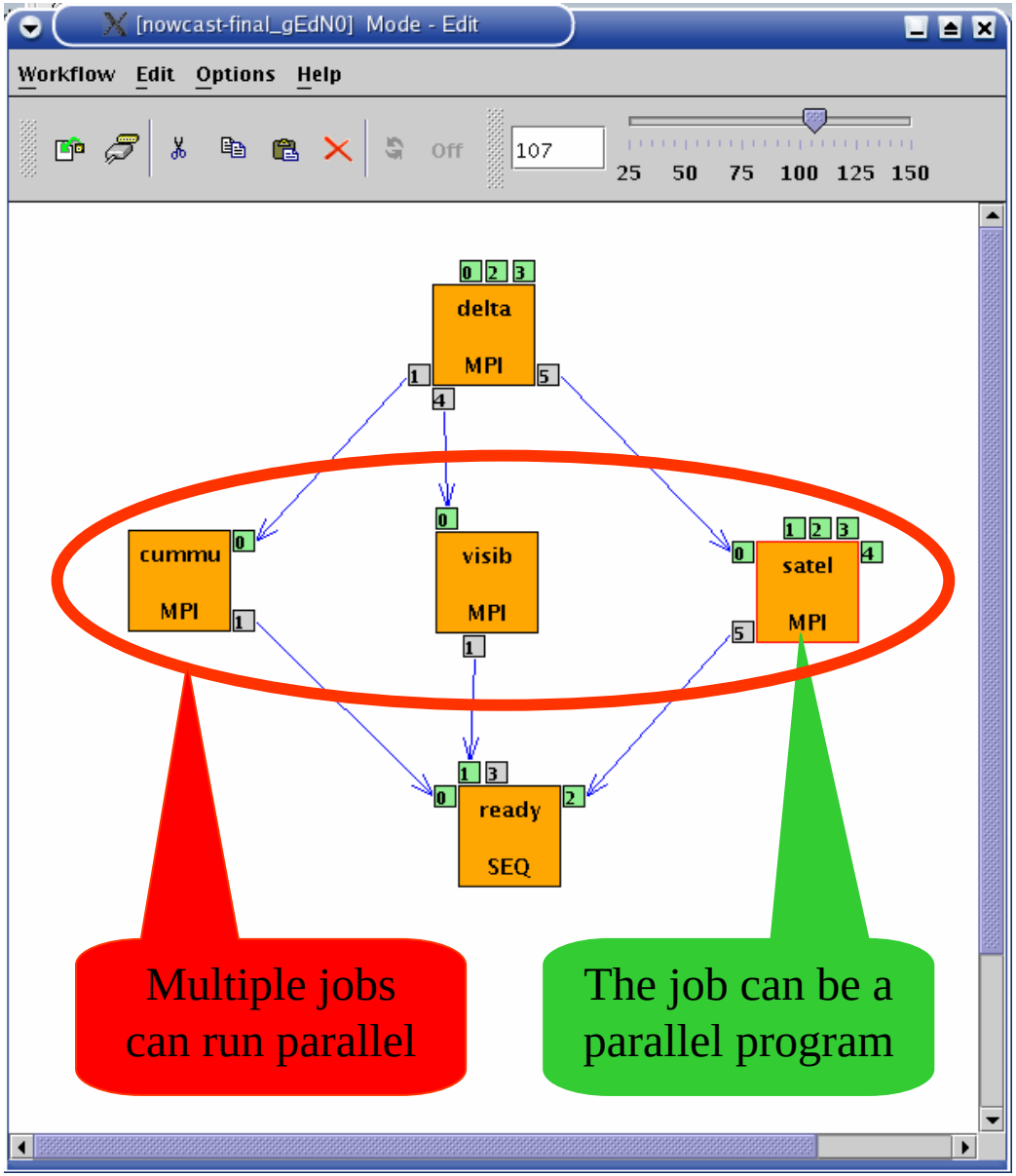
What is a P-GRADE Portal workflow?

- **a directed acyclic graph where**
  - Nodes represent jobs (batch programs to be executed on a computing element)
  - Ports represent input/output files the jobs expect/produce
  - Arcs represent file transfer operations
- **semantics of the workflow:**
  - A job can be executed if all of its input files are available

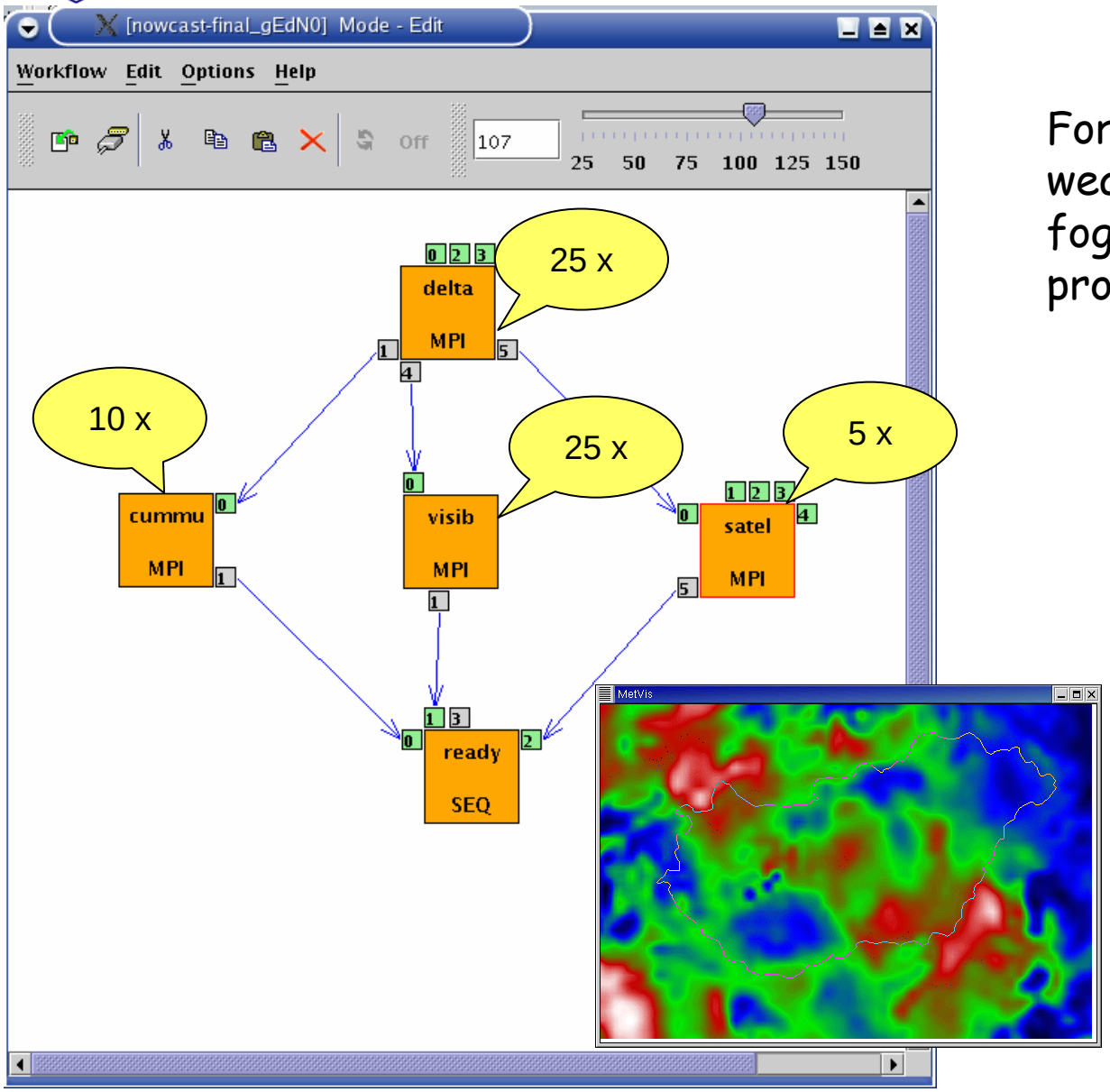


Two levels of parallelism by a workflow

- The workflow concept of the P-GRADE Portal enables the efficient parallelization of complex problems
- Semantics of the workflow enables two levels of parallelism:
  - Parallel execution inside a workflow node
  - Parallel execution among workflow nodes



Ultra-short range weather forecast (Hungarian Meteorology Service)

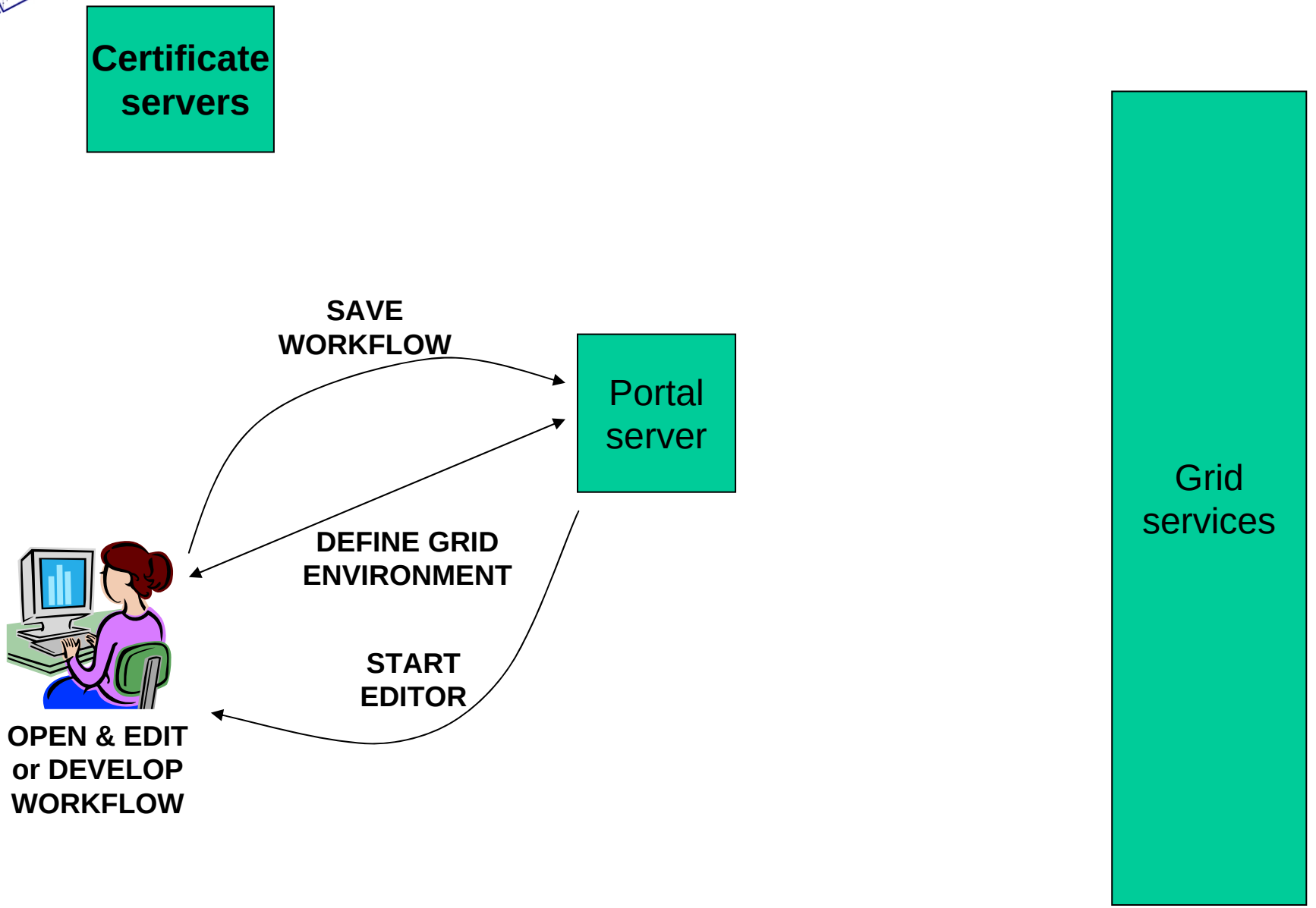


Forecasting dangerous weather situations (storms, fog, etc.), crucial task in the protection of life and property

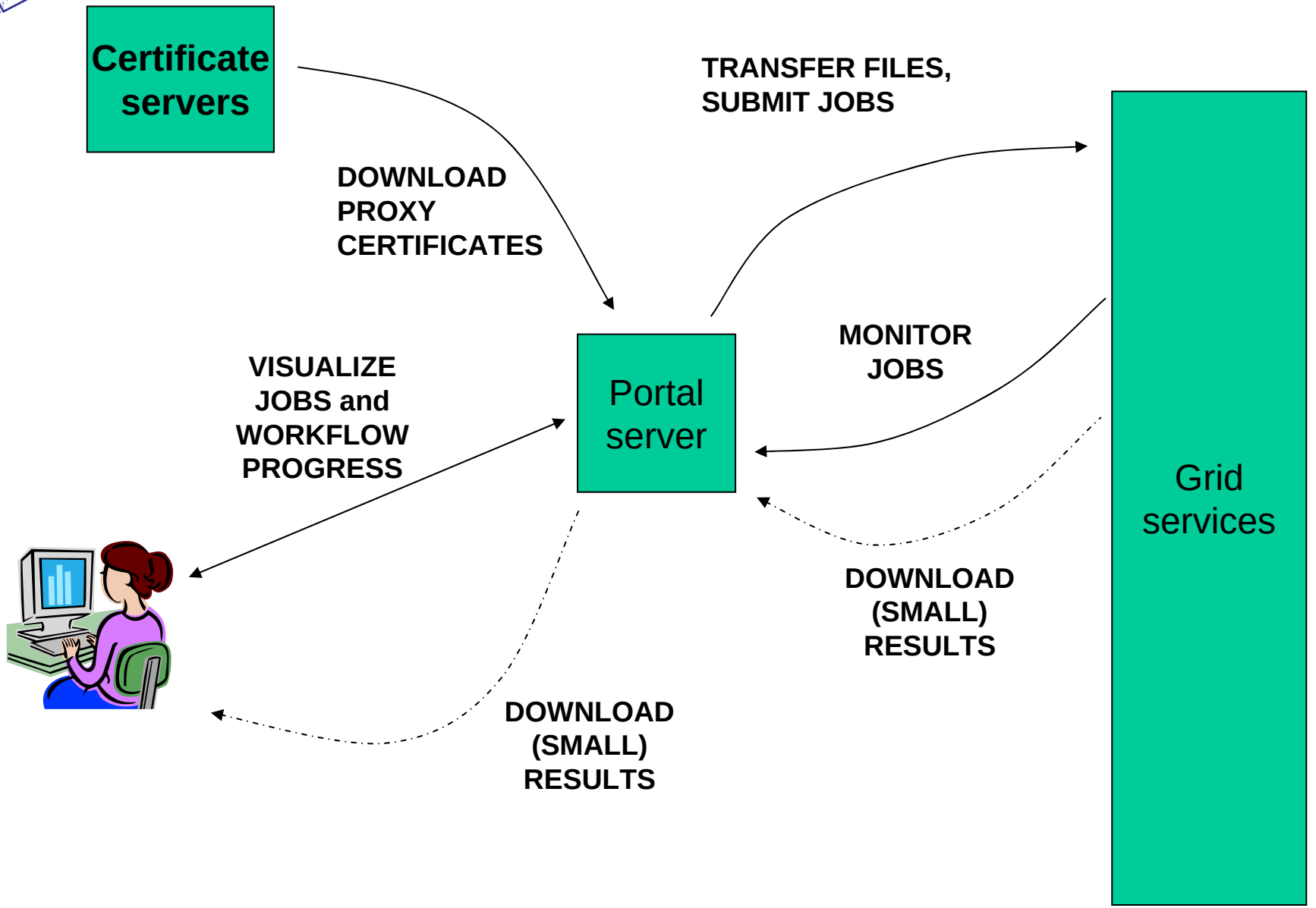
Processed information: surface level measurements, high-altitude measurements, radar, satellite, lightning, results of previous computed models

Requirements:
- Execution time < 10 min
- High resolution (1km)

