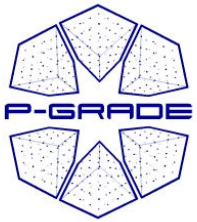


P-GRADE Portal tutorial at EGEE'09

Gergely Sipos
MTA SZTAKI
sipos@sztaki.hu

EGEE Training and Induction
EGEE Application Porting Support

www.lpds.sztaki.hu/gasuc
www.portal.p-grade.hu

Agenda of the morning

- Introduction to workflow concept
- Workflow hands-on

~ **Break**

- Parameter studies
- Parameter study hands-on

- Further information and next steps

Workflow

The automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules to achieve, or contribute to, an overall business goal.

Workflow Reference Model, 19/11/1998

WORKFLOW MANAGEMENT COALITION

www.wfmc.org

- Workflow management system (WFMS) is the software that does it

Why use workflows in Grid?

- Build distributed applications through orchestration of multiple services
  - A single job or a single service is good for nothing…
- Integration of multiple teams involved
  - Collaborative work
- Unit of reusage
  - (E-)science requires traceable, repetable analysis
- (Typically) ease of use grids
  - Graphical representation

Grid Workflow definition examples

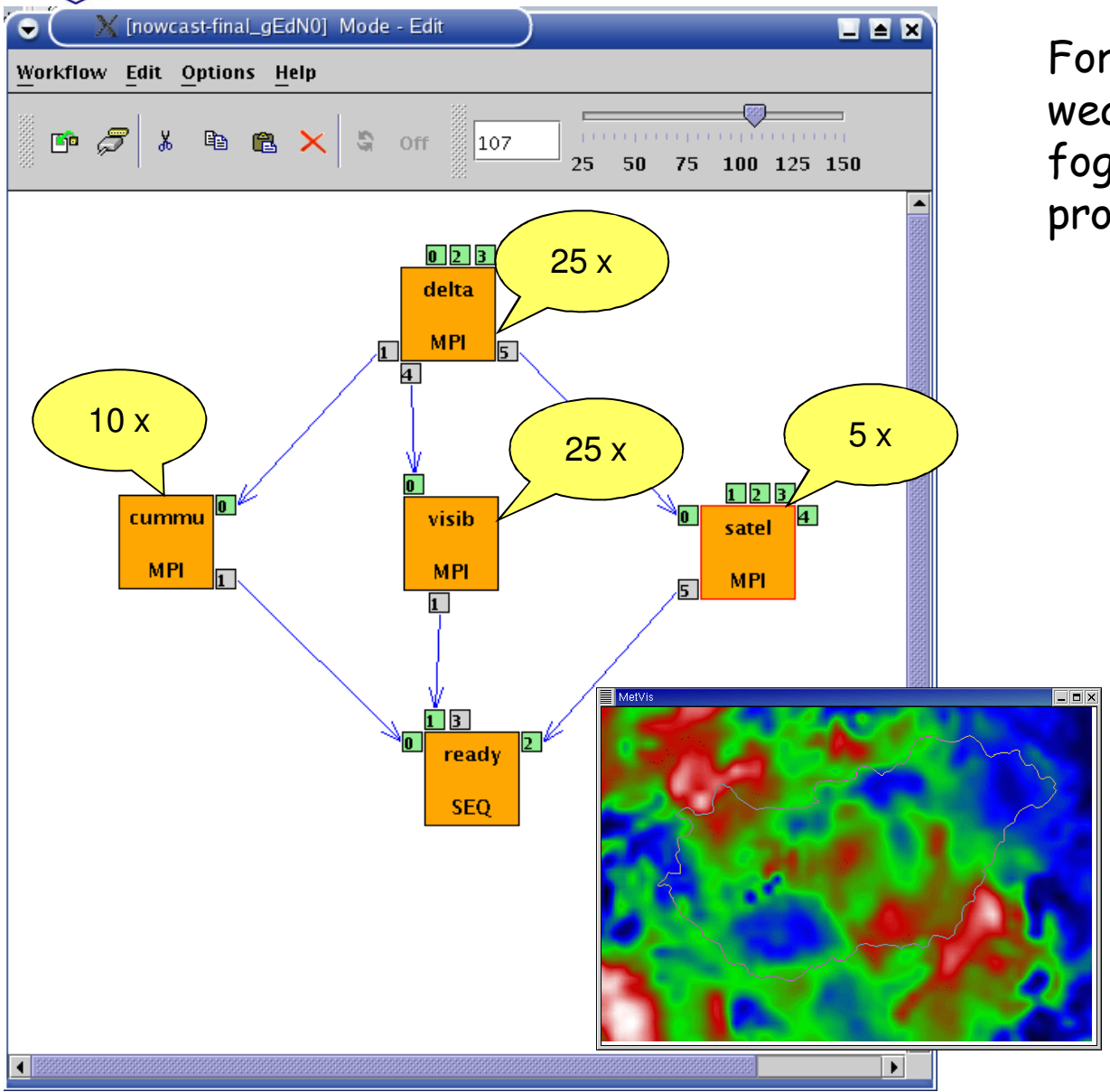
Grid workflow can be defined as the composition of grid application services which execute on heterogeneous and distributed resources in a well-defined order to accomplish a specific goal.

R. Buyya

The automation of the processes, which involves the orchestration of a set of Grid services, agents and actors that must be combined together to solve a problem or to define a new service.

Geoffrey Fox [GGF 10]

Example: Ultra-short range weather forecast with P-GRADE Portal



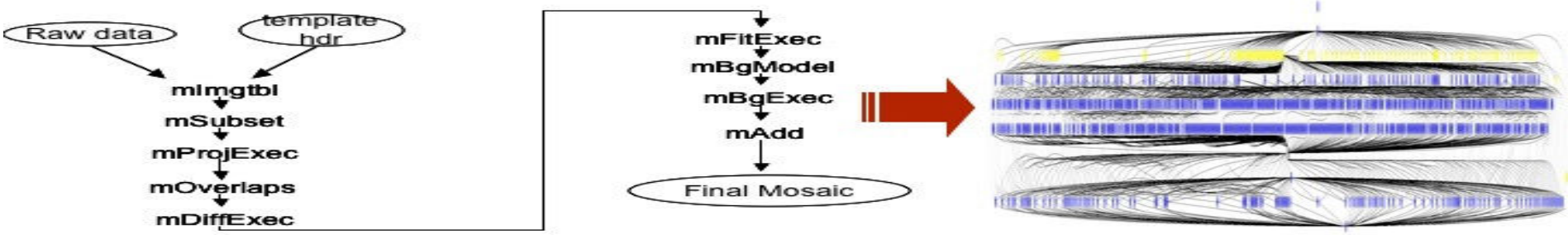
Forecasting dangerous weather situations (storms, fog, etc.), crucial task in the protection of life and property