The typical user scenario
Part 1 - development phase

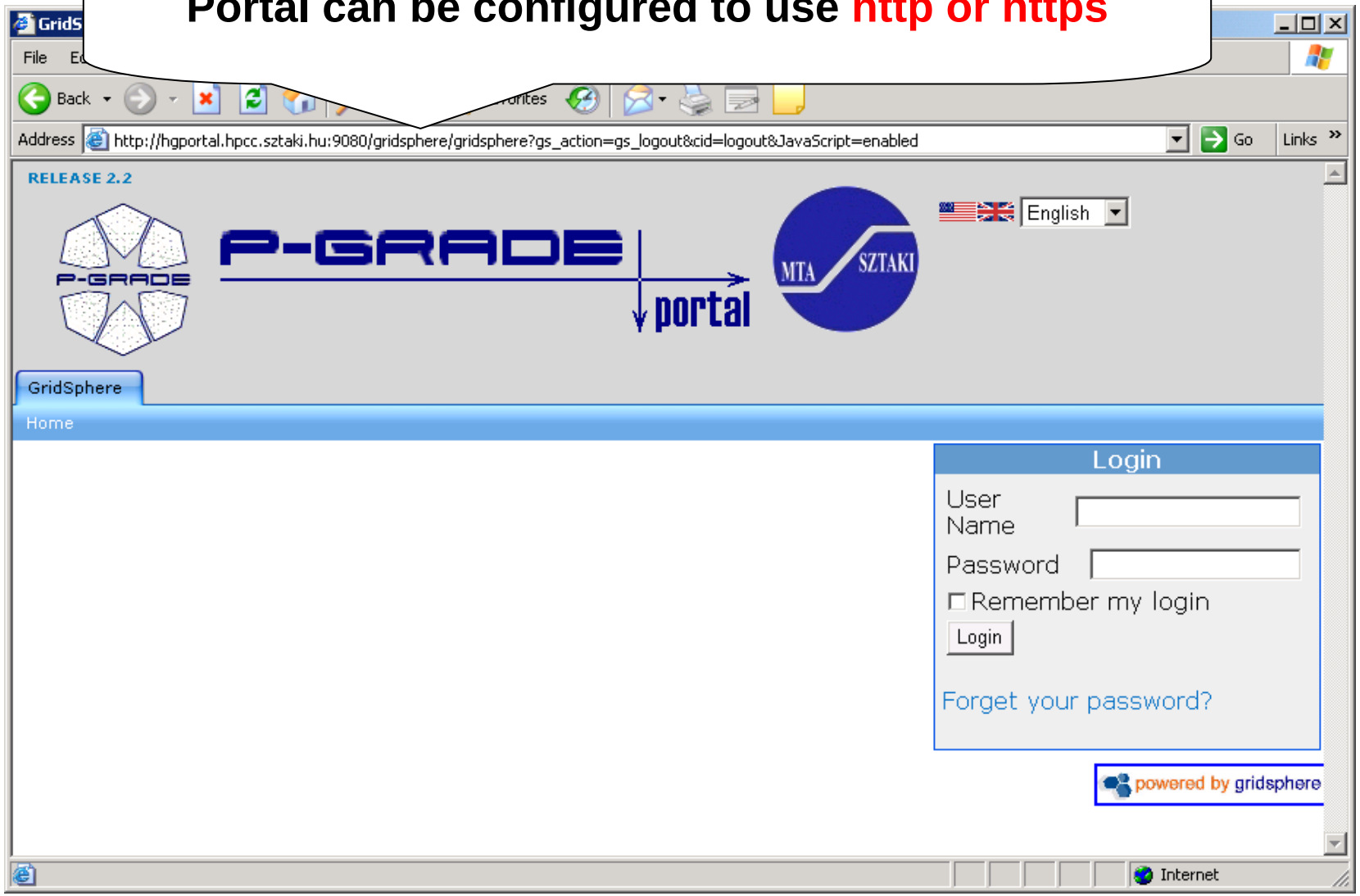


The typical user scenario
Part 2 - execution phase



0. step: login

Portal can be configured to use http or https



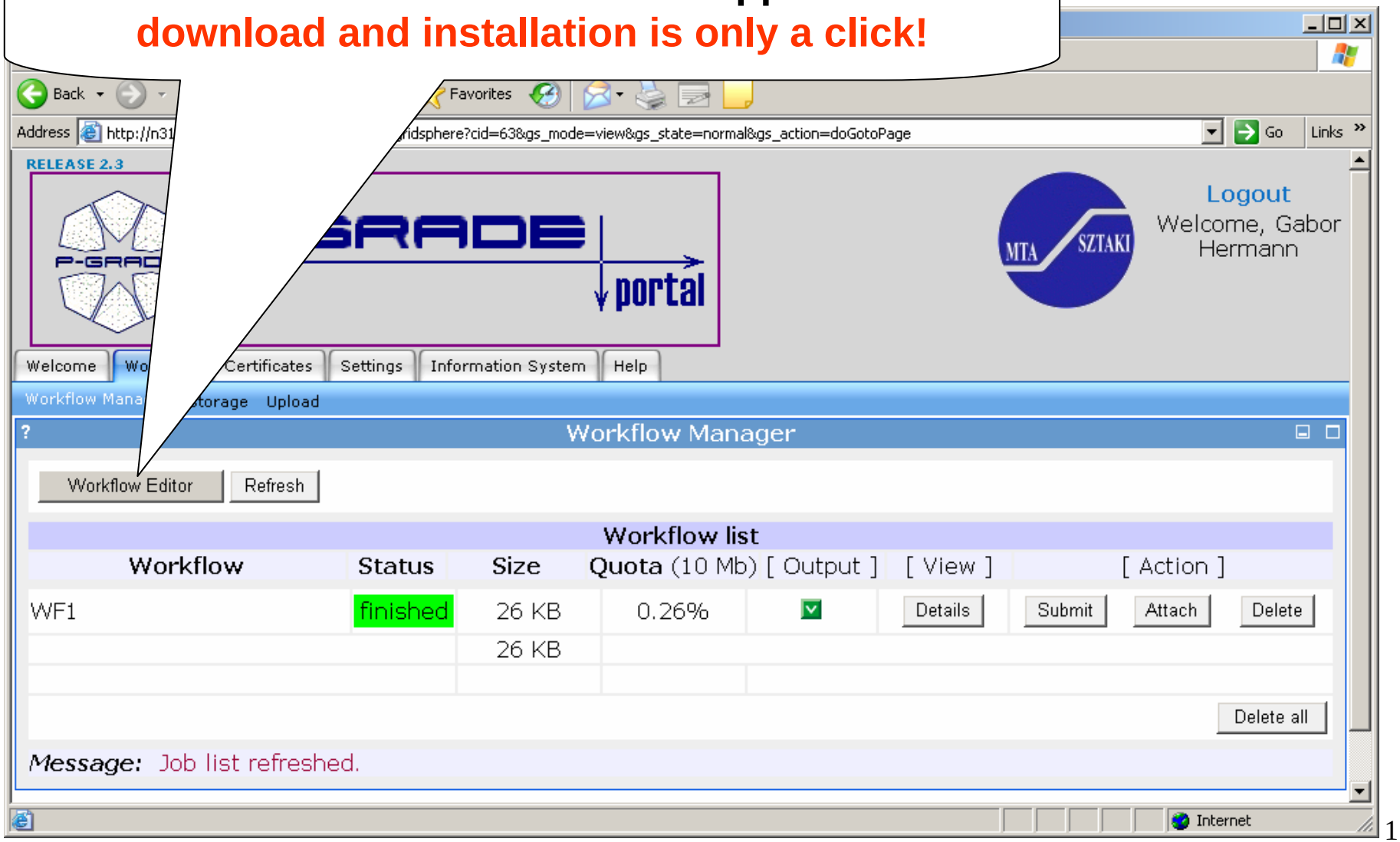
Developing workflows with the P-GRADE Portal

Main steps

1. Open the workflow editor
2. Define workflow
    1. Define graph structure
    2. Define jobs and input/output data
    3. Save workflow

Opening the workflow editor

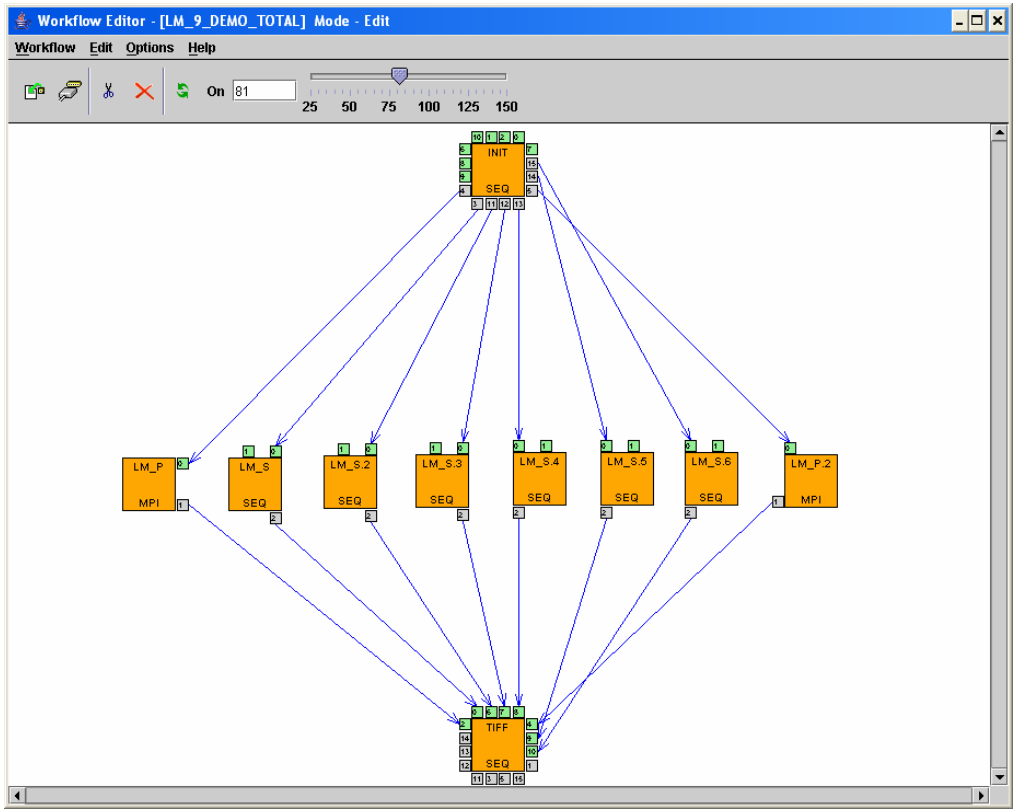
The editor is a Java Webstart application
download and installation is only a click!



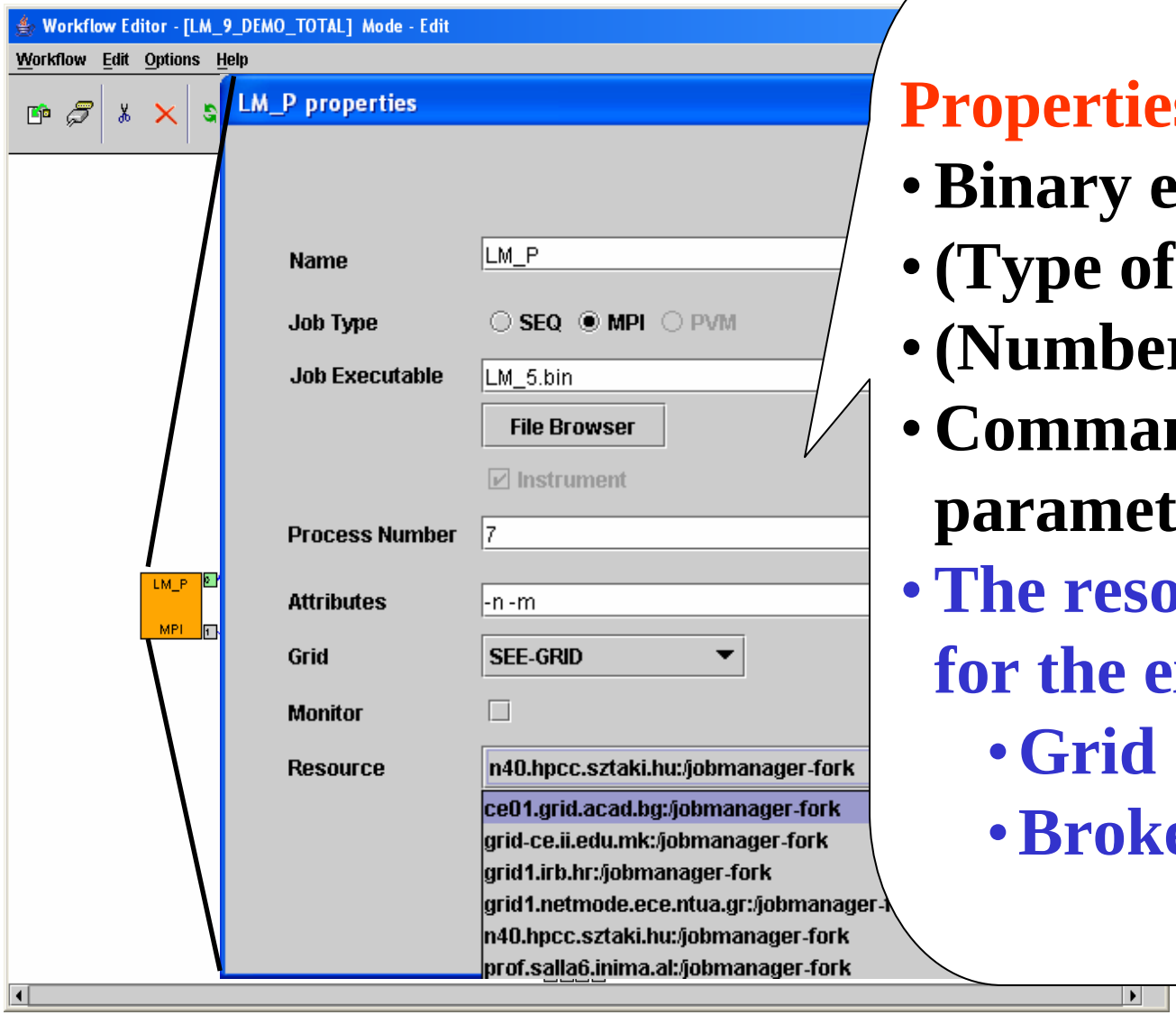
Workflow Editor

Define DAG of batch jobs:

1. **Drag & drop components:** jobs and ports
2. **Connect ports by channels** (no cycles, no loops)
3. **Define job and port properties**