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

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

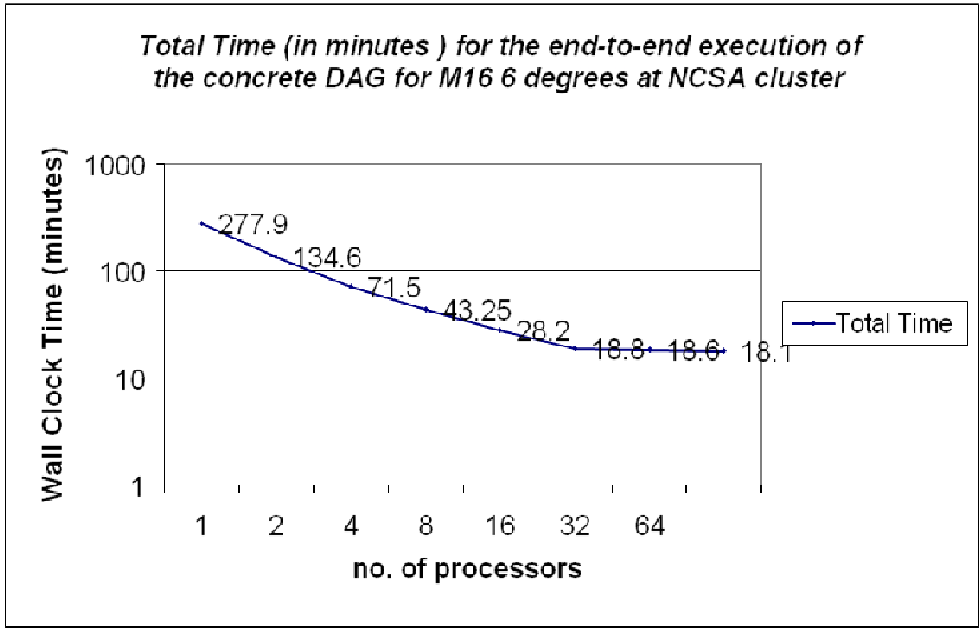
Execution on a GT2 based Hungarian Grid

Example: Montage workflow with Pegasus (and DAGMan)



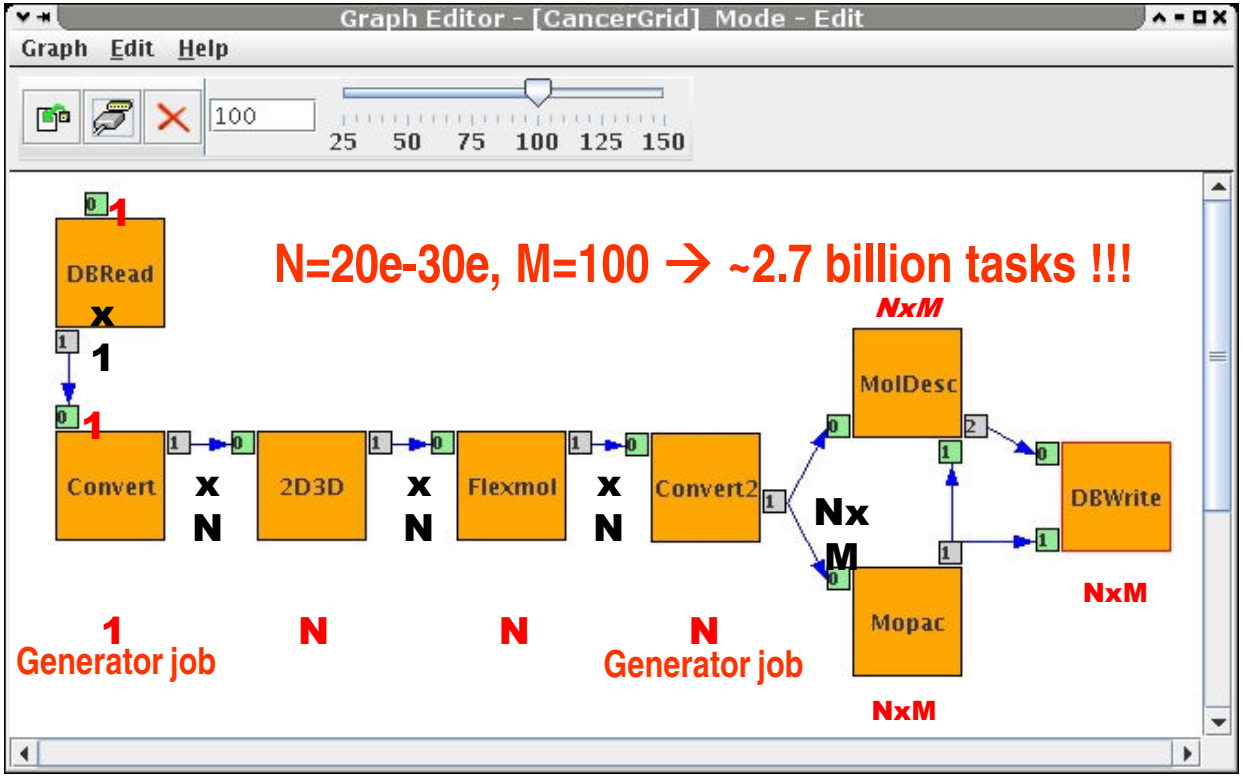
Tasks run on NSF's TeraGrid

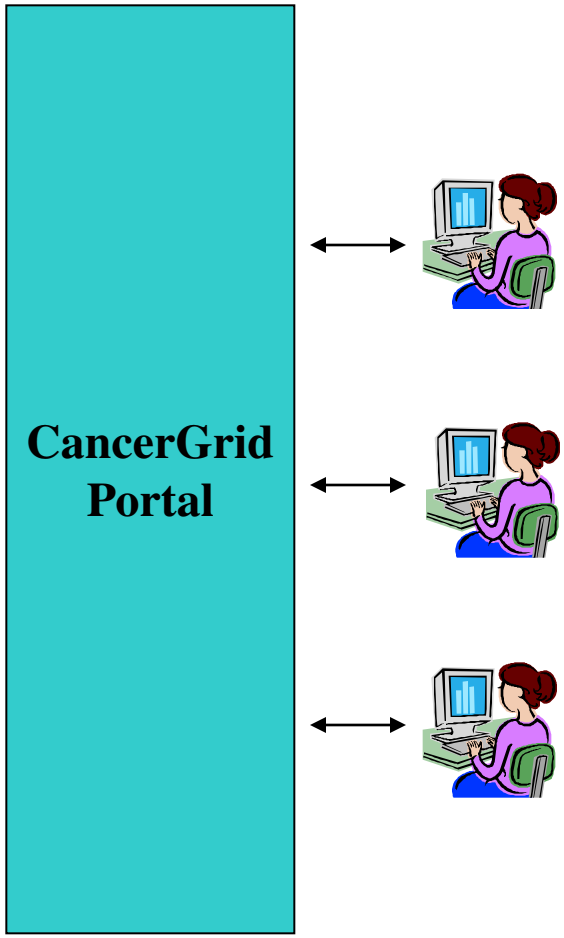
Montage application
~7,000 compute jobs in instance
~10,000 nodes in the executable workflow
same number of clusters as processors
speedup of ~15 on 32 processors



Pegasus: a Framework for Mapping Complex Scientific Workflows onto Distributed Systems, Ewa Deelman, Gurmeet Singh, Mei-Hui Su, James Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Karan Vahi, G. Bruce Berriman, John Good, Anastasia Laity, Joseph C. Jacob, Daniel S. Katz, Scientific Programming Journal, Volume 13, Number 3, 2005

Example: CancerGrid workflow with gUSE (and WS-PGRADE)



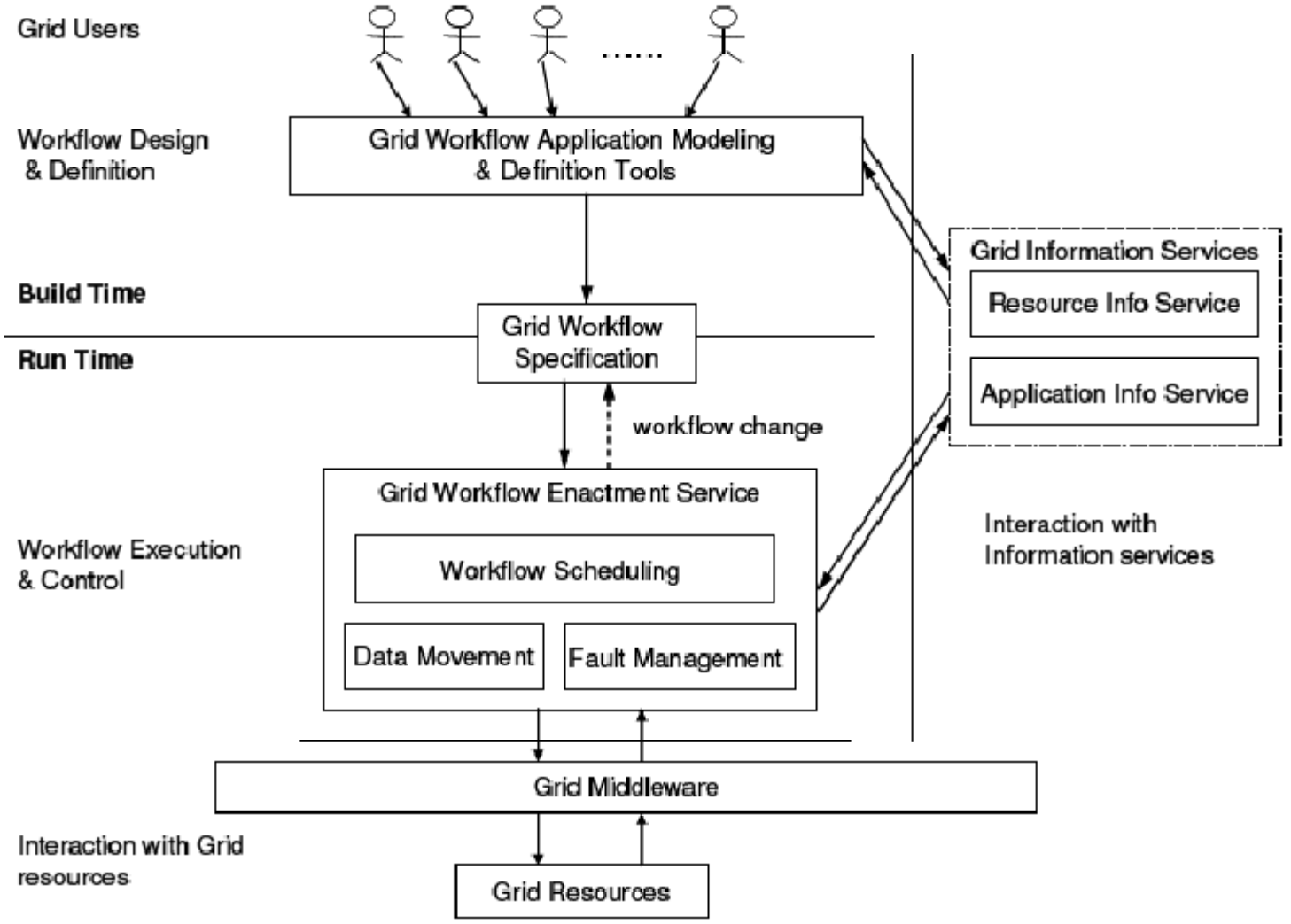


Workflow is hidden from end users
Tasks run on Desktop Grids and RDBMS

CancerGrid

http://www.cancergrid.eu/

Grid WFMS

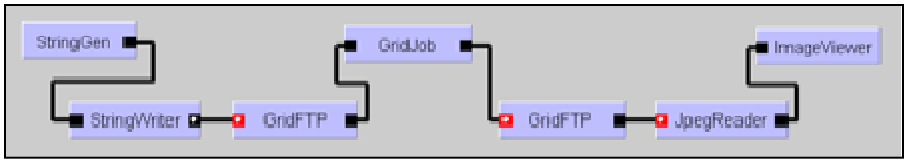


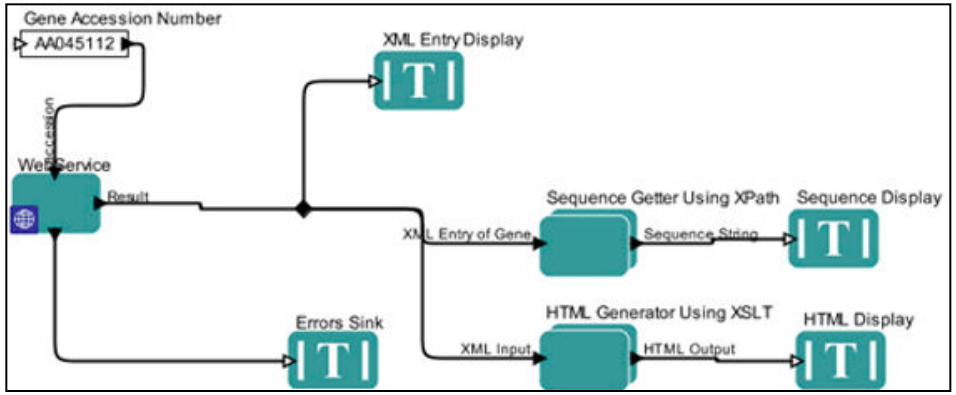
Source: Jia Yu and Rajkumar Buyya: A Taxonomy of Workflow Management Systems for Grid Computing, Journal of Grid Computing, *Volume 3, Numbers 3-4 / September, 2005*

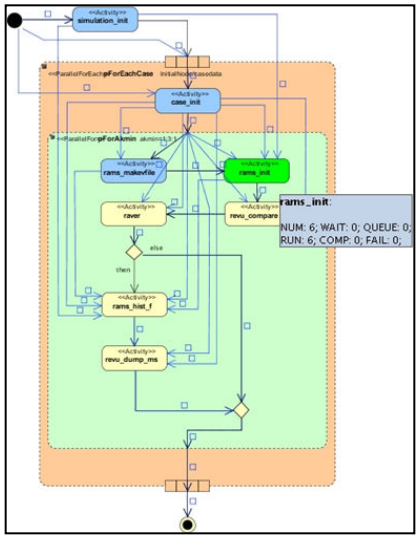
What does a typical Grid WFMS provide?