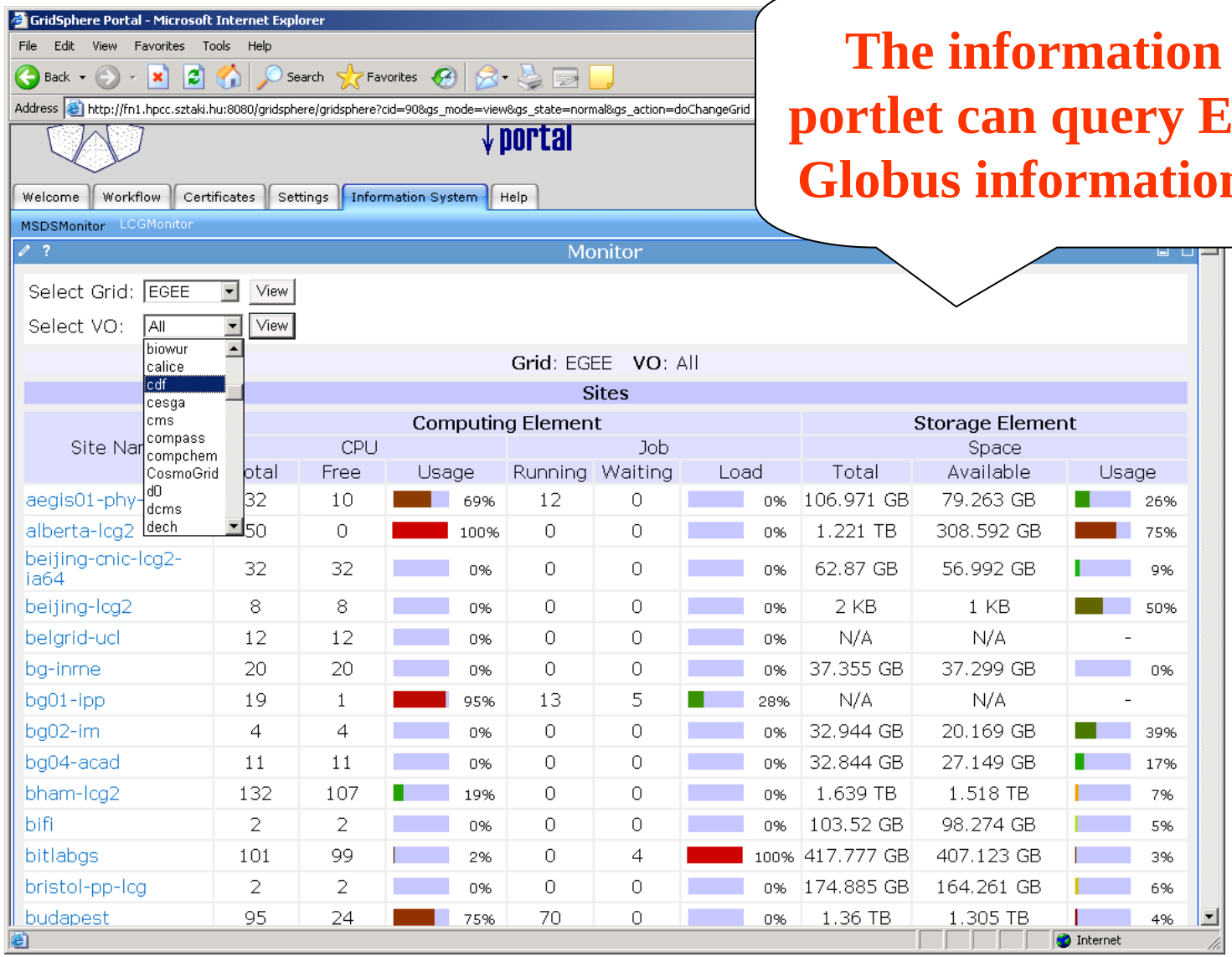
Job property window



Properties of a job:

- **Binary executable**
- **(Type of executable)**
- **(Number of processors)**
- **Command line parameters**
- **The resource to be used for the execution:**
  - **Grid**
  - **Broker / resource**

Support for manual resource selection: information system browser



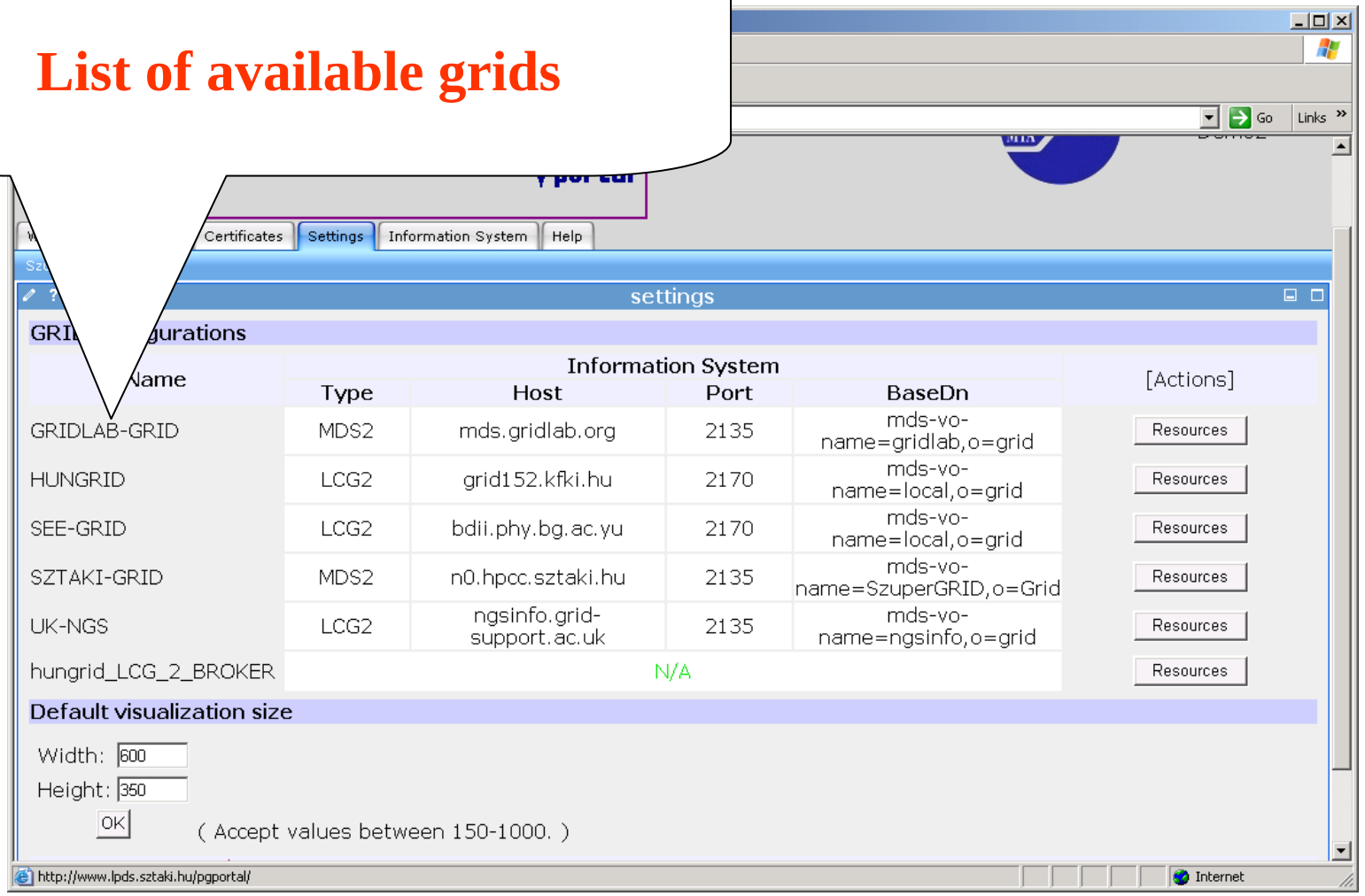
The information system portlet can query EGEE and Globus information systems

Support for manual resource selection: Settings portlet

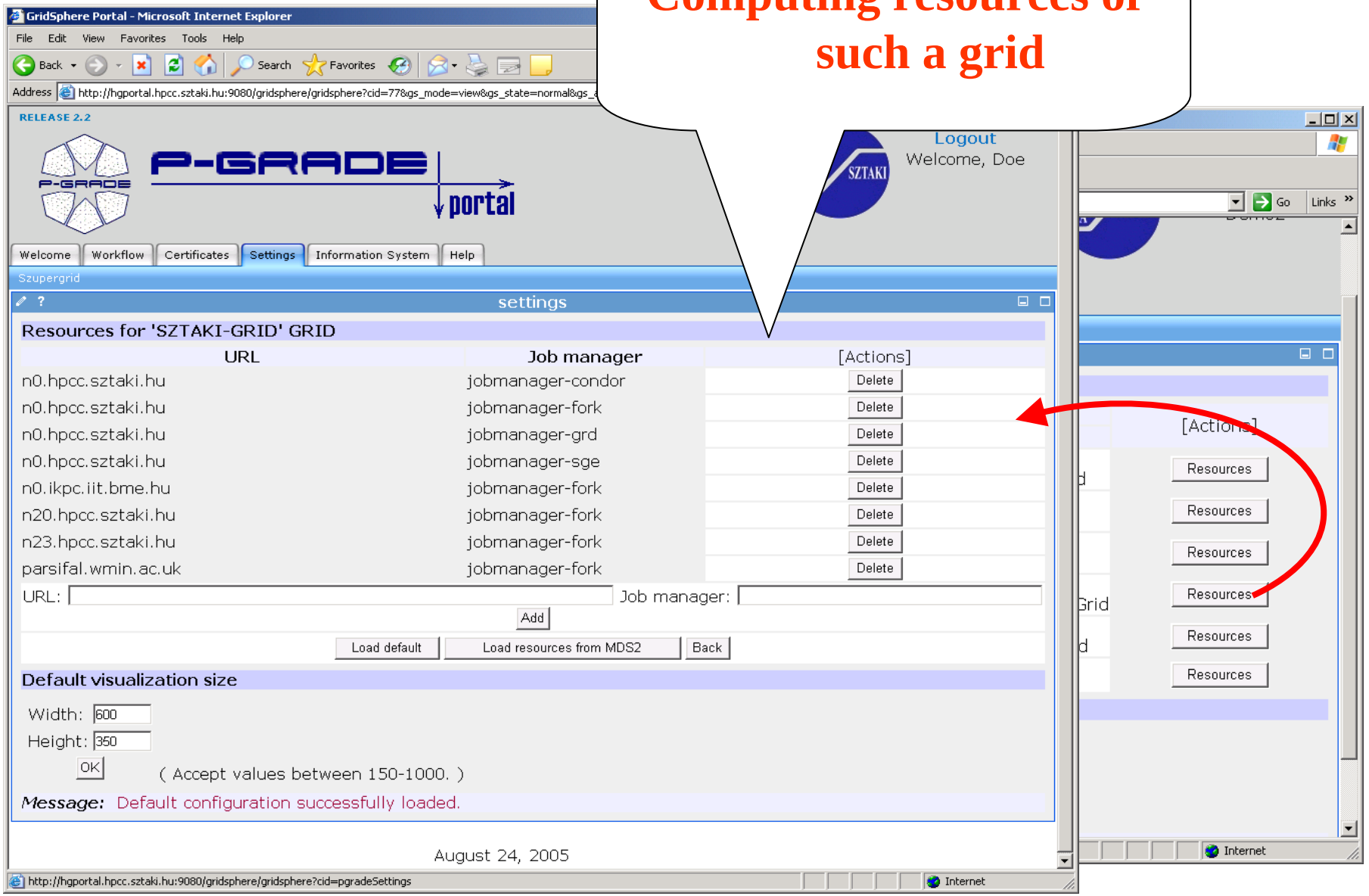
- **Here you can define those computing elements that your jobs can access directly** *(by skipping the broker)*
- Two levels:
  1. **Define grids** → portal administrator
     1. **Name** *(e.g. gridats)*
     2. **Information system** *(e.g. egrid-2.egrid.it)*
  2. **Define Computing Elements for each grid:**
     1. **Default list can be set by the portal administrator**
     2. **Users can customize the list**

Support for manual resource selection: Settings portlet

List of available grids



Support for manual resource selection: Sc

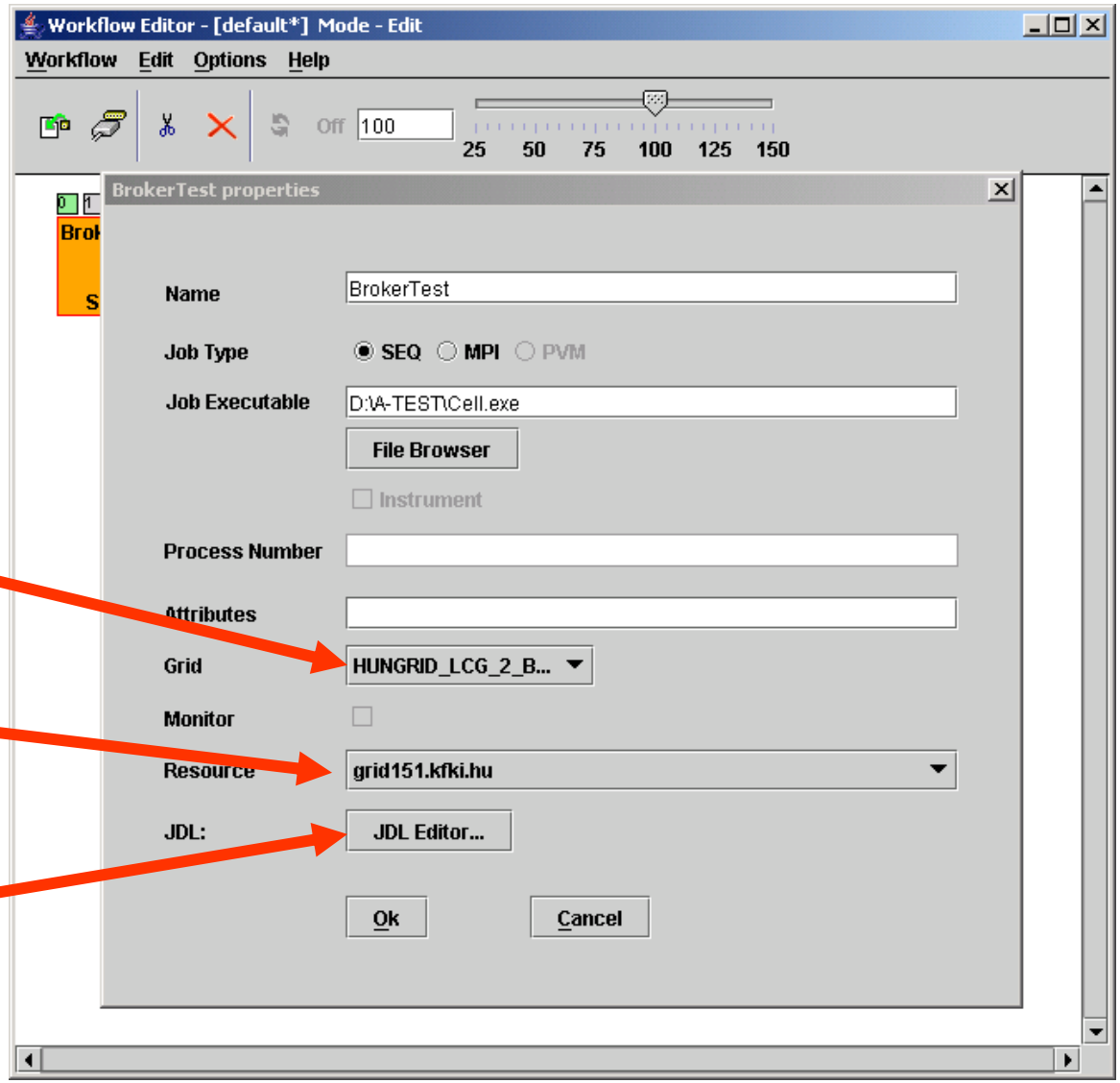


Support for broker-based resource selection

(currently not supported by the NGS)

1. Select a broker Grid for the job
2. (Specify extra ranks & requirements for the job in Job Description Language)
3. The broker will find the best resource for your job!