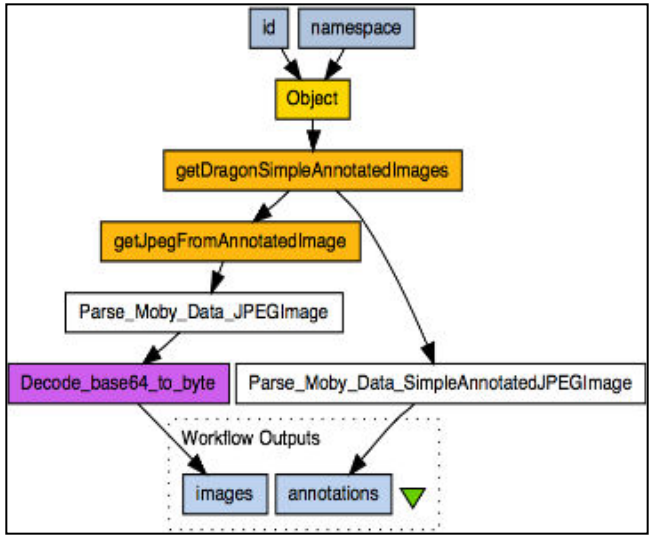
- A level of abstraction above grid processes
  - gridftp, lcg-cr, lfc-mkdir, ...
  - condor-submit, globus-job-run, glite-wms-job-submit, ...
  - lcg-infosites, ...
- A level of abstraction above „legacy processes”
  - SQL read/write
  - HTTP file transfer
  - ...
- Automated mapping and execution of tasks grid resources
  - Submission of jobs
  - Invocation of (Web) services
  - Manage data
  - Catalog intermediate and final data products
- Improve successful application execution
- Improve application performance
- Provide provenance tracking capabilities

What does a typical grid workflow consist of?



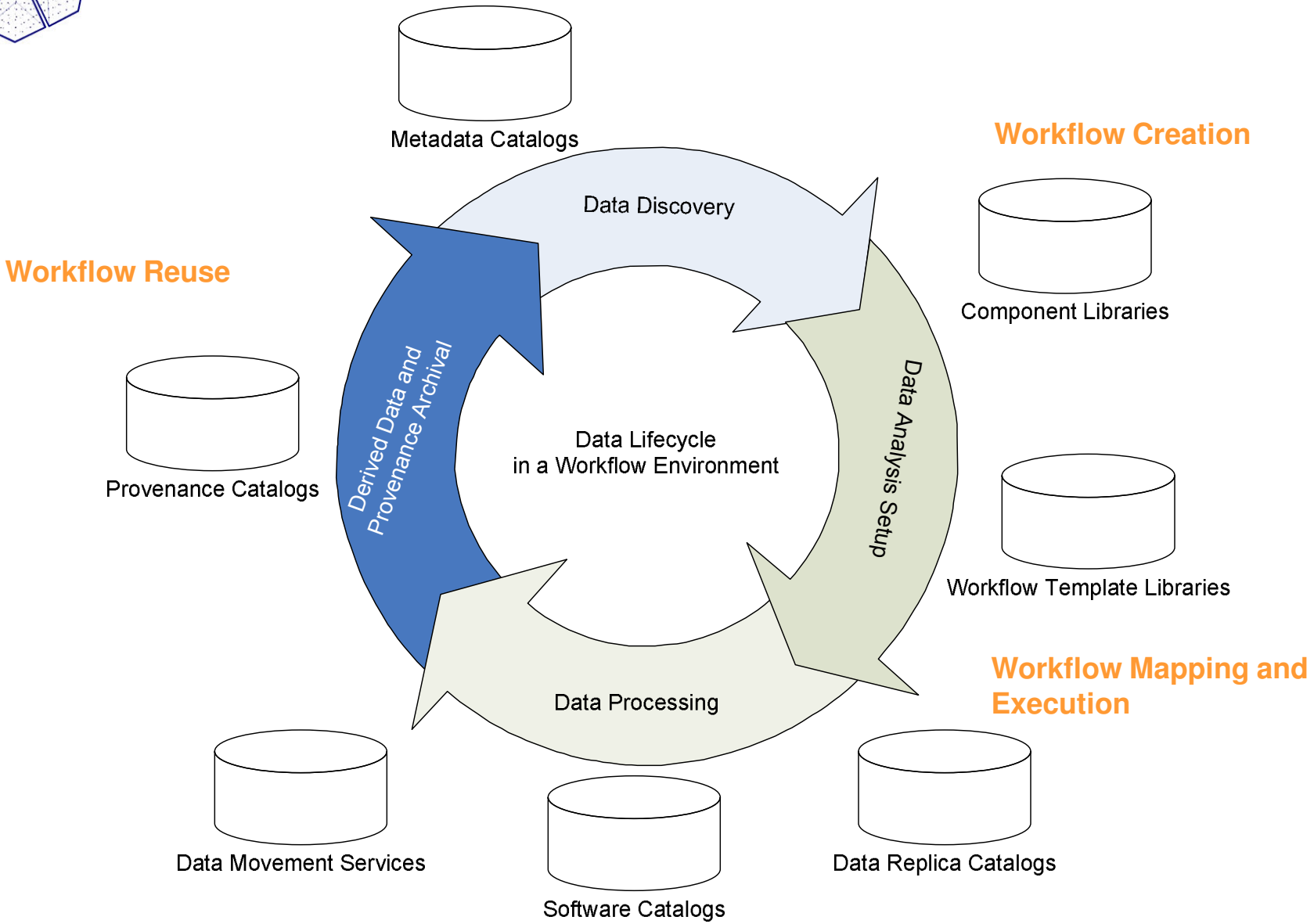




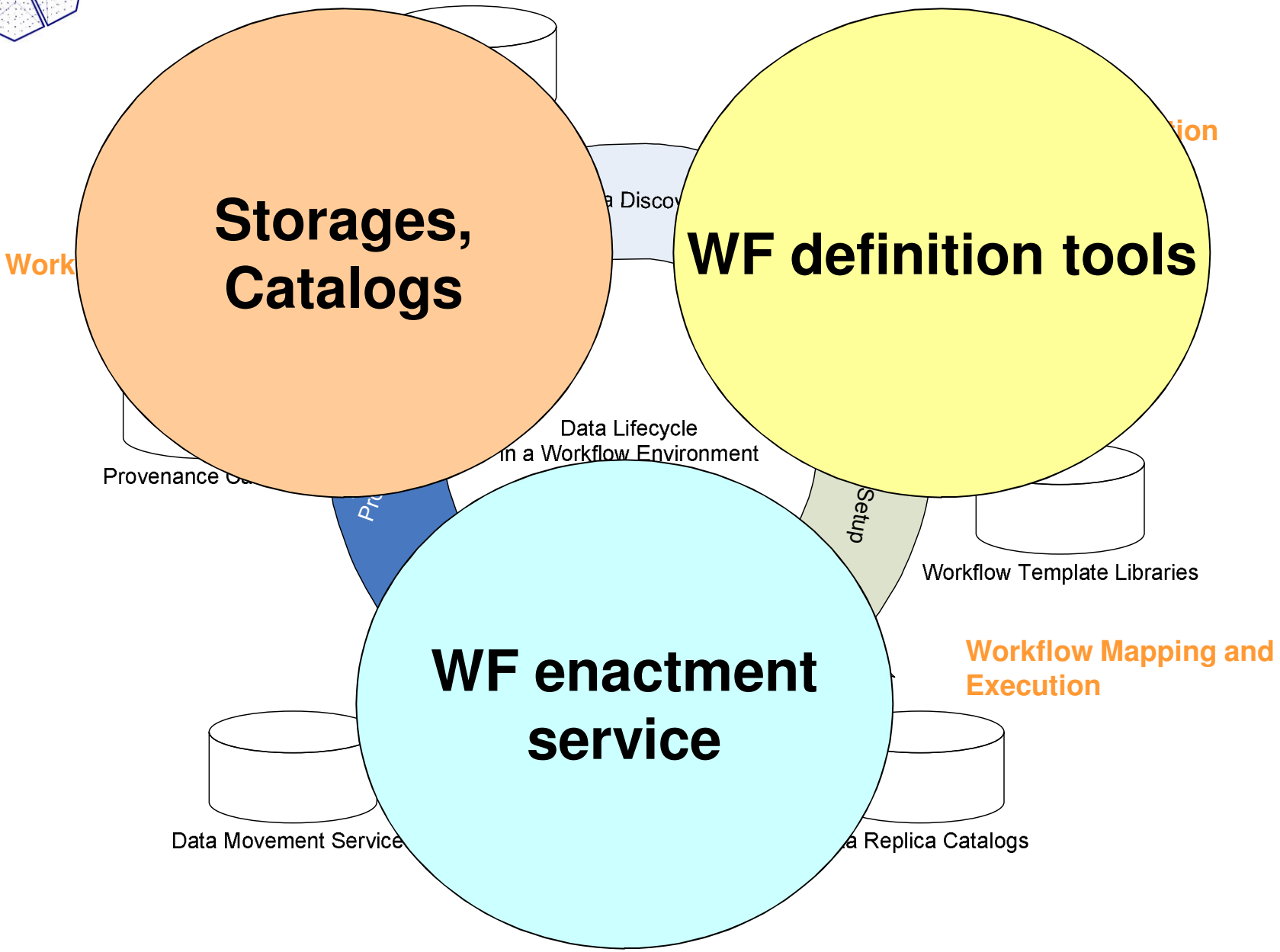


- Dataflow graph
- Activities
  - Definition of Jobs
  - Specification of services
- Data channels
  - Data transfer
  - Coordination
- Cyclic (DAG) /acyclic
- Conditional statements

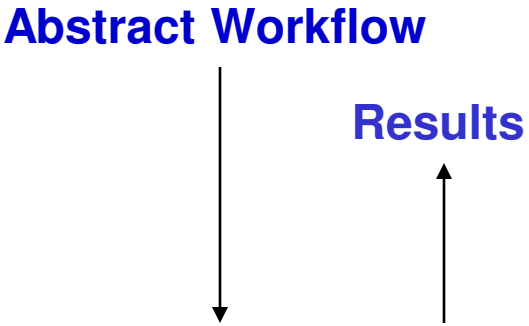
Data lifecycle in workflows



User interaction



Layered architecture of WFMS



WF optimizer e.g. Pegasus Mapper
WF scheduler e.g. Condor DAGMan
Grid scheduler e.g. Condor Schedd

A decision system that develops strategies for reliable and efficient execution in a variety of environments

Reliable and scalable execution of ***dependent*** tasks

Reliable, scalable execution of ***independent*** tasks (locally, across the network), priorities, scheduling

Cyberinfrastructure: Cluster, Condor pool, OSG, EGEE, TeraGrid

(Some of the) available grid workflow systems

http://www.gridworkflow.org

Categories for

- Composition tools
- Description languages
  - Scientific
  - Industrial
  - Formalism
- Engines

Some relevant tools for ARC, gLite, Globus, UNICORE grid users

- Condor DAGMan
  - Used as an enactor in P-GRADE Portal, Pegasus, …
  - Uses DAGMan WF language (DAG = Directed Acyclic Graph)
- MOTEUR
  - Interfaced with “pilot job” framework on EGEE (pull style job execution)
  - Uses SCUFL WF language
- gLite WMS
  - Describe workflows in JDL
  - Share Input-Output sandboxes with multiple jobs
- Taverna
  - Mainly for cluster computing
  - ARC interface is available by Lubeck University
- …

P-GRADE Portal

A Grid WFMS

www.portal.p-grade.hu

Short History of P-GRADE portal

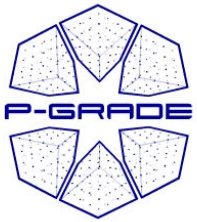
- Parallel Grid Application Development Environment
- Initial development started in the Hungarian SuperComputing Grid project in 2003
- It has been continuously developed since 2003
  - Around 30 manyear development + training + user support
- Detailed information: http://portal.p-grade.hu/
- Open Source community development since January 2008: https://sourceforge.net/projects/pgportal/

- Current version: 2.8

Current P-GRADE Portal related projects

- **GGF GIN** (Since 2006)
  - Providing the GIN Resource Testing portal
- **EU EGEE-II, EGEE-III** (2006-2010)
  - Tool recommended for application development
  - Intensively used in new users' training
- **EU SEE-GRID-SCI** (2008-2010)
  - Interfacing to DSpace-based workflow storage
  - Infrastructure testing workflows
- **EU CancerGrid** (2007-2009)
  - Development of new generation P-GRADE (gUSE and WS-PGRADE)
  - Integration with desktop grids
- **EU EDGeS** (2008-2009)
  - Transparent access to Desktop Grid systems

Portal installations



P-GRADE Portal services:

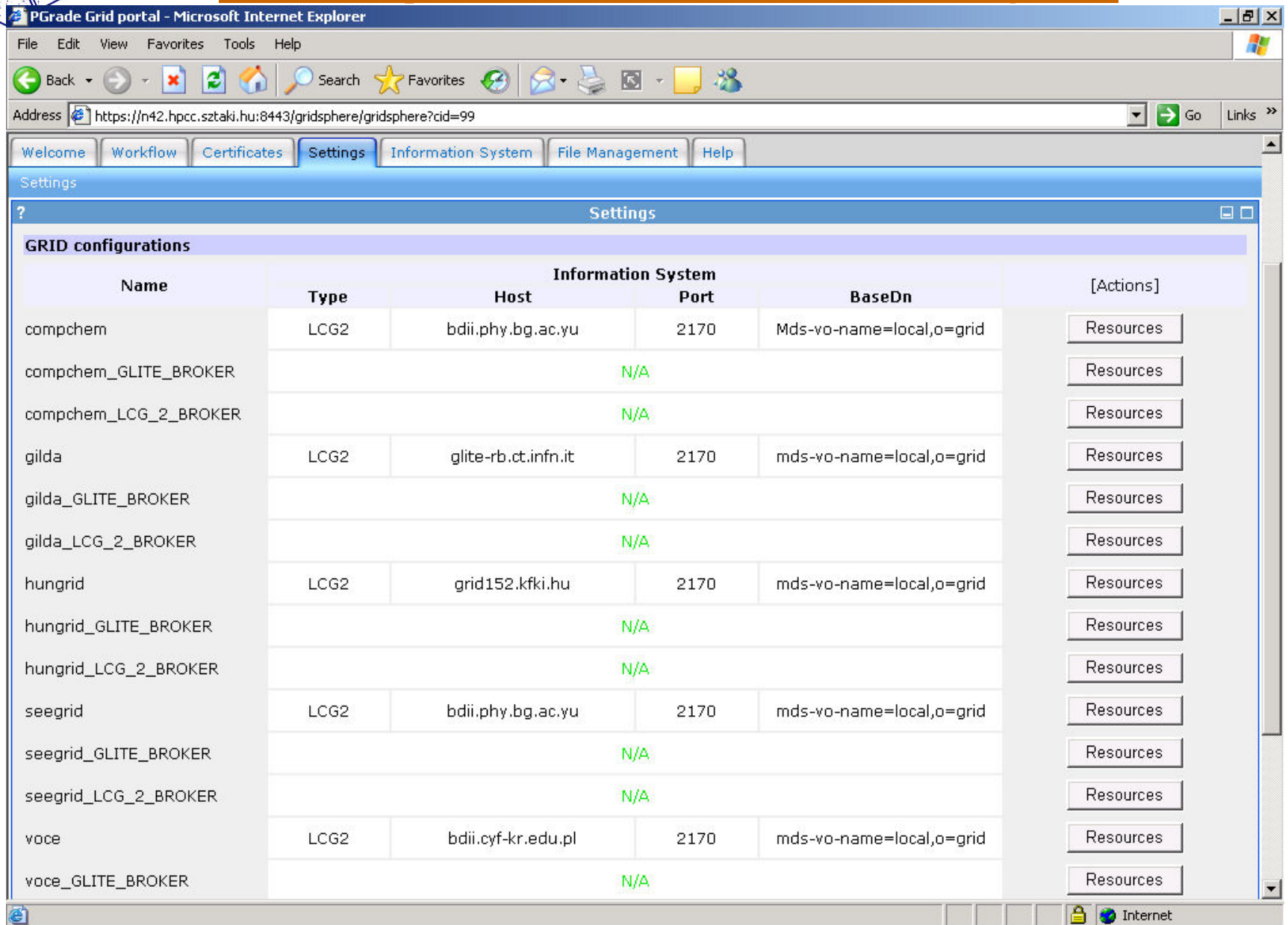
- SEE-GRID infrastructure
- Several VOs of EGEE:
  - Biomed, Astronomy, Central European, NA4,...
- GILDA: Training VO of EGEE
- Many national Grids (UK National Grid Service, HunGrid, Turkish Grid, etc.)
- US Open Science Grid, TeraGrid
- OGF Grid Interoperability Now (GIN) VO
- …

Portal services and account request:

http://portal.p-grade.hu/index.php?m=3&s=0

Account request form on portal login page

Multi-Grid portal installation: www.lpds.sztaki.hu/multi-grid

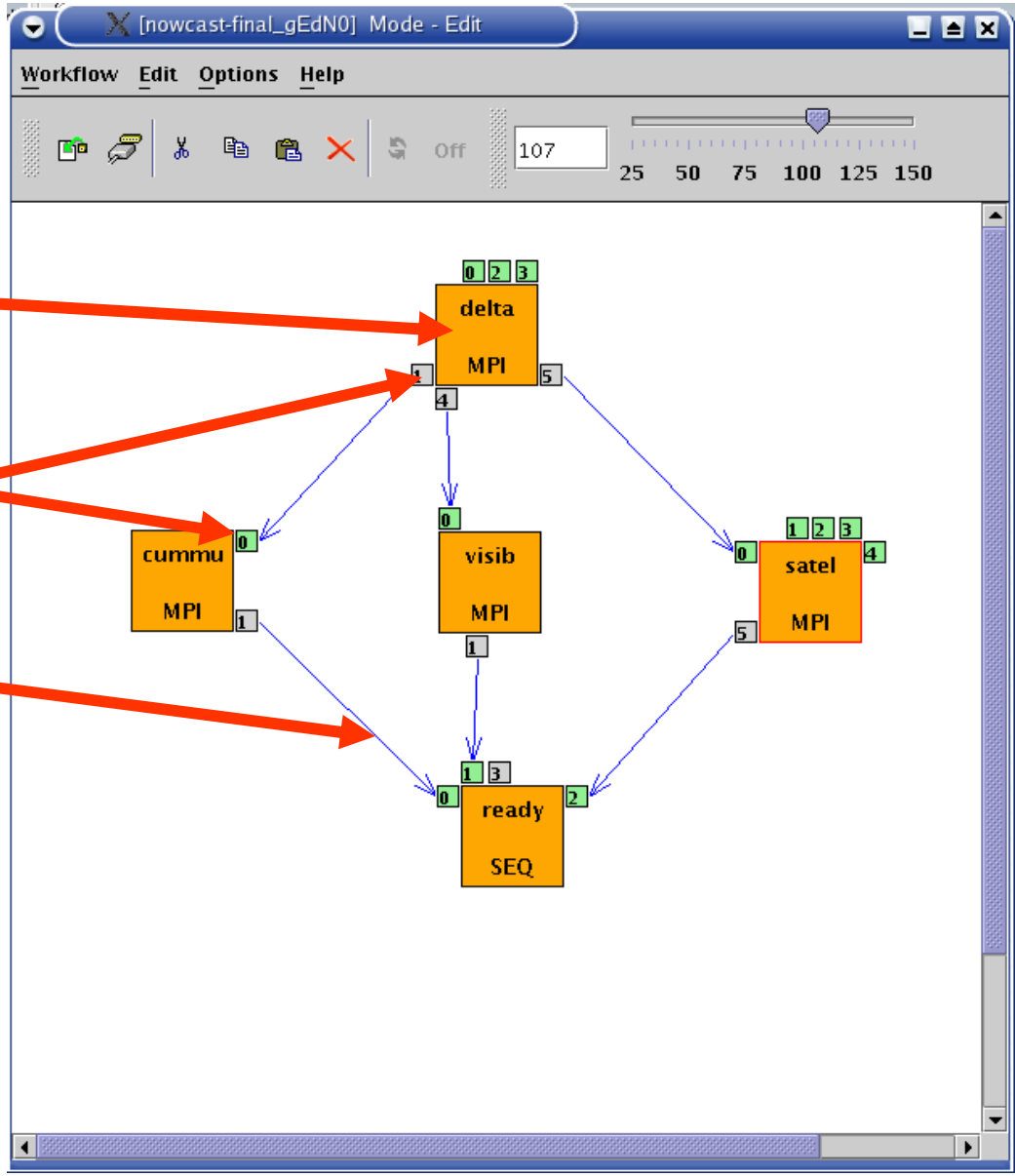


Design principles of P-GRADE portal

- **P-GRADE Portal is not only a user interface**, it is a
  - General purpose
  - Workflow-level
  - Multi-Grid
  - Application Development and Execution Environment
- **P-GRADE Portal includes a high-level middleware layer** for orchestrating jobs on grid resources
  - inside a grid
  - among several different grids (and several VOs)
- **P-GRADE Portal is grid-neutral**:
  - Unlike many existing grid portals it is not tailored to any particular grid type
  - Can be connected to various grids based on different grid middleware
    - LCG-2, gLite, GT2, GT4, ARC, Unicore, etc.
  - Implements the high-level grid middleware services on top of the existing grid middleware services
  - The workflow interface is the same no matter which type of grid is connected to it

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

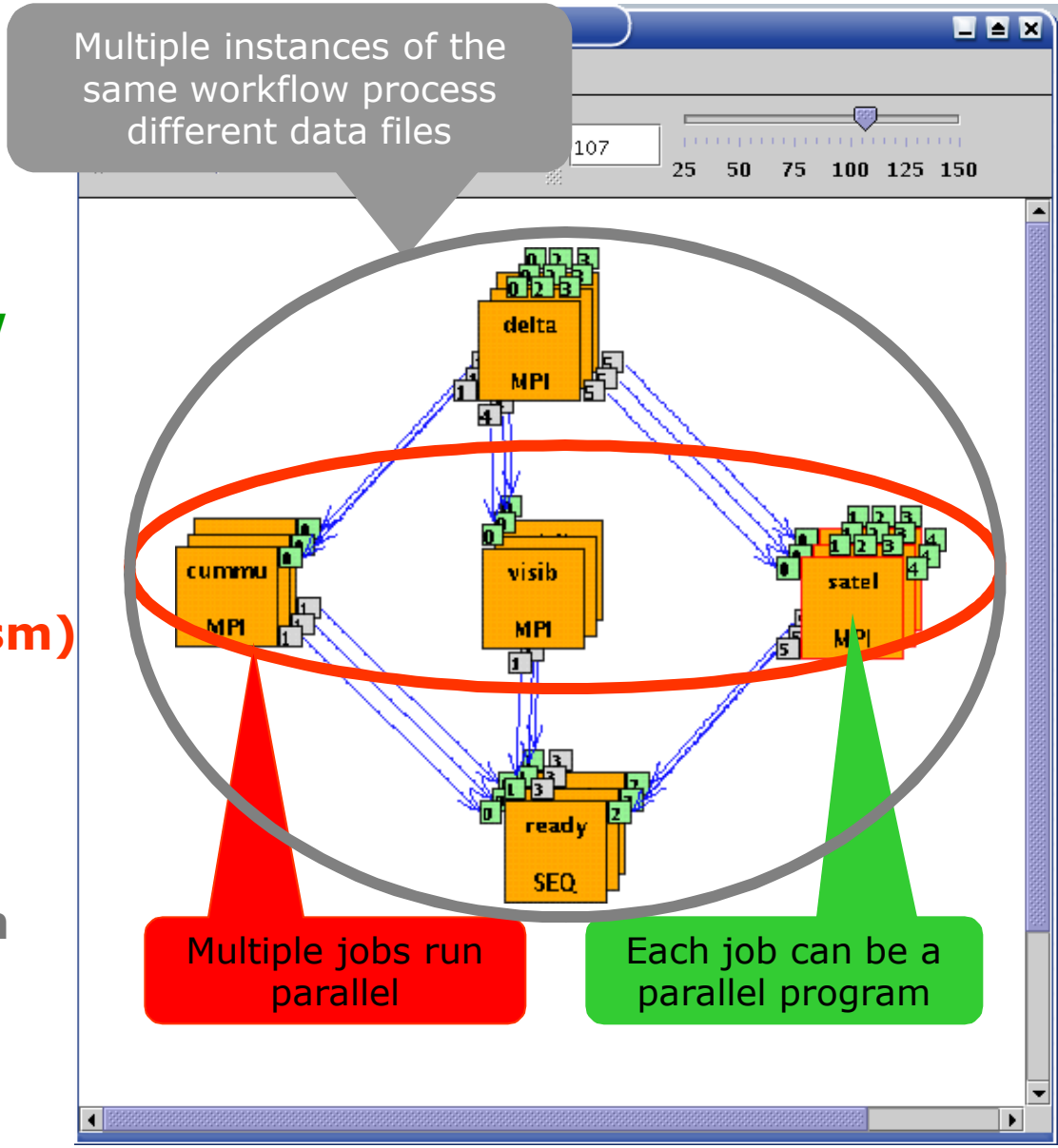


Three levels of parallelism

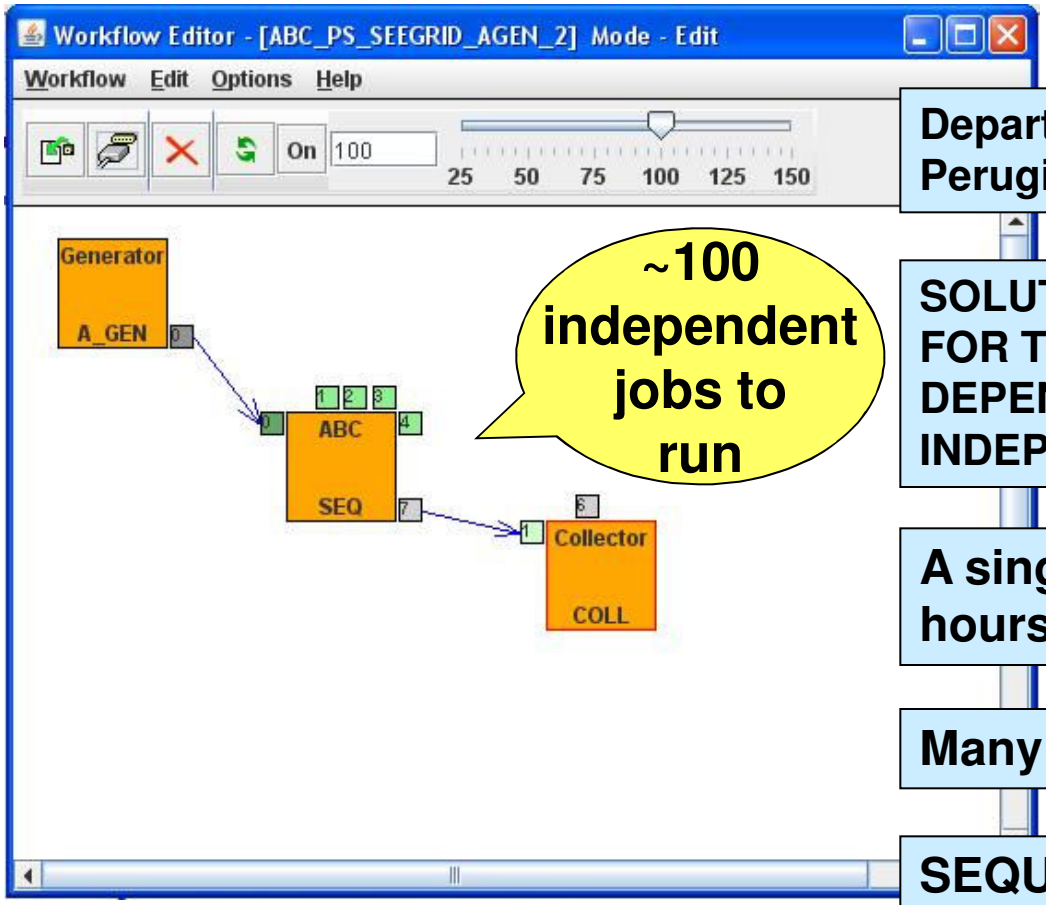
– **Job level: Parallel execution inside a workflow node (MPI job as workflow component)**