Support for broker-based resource selection

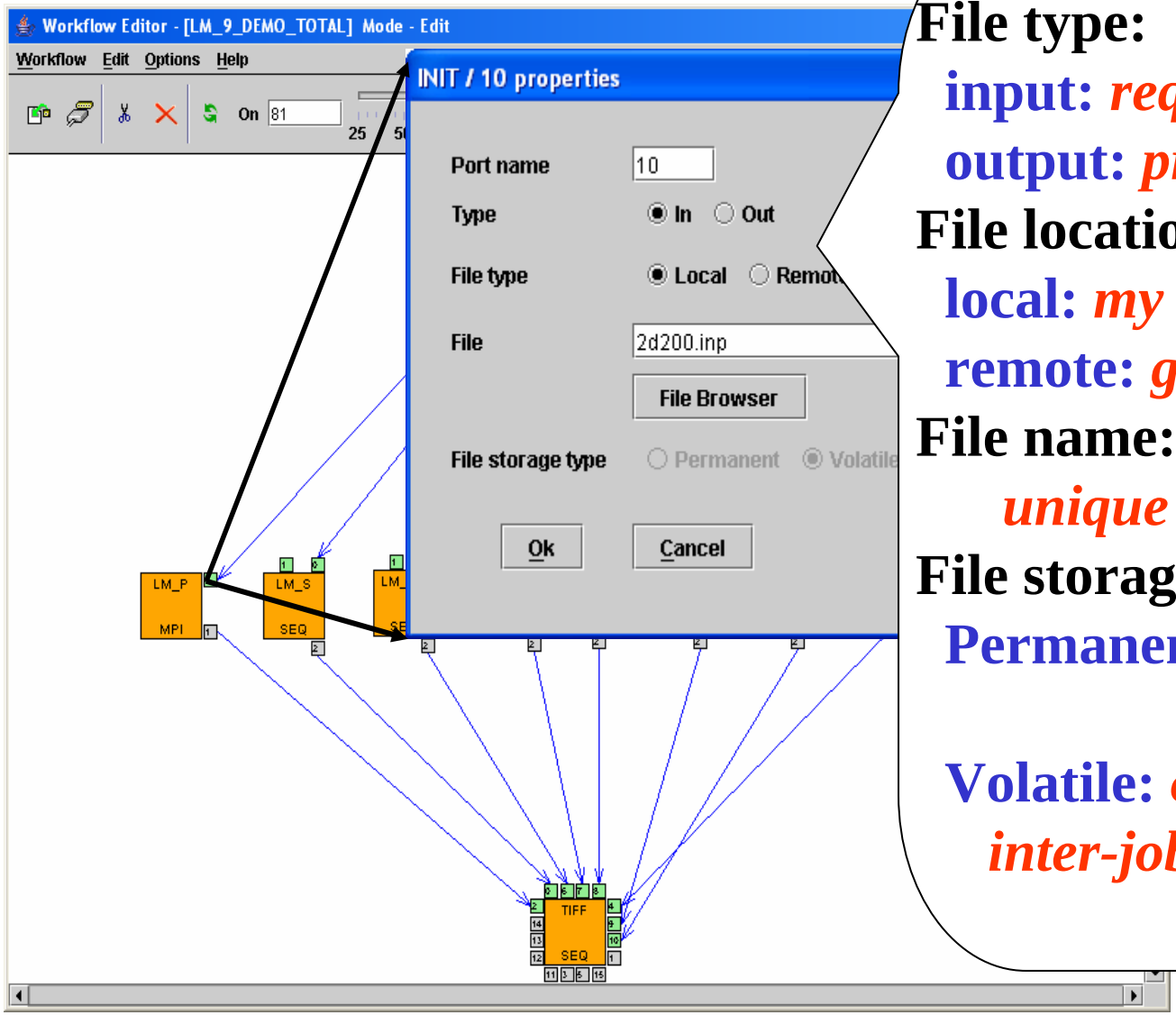


Select a Grid with broker! (*_BROKER)

Ignore the resource field!

If default is not sufficient define ranks & requirements using the built-in JDL editor!

Defining input/output data for jobs



File type:

- **input:** ***required by the job***
- **output:** ***produced by the job***

File location:

- **local:** ***my desktop***
- **remote:** ***grid storage resource***

File name:

- ***unique name of the file***

File storage type:

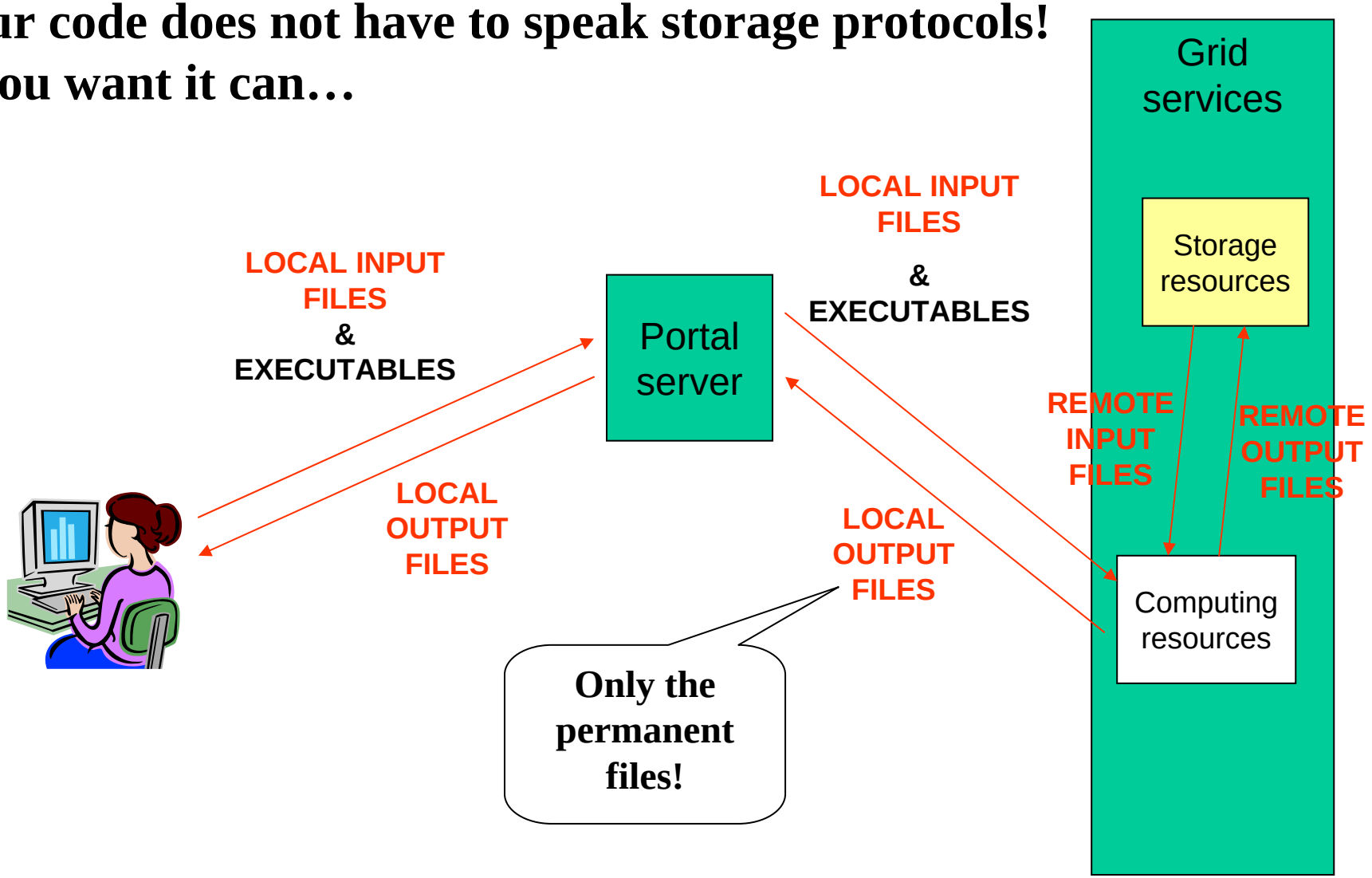
- **Permanent:** ***final result of the WF***
- **Volatile:** ***only used for inter-job data transfer***

Possible values for file location

Input file	Output file
Local file	
• **Client side location** c:\experiments\11-04.dat	• **Client side location** result.dat
• **Grid Unique IDentifier (GUID)** *(In any EGEE Grid)* guid:1fd75fdf-dccc-4603-998b-e17facb0d034 • **LRS logical file name** *(In RMC-enabled EGEE Grids)* lfn:/sipos_11_04.dat • **LFC logical file name** *(In LFC-enabled EGEE Grids)* lfn:/grid/egrid/sipos/11-04.dat • **GSIFTP reference** *(In Globus Grids)* gsiftp://lpds.sztaki.hu/sipos/11-04.dat	• **LRS logical file name** *(In RMC-enabled EGEE Grids)* lfn:/sipos_11_04_-_result.dat • **LFC logical file name** *(In LFC-enabled EGEE Grids)* lfn:/grid/egrid/sipos/11-04_-_result.dat • **GSIFTP reference** *(In Globus Grids)* gsiftp://lpds.sztaki.hu/sipos/11-04_-_result.dat
Remote file	

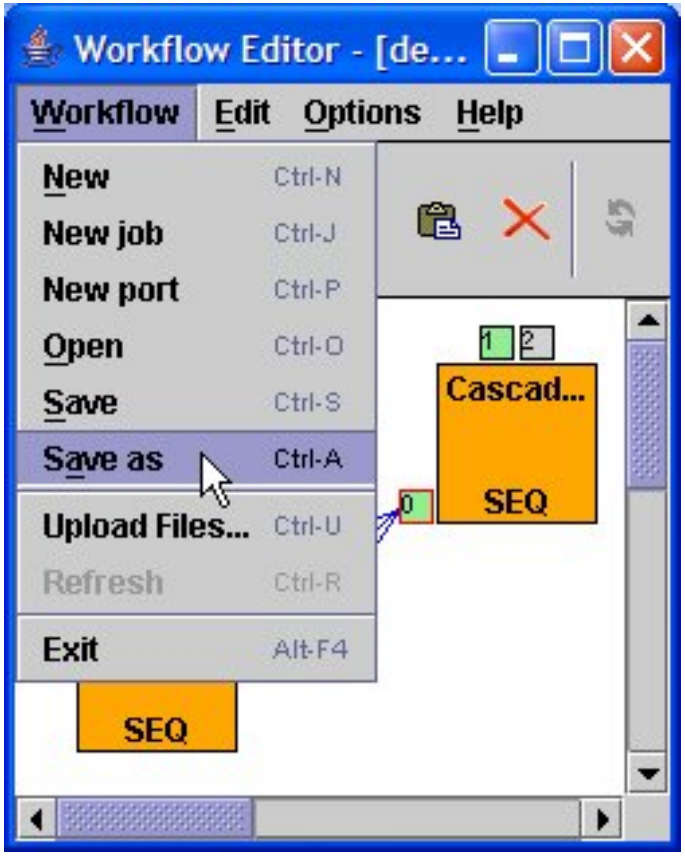
Local vs. remote files

Your code does not have to speak storage protocols!
If you want it can…



Workflow Editor

Saving the workflow



Workflow is defined!

Let's execute it!

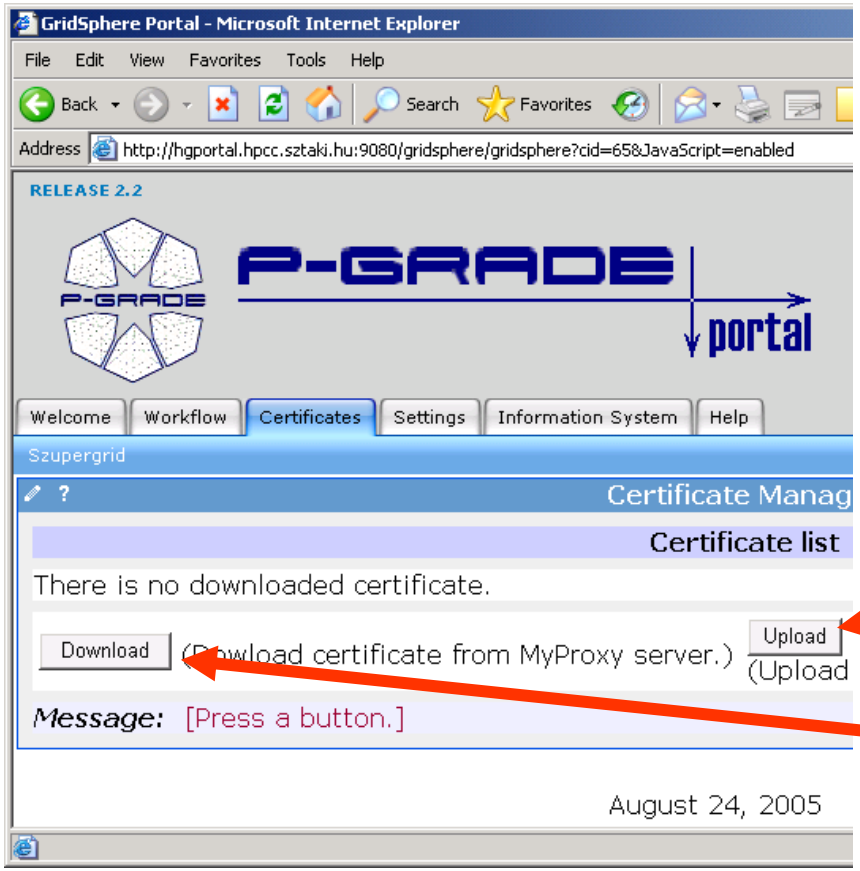
Executing workflows with the P-GRADE Portal

Main steps

1. Download proxies
2. Submit workflow
3. Observe workflow progress
4. If some error occurs correct the graph
5. Download result

The typical user scenario
Execution phase – step 1:

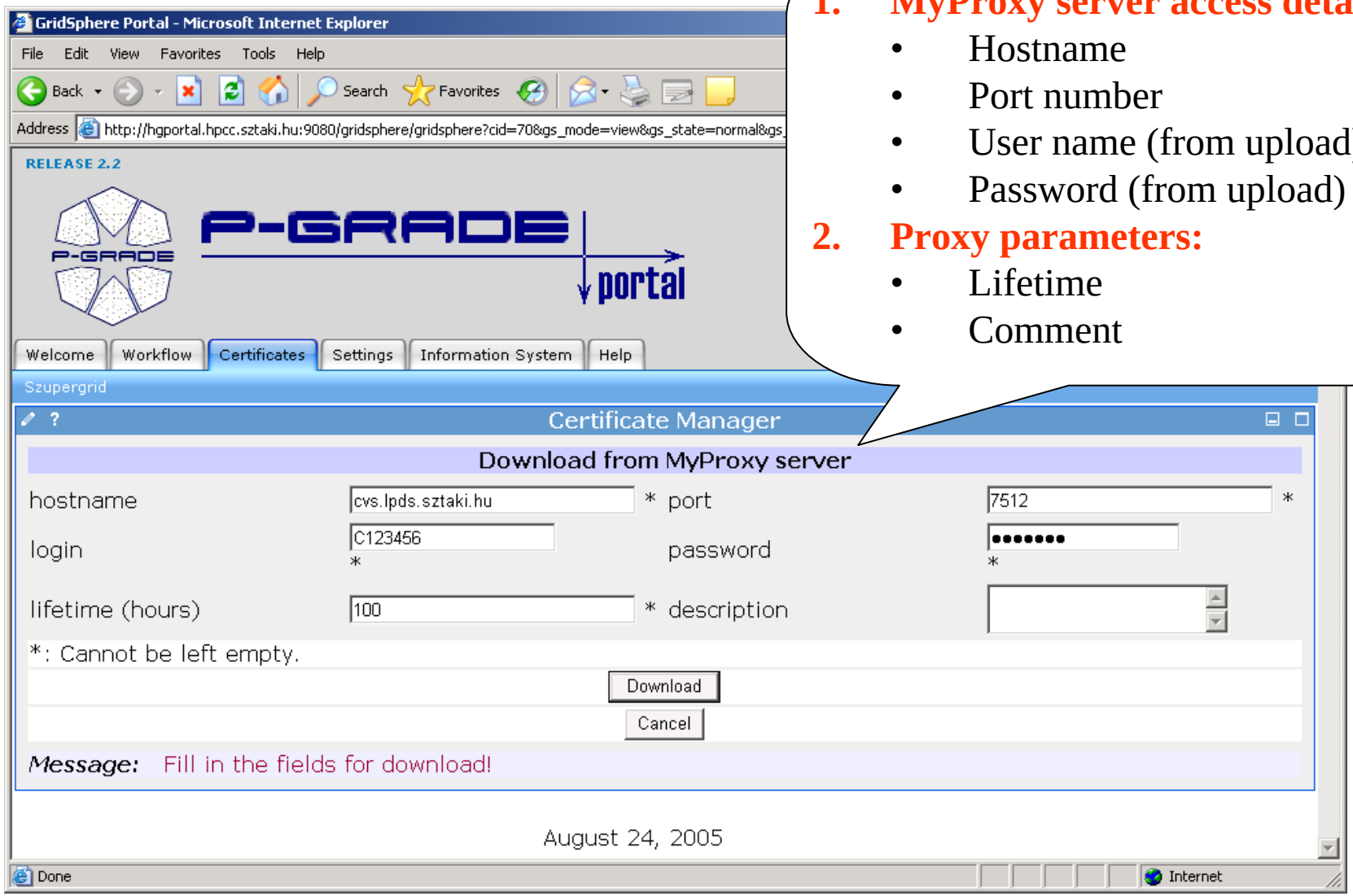
Certificate manager portlet



- To access GSI-based Grids the portal server application needs proxy certificates
- "Certificates" portlet:
  - to upload X.509 certificates into MyProxy servers
  - to download short-term proxy credentials into the portal server application

Certificate manager portlet

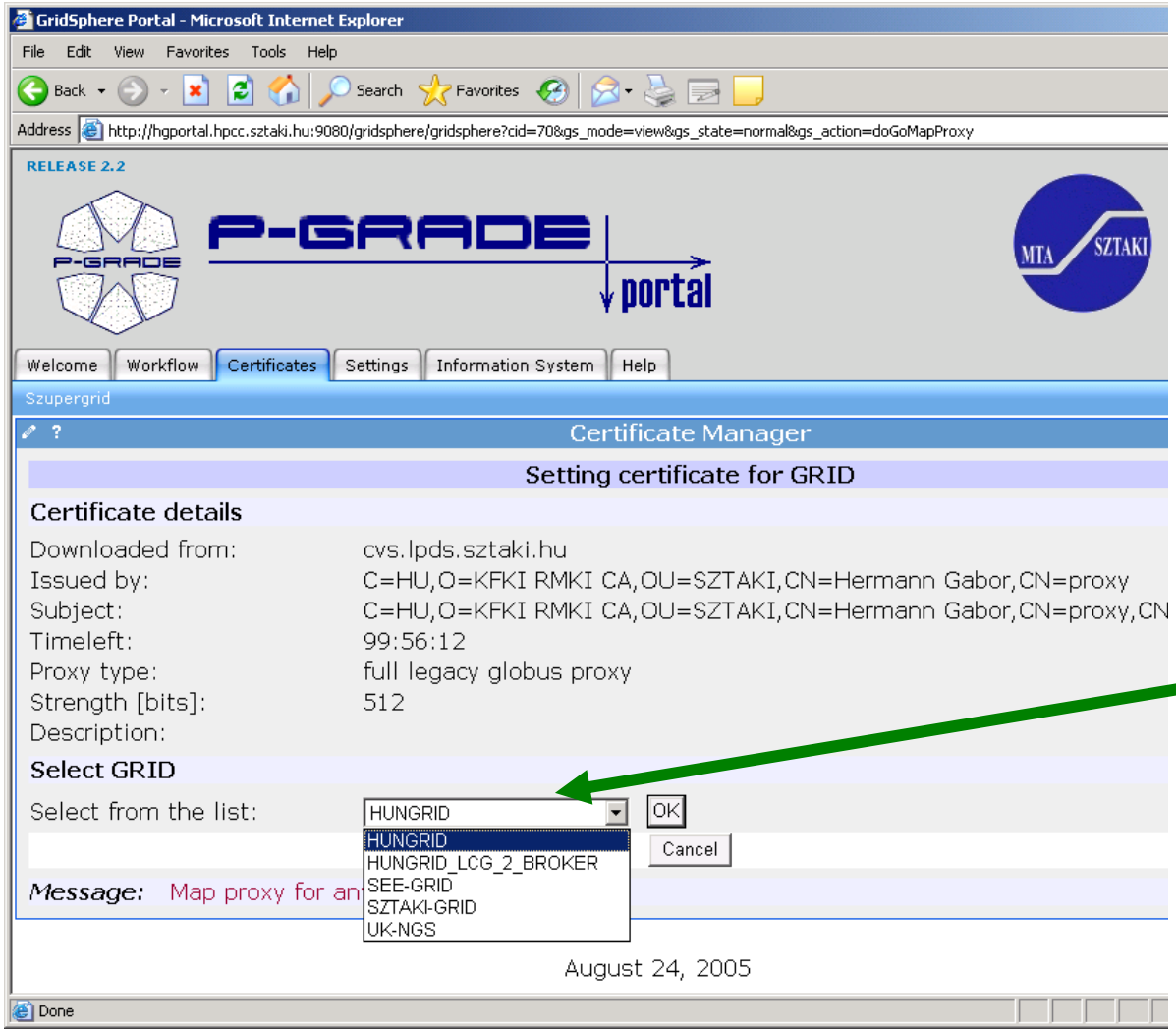
Downloading a proxy



1. **MyProxy server access details:**
   - Hostname
   - Port number
   - User name (from upload)
   - Password (from upload)
2. **Proxy parameters:**
   - Lifetime
   - Comment

Certificate manager portlet

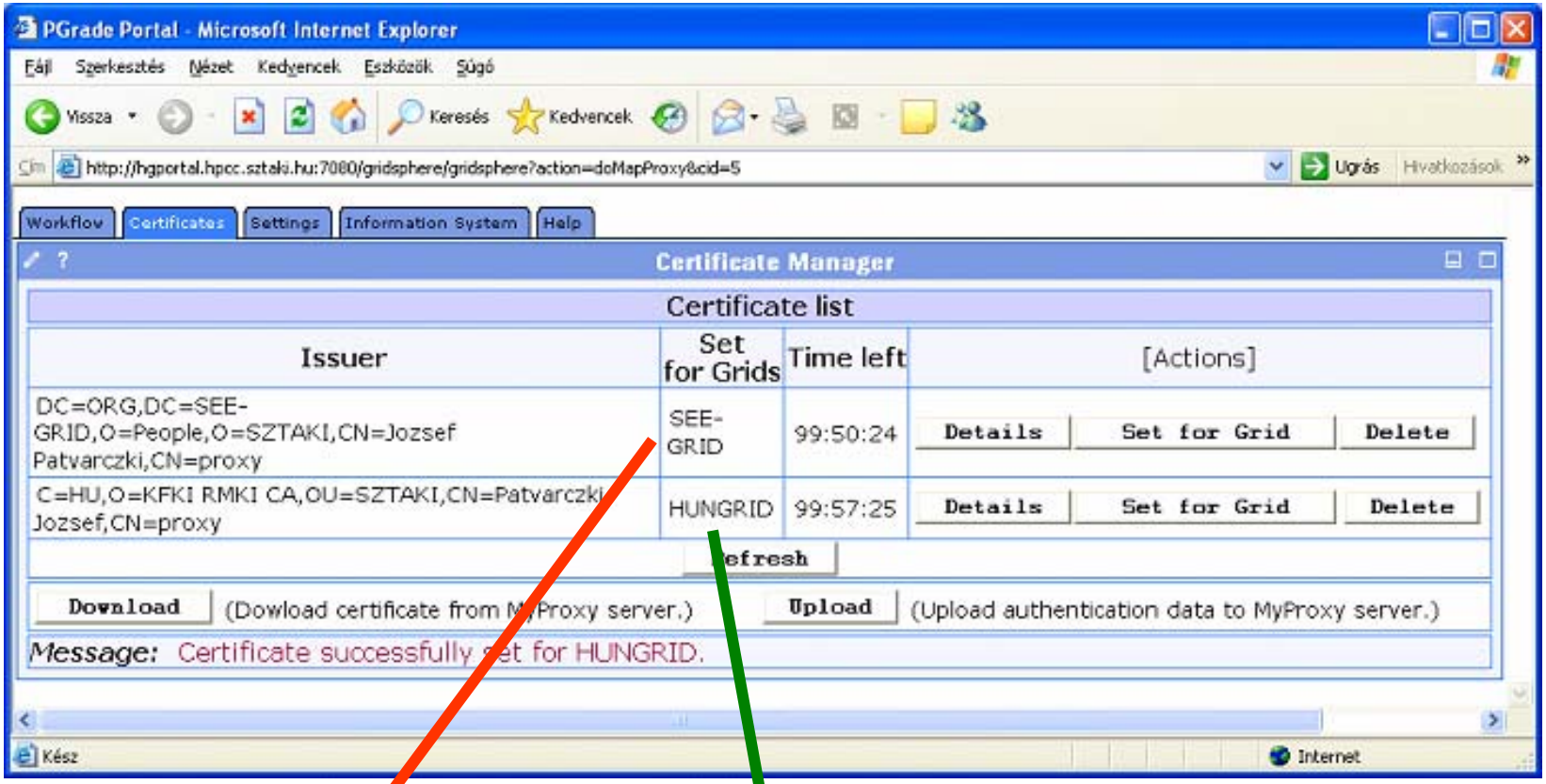
Associating the proxy with a grid



This operation displays the details of the certificate and the list of available Grids

Certificate Manager

browsing proxies



SEE-GRID resources

HUNGRID resources

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

The typical user scenario
Execution phase - step 2:

Certificate servers

TRANSFER FILES, SUBMIT JOBS

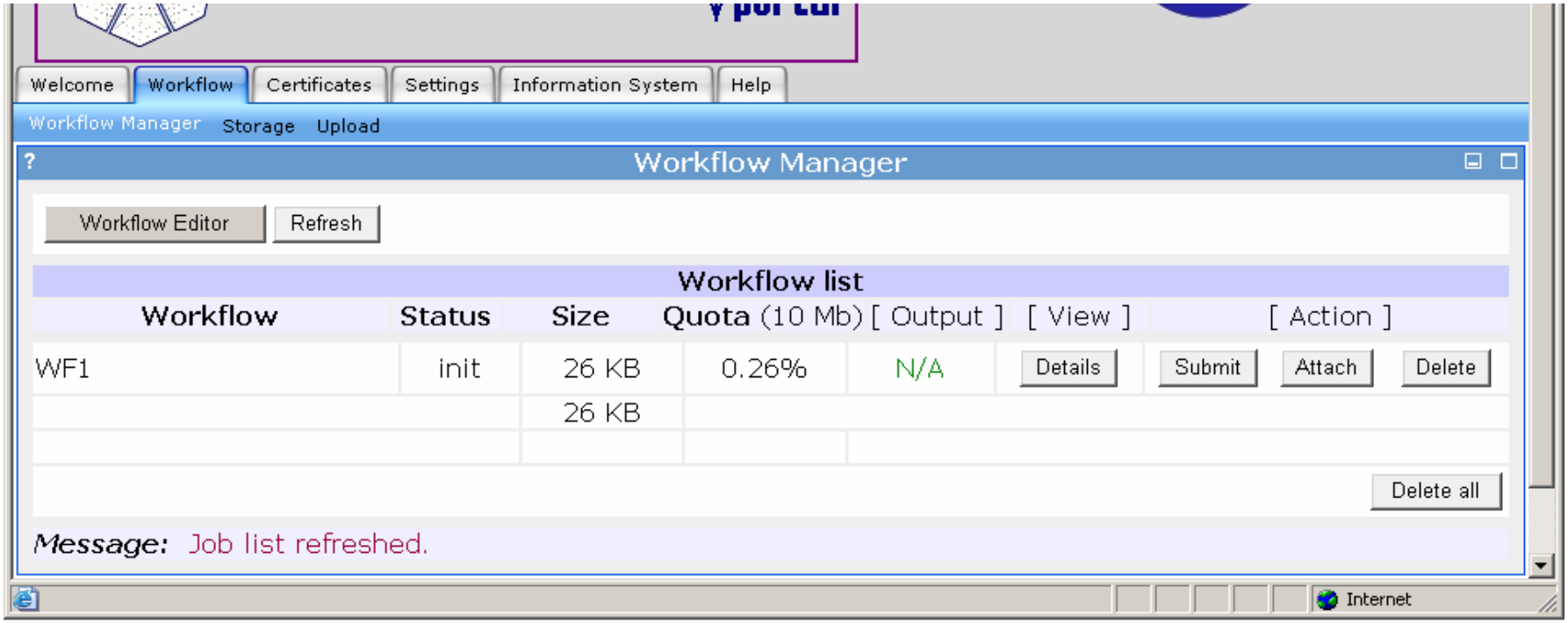
Portal server

Grid services

Workflow Management

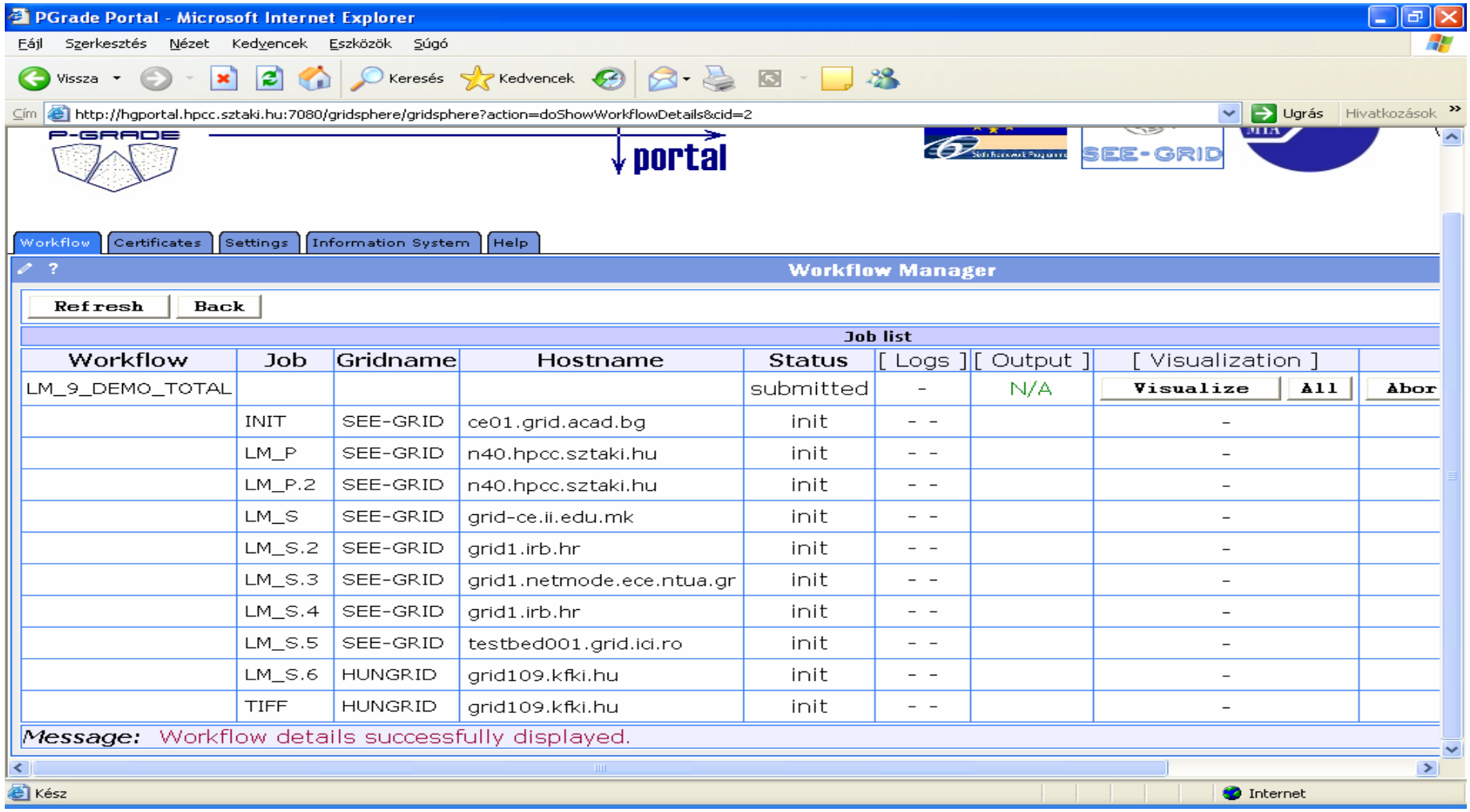
(workflow portlet)

- The portlet presents the status, size and output of the available workflow in the "Workflow" list
- It has a Quota manager to control the users' storage space on the server
- The portlet also contains the "Abort", "Attach", "Details", "Delete" and "Delete all" buttons to handle execution of workflows
- The "Attach" button opens the workflow in the Workflow Editor
- The "Details" button gives an overview about the jobs of the workflow



Workflow Execution

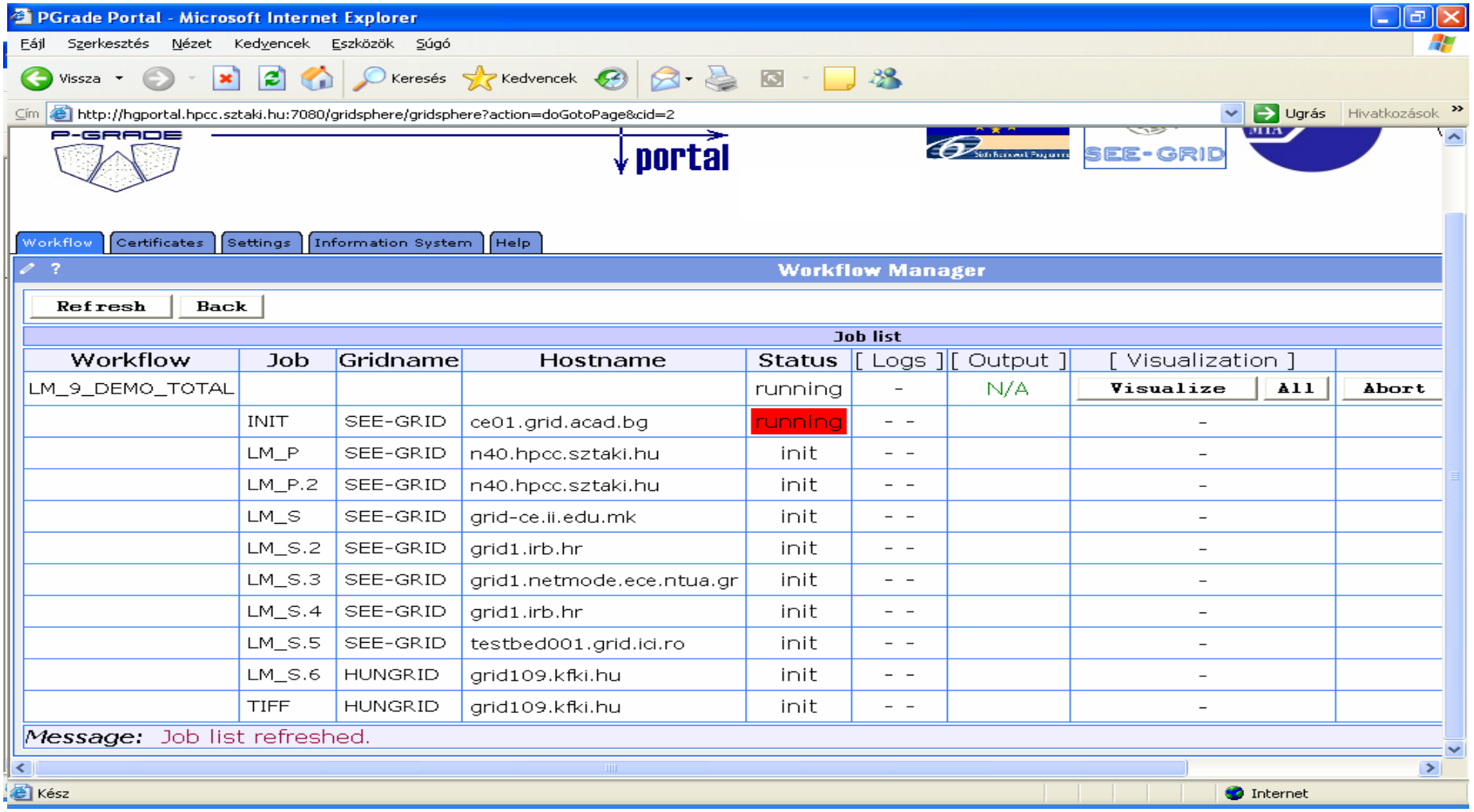
(observation by the workflow portlet)



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

Workflow Execution

(observation by the workflow portlet)

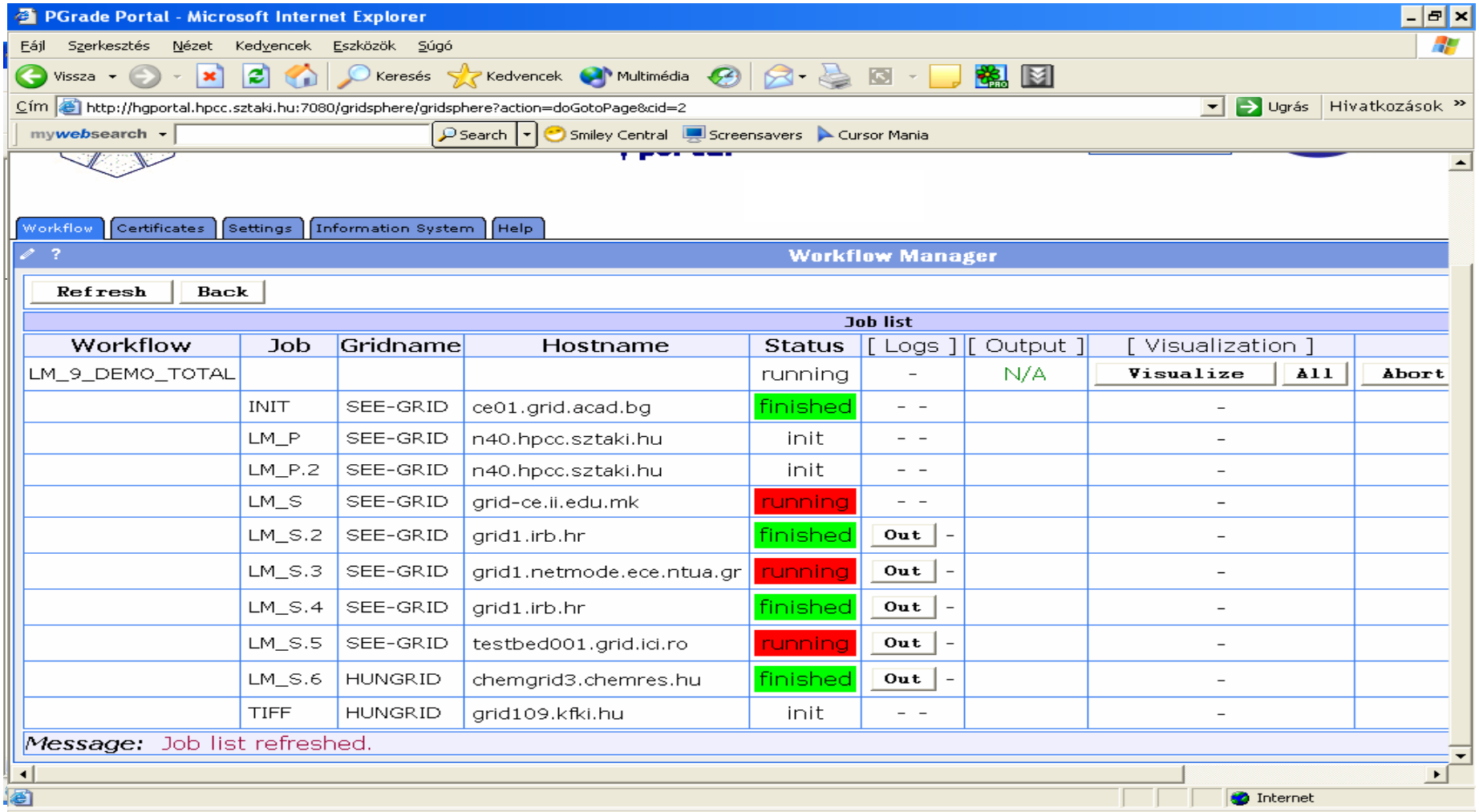


Workflow	Job	Gridname	Hostname	Status	[Logs]	[Output]	[Visualization]	
LM_9_DEMO_TOTAL				running	-	N/A	Visualize All	Abort
	INIT	SEE-GRID	ce01.grid.acad.bg	running	- -		-	
	LM_P	SEE-GRID	n40.hpcc.sztaki.hu	init	- -		-	
	LM_P.2	SEE-GRID	n40.hpcc.sztaki.hu	init	- -		-	
	LM_S	SEE-GRID	grid-ce.ii.edu.mk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

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

Workflow Execution

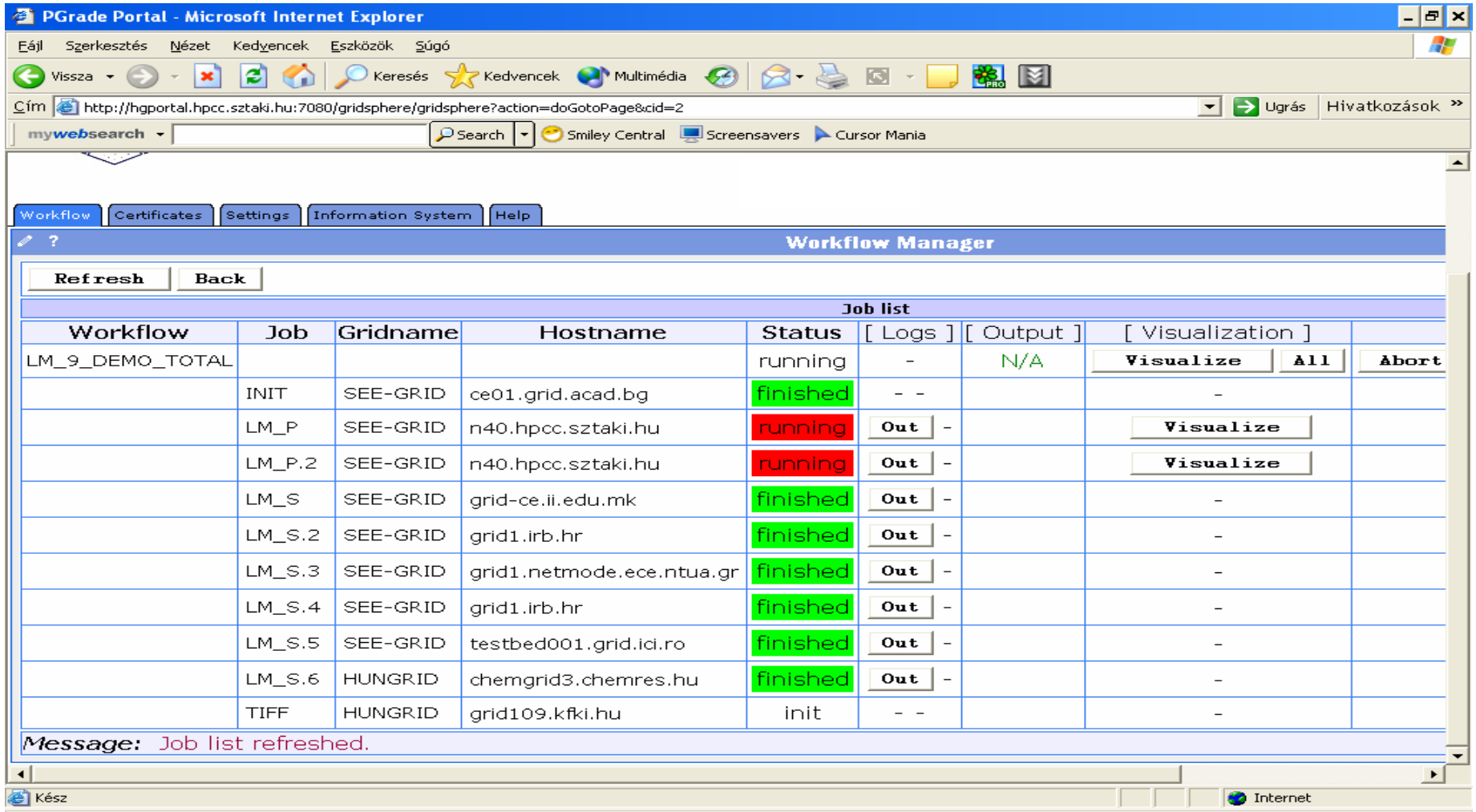
(observation by the workflow portlet)



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

Workflow Execution

(observation by the workflow portlet)

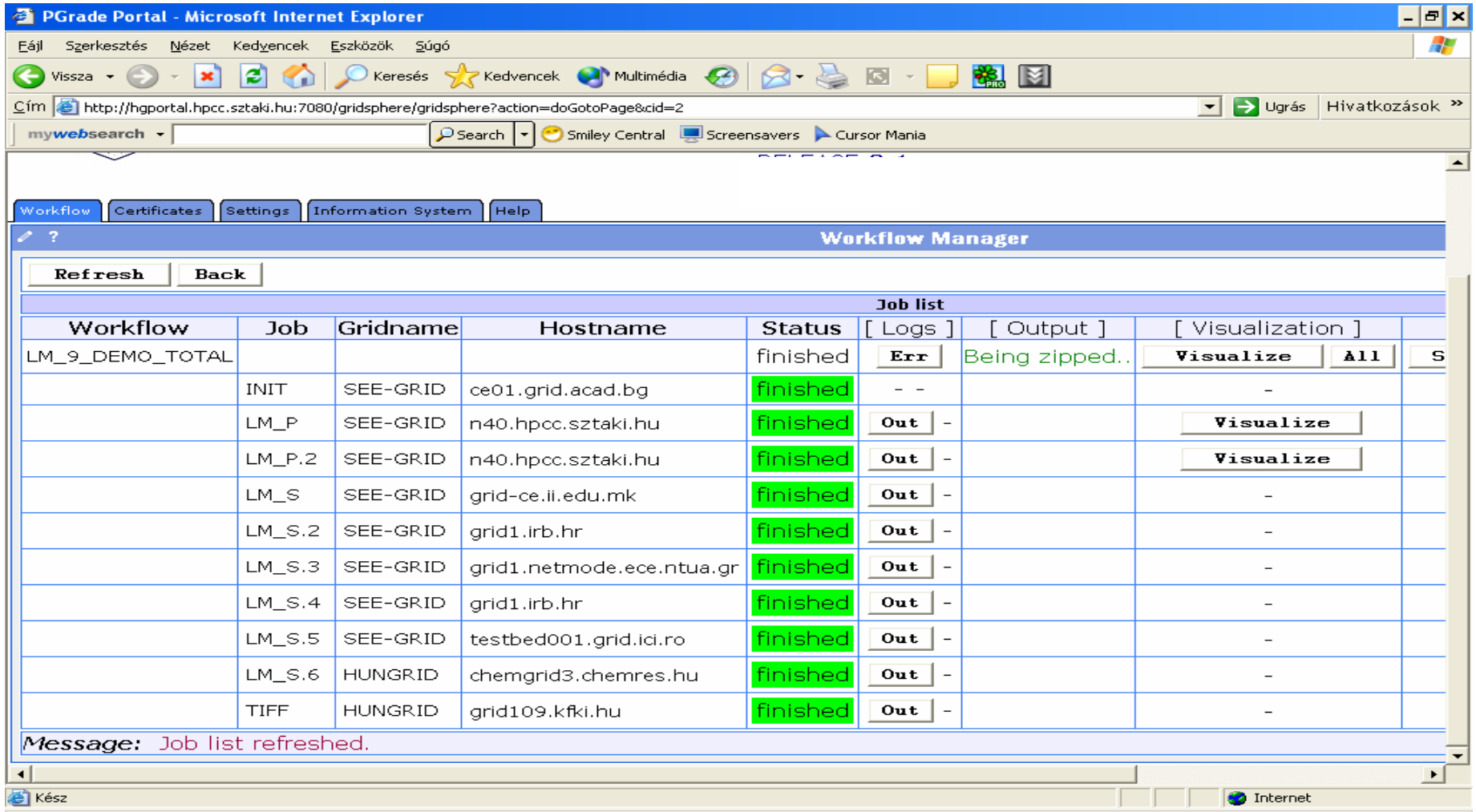


Workflow	Job	Gridname	Hostname	Status	[Logs]	[Output]	[Visualization]	
LM_9_DEMO_TOTAL				running	-	N/A	Visualize All	Abort
	INIT	SEE-GRID	ce01.grid.acad.bg	finished	- -		-	
	LM_P	SEE-GRID	n40.hpcc.sztaki.hu	running	Out -		Visualize	
	LM_P.2	SEE-GRID	n40.hpcc.sztaki.hu	running	Out -		Visualize	
	LM_S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

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

Workflow Execution

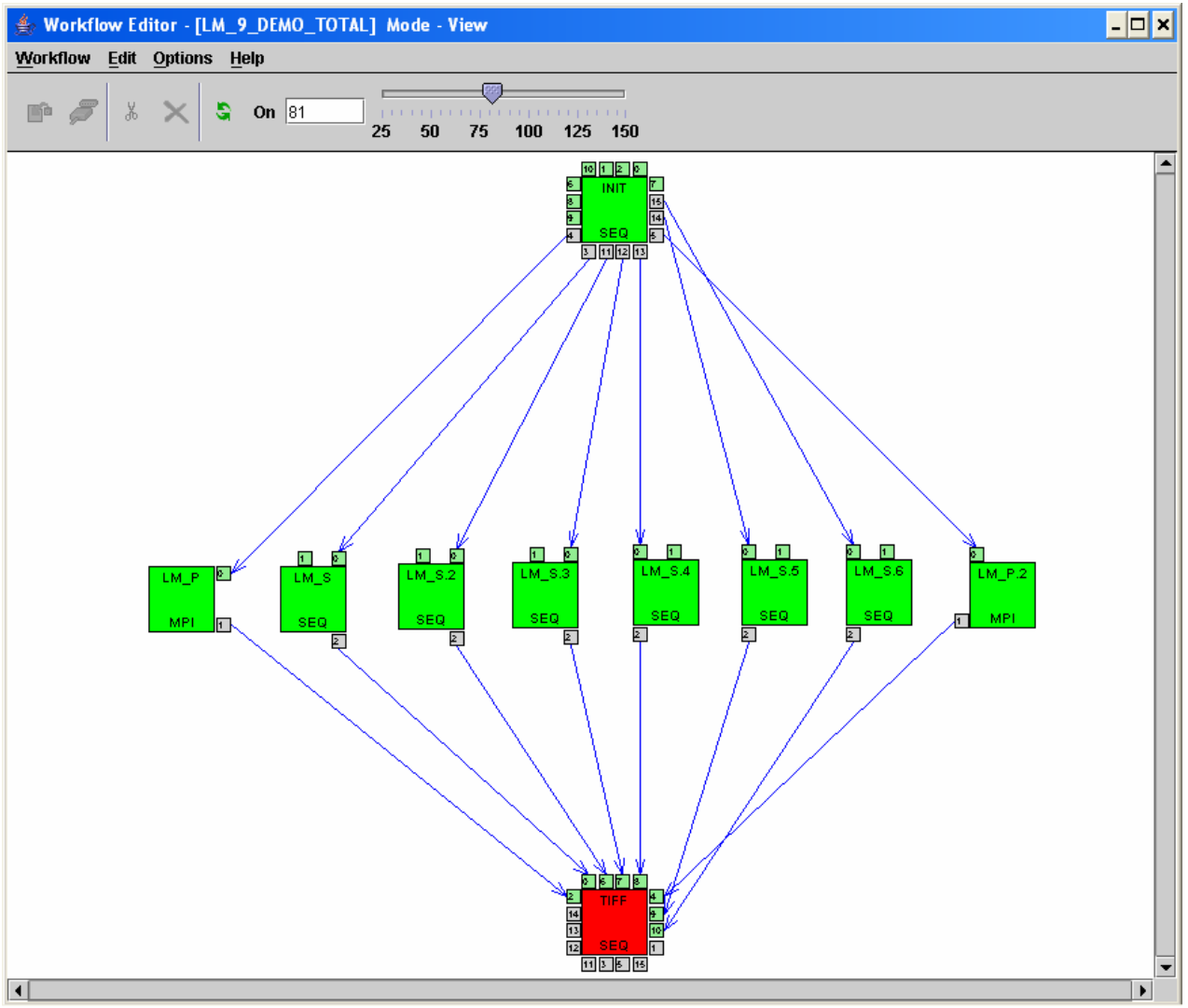
(observation by the workflow portlet)



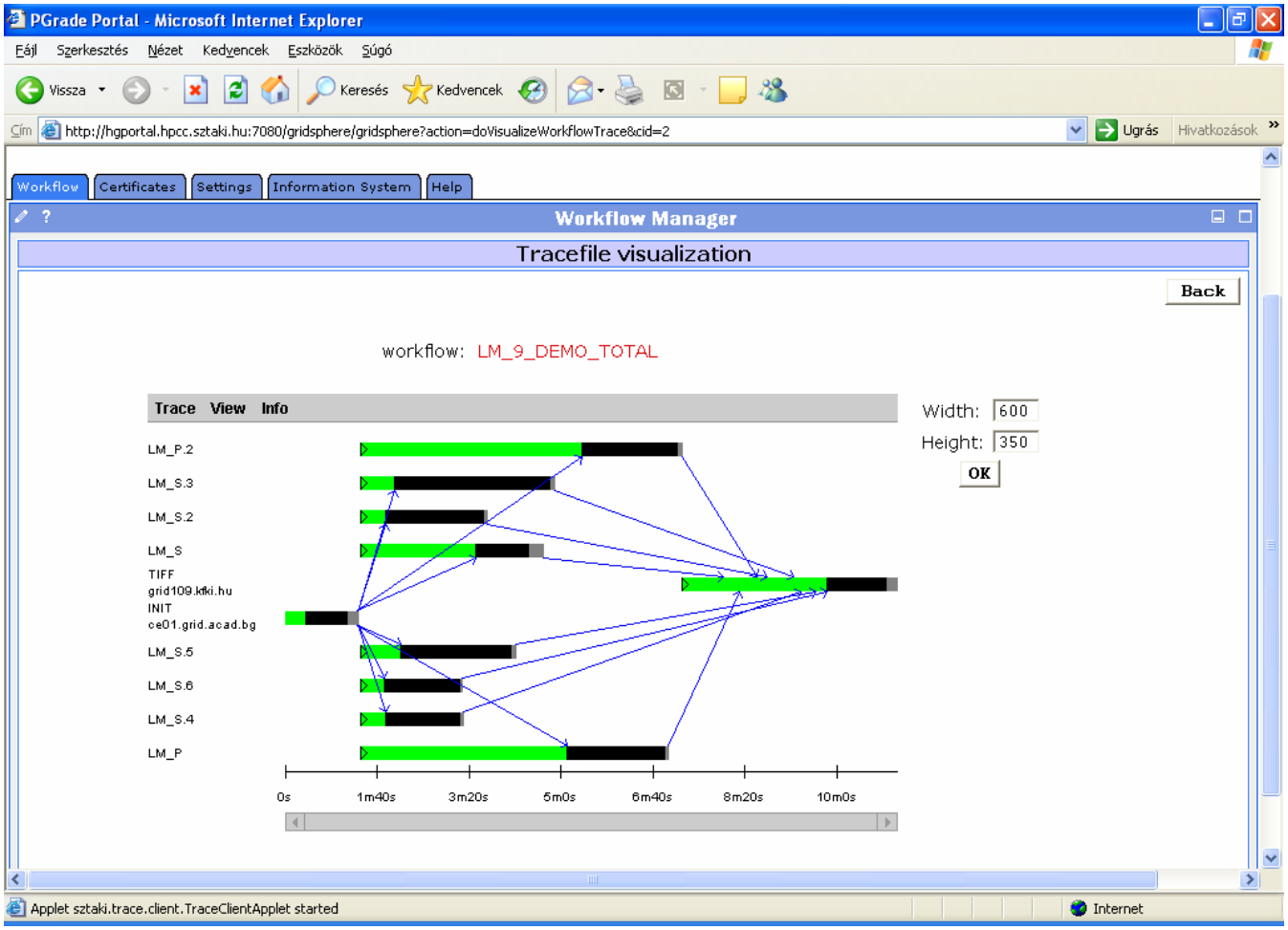
White/Red/Green color means the job is initialised/running/finished

Workflow Execution

(observation by the workflow editor)

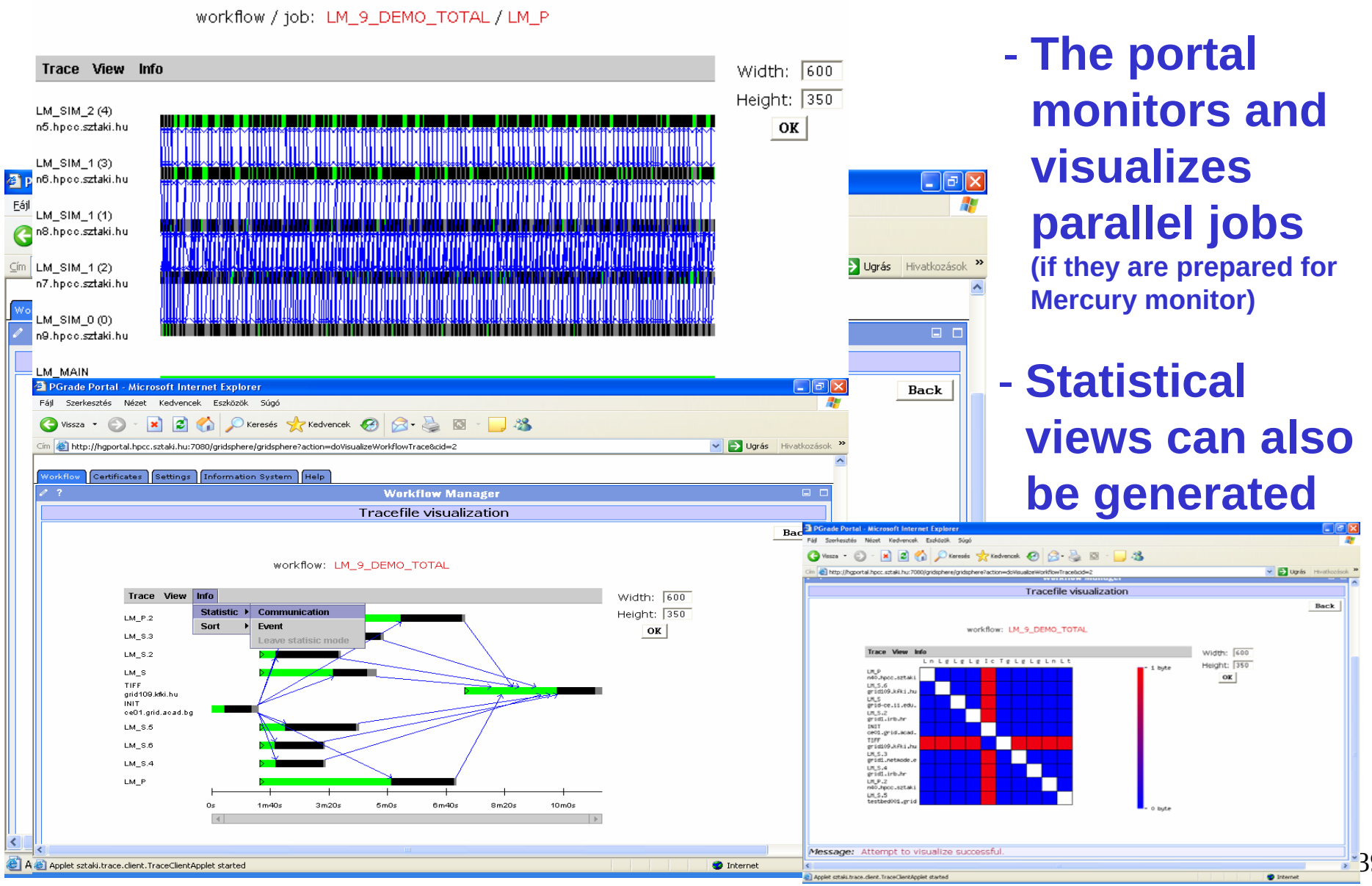


On-line application monitoring: workflow and job level



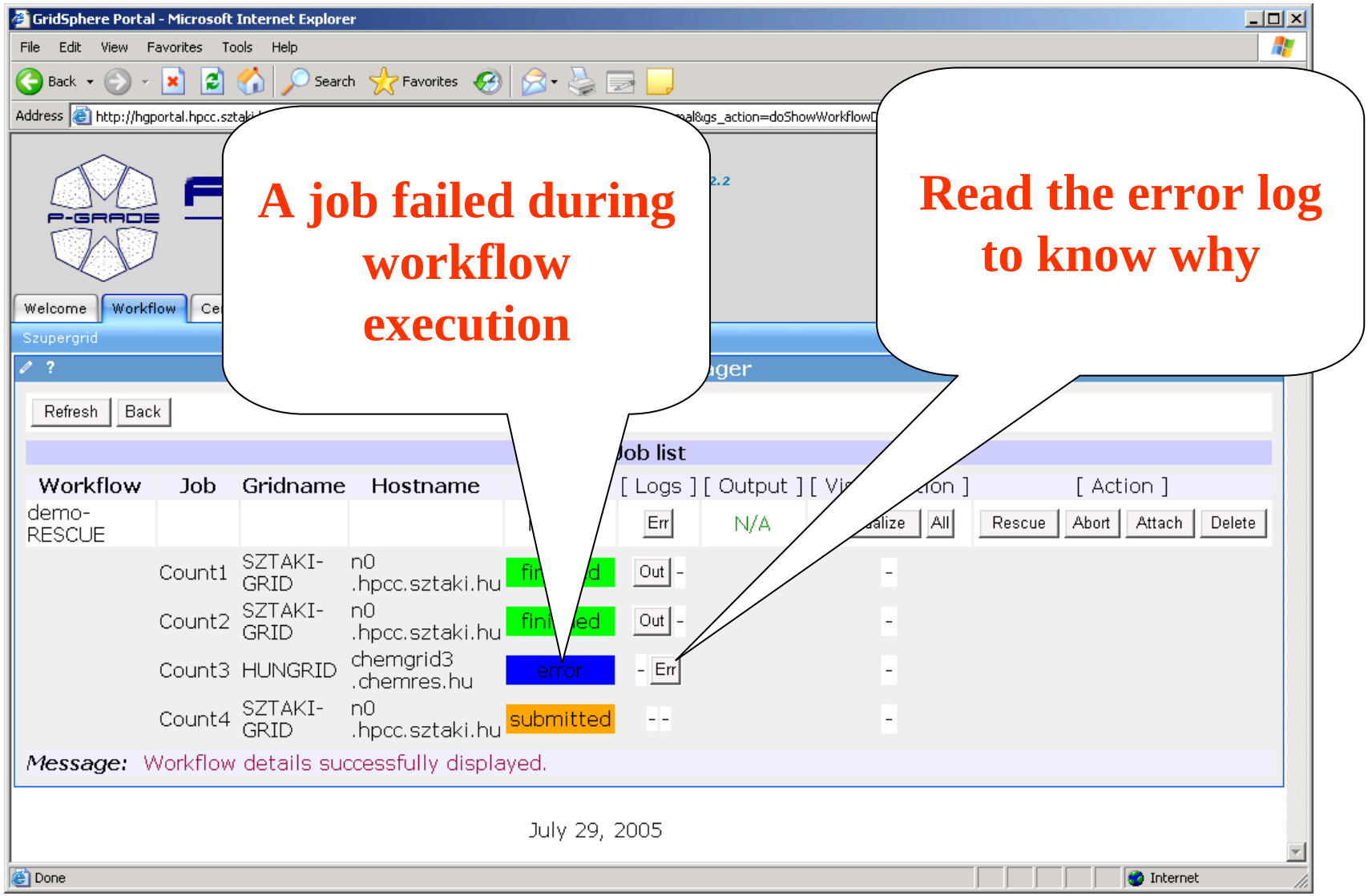
- The portal monitors and visualizes workflow progress

On-line application monitoring: workflow and job level



- The portal monitors and visualizes parallel jobs
  (if they are prepared for Mercury monitor)
- Statistical views can also be generated

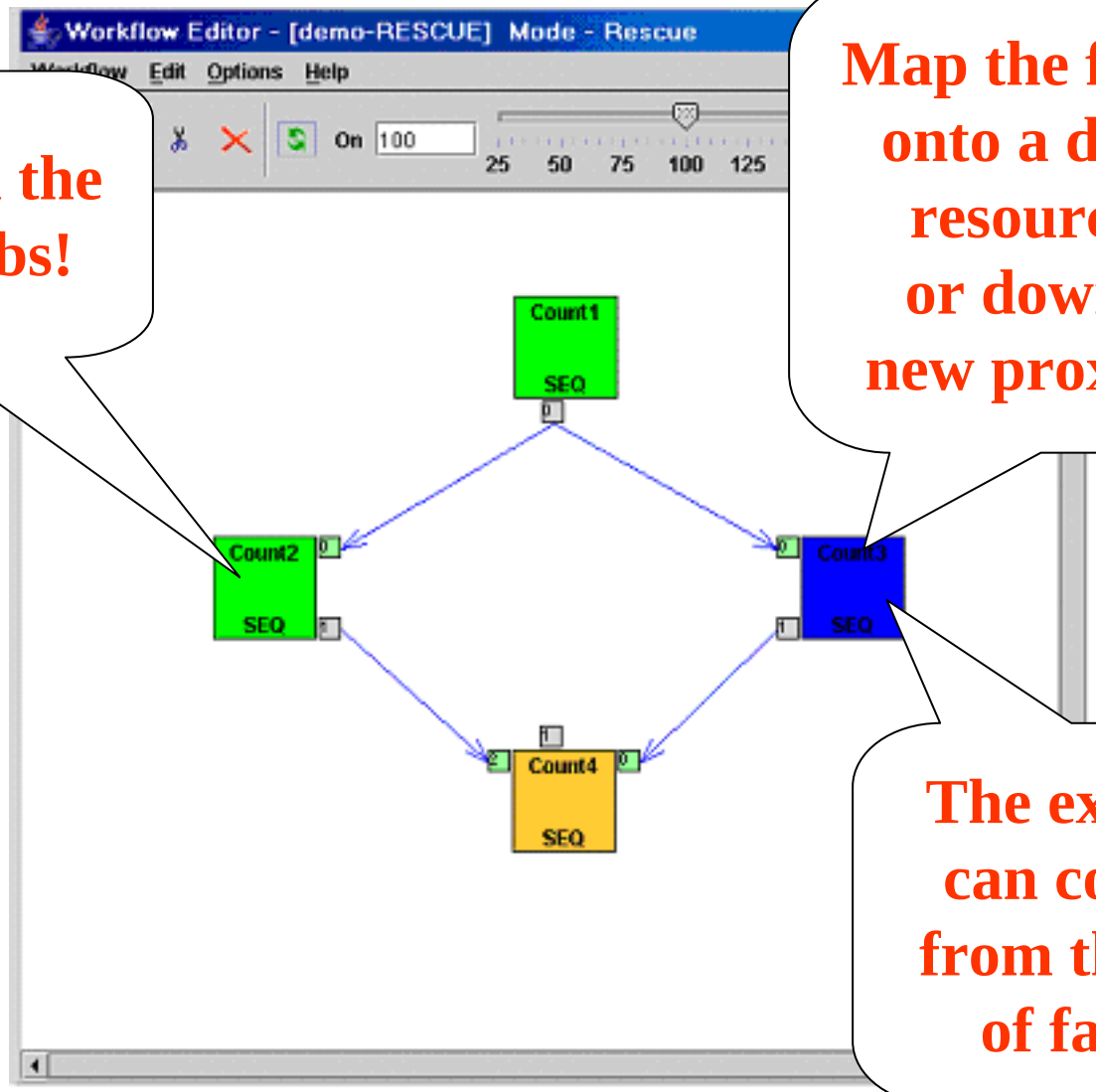
Rescuing a failed workflow 1.



A job failed during workflow execution

Read the error log to know why

Rescuing a failed workflow 2.

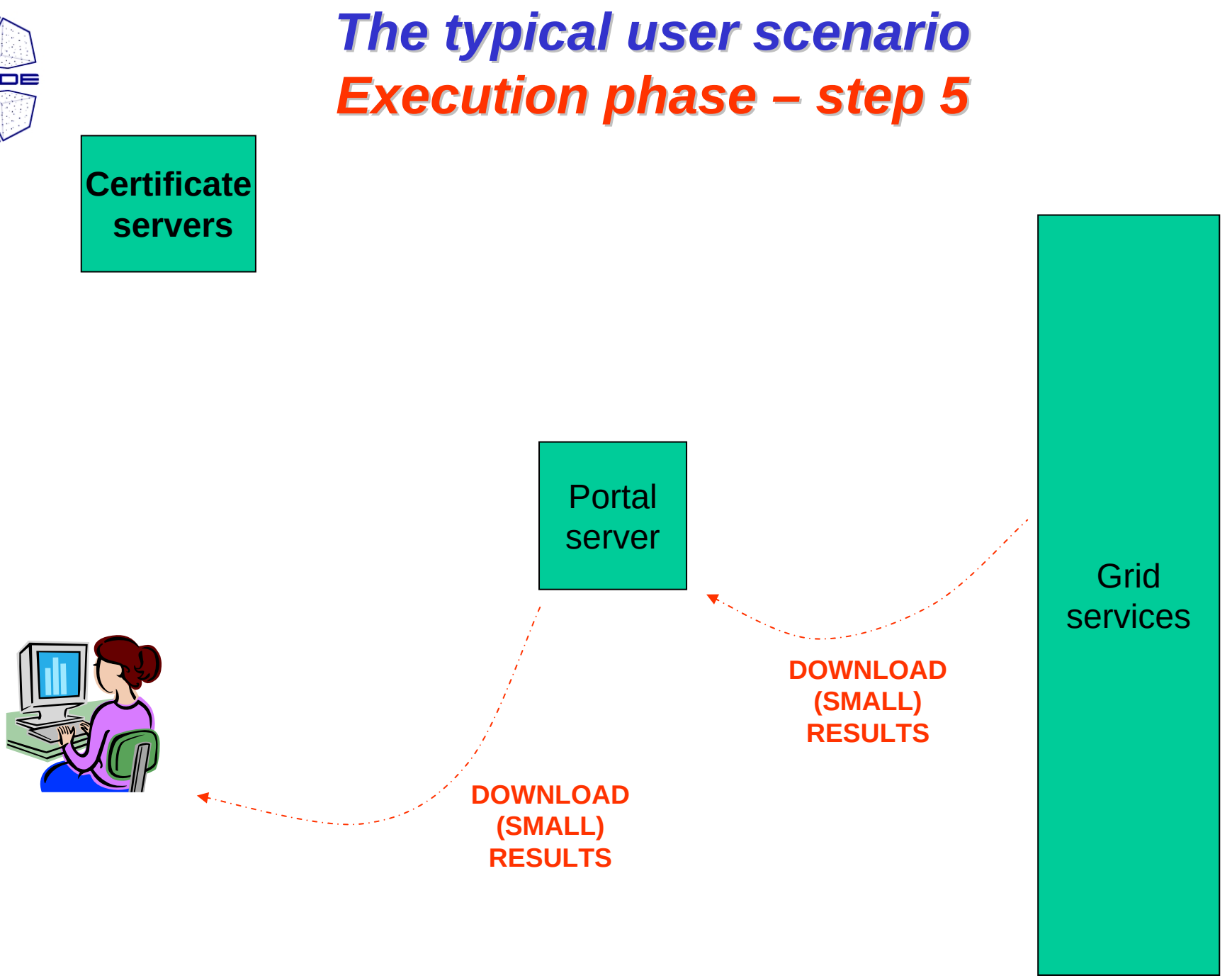


Don't touch the finished jobs!

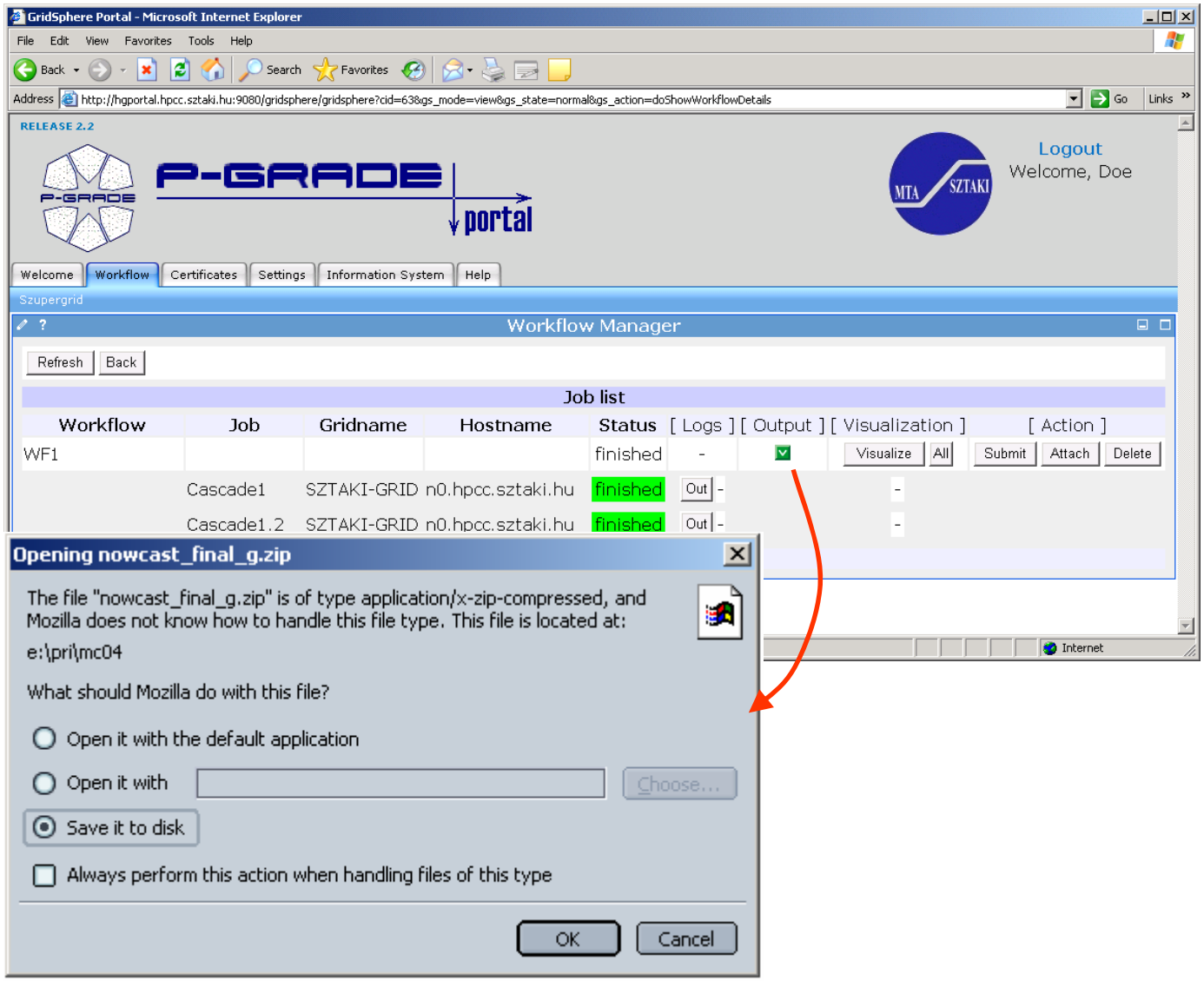
Map the failed job onto a different resource/grid or download a new proxy for it.

The execution can continue from the point of failure!

The typical user scenario
Execution phase – step 5



Downloading the results…



- **Download small files in a single ZIP file**

- **Keep large files in the Grid for future analysis!**

Extra features

- **Workflows and traces** can be exported from the portal server onto your client machine
- **Workflows and traces** can be imported into the Portal

- **Share your workflows or results with other researchers!**
- **Migrate your application from one portal into another!**

References

- P-GRADE Portal service is available for
  - UK National Grid Service (with GEMLCA support)
  - SEE-GRID infrastructure
  - HUNGRID VO of EGEE
  - Central European VO of EGEE
  - GILDA: Training VO of EGEE
- Under preparation for
  - US Open Science Grid, Economy-Grid, Swiss BioGrid, Bio and Biomed EGEE VOs
- P-GRADE portal can be installed for any public/private EGEE or Globus Grid/VO

  Please contact us!

How to get access?

- Take a look at **www.lpds.sztaki.hu/pgportal** *(manuals, training events, client & server requirements, installation procedure, etc.)*
- **Visit or request a training event!** (pgportal@sztaki.hu)
  - Lectures, demos, hands-on tutorials, application development support
- **Get an account for one of its production installations:**
  - NGS portal – University of Westminster } **www.cpc.wmin.ac.uk/gngsportal**
  - VOCE portal - SZTAKI
  - SEEGRID portal – SZTAKI
  - HUNGrid portal – SZTAKI

  } **www.lpds.sztaki.hu/pgportal**
- **If you are the administrator of a Grid/VO then contact SZTAKI to get your own P-GRADE Portal!**
- **If you know the administrator of a P-GRADE Portal you can ask him/her to give access to your Grid through his/her portal installation!** ***(Multi-Grid portal)***

Conclusion: Easy-to-use, technology-neutral Grid portal for e-Scientists

- The P-GRADE Portal hides differences of Grids
  - Globus X – LCG2 – gLite Grid interoperability
  - Transparent switching between Grid technologies
- Graphical tools for application development, execution and monitoring
  - Sequential & parallel components can be integrated into large Grid applications
  - Manual or broker based resource allocation
- Your code does not have to contain grid specific calls
- Support for collaborative research
  - Share workflows
- Built by standard portlet API
  - customizable to specific application areas and user groups

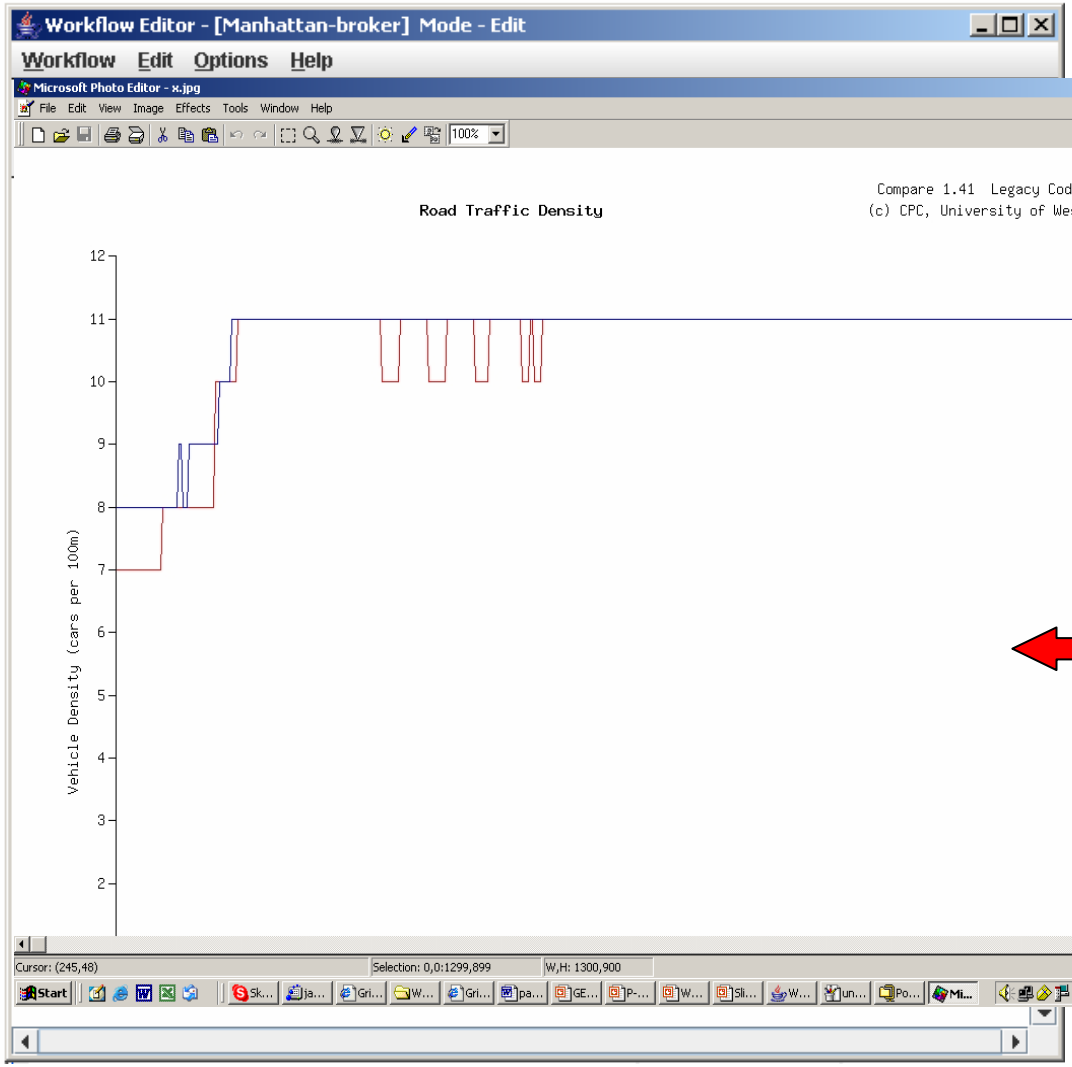
Learn once, use everywhere
Develop once, execute anywhere

Thank you!

www.lpds.sztaki.hu/pgportal
pgportal@lpds.sztaki.hu

Live Demonstration I.

Workflow to analyse road traffic



Manhattan road network generator

Traffic simulators

Analyser