– **Workflow level: Parallel execution among workflow nodes (WF branch parallelism)**

– **PS workflow level: Parallel execution of the workflow**



Example: Computational Chemistry



Department of Chemistry, University of Perugia

SOLUTION OF SCHRODINGER EQUATION FOR TRIATOMIC SYSTEMS USING TIME-DEPENDENT (RWAVEPR) OR TIME INDEPENDENT (ABC) METHOD

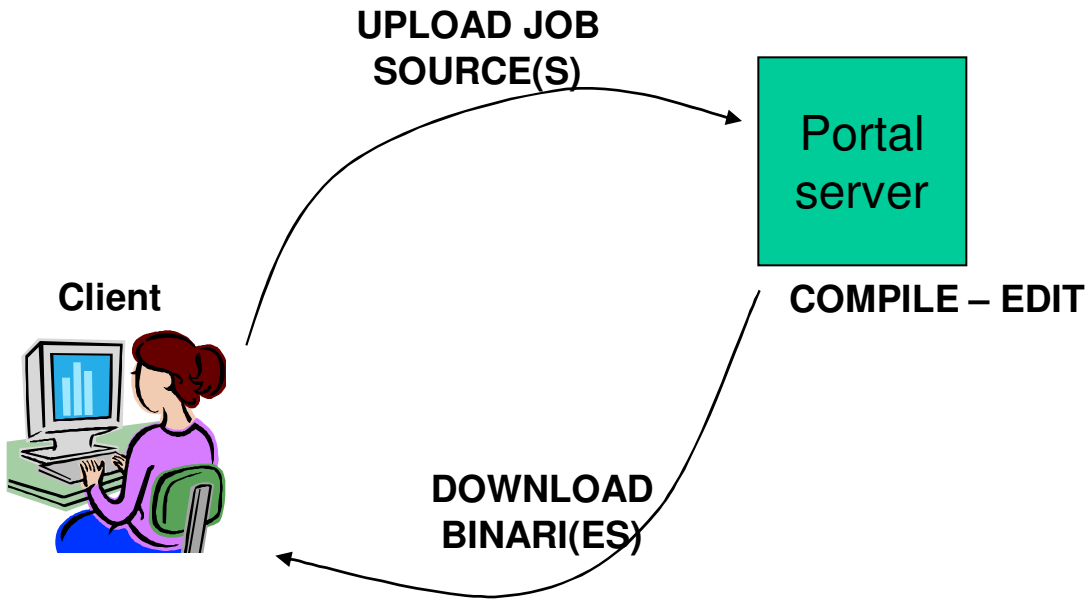
A single execution can be between 5 hours and 10 hours

Many simulations at the same time

SEQUENTIAL FORTRAN 90

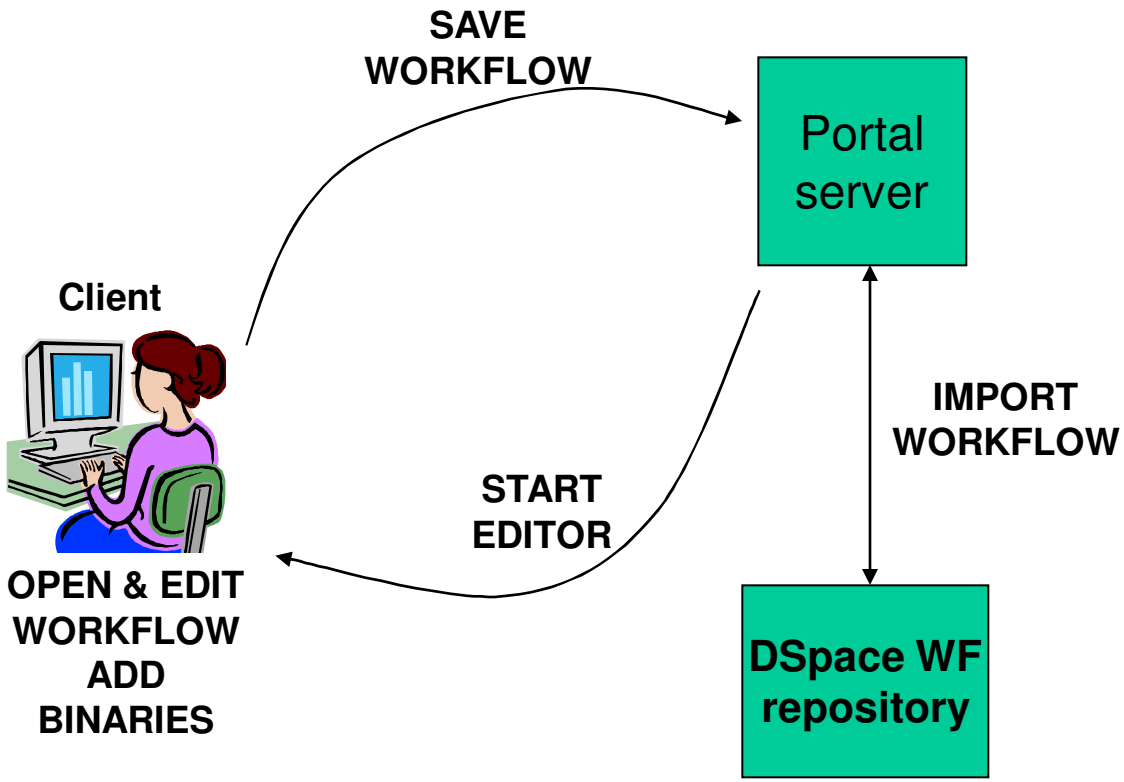
Full story: EGEE Grid Application Porting Support - http://www.lpds.sztaki.hu/gasuc/index.php?m=7&s=3

Typical user scenario
Job compilation phase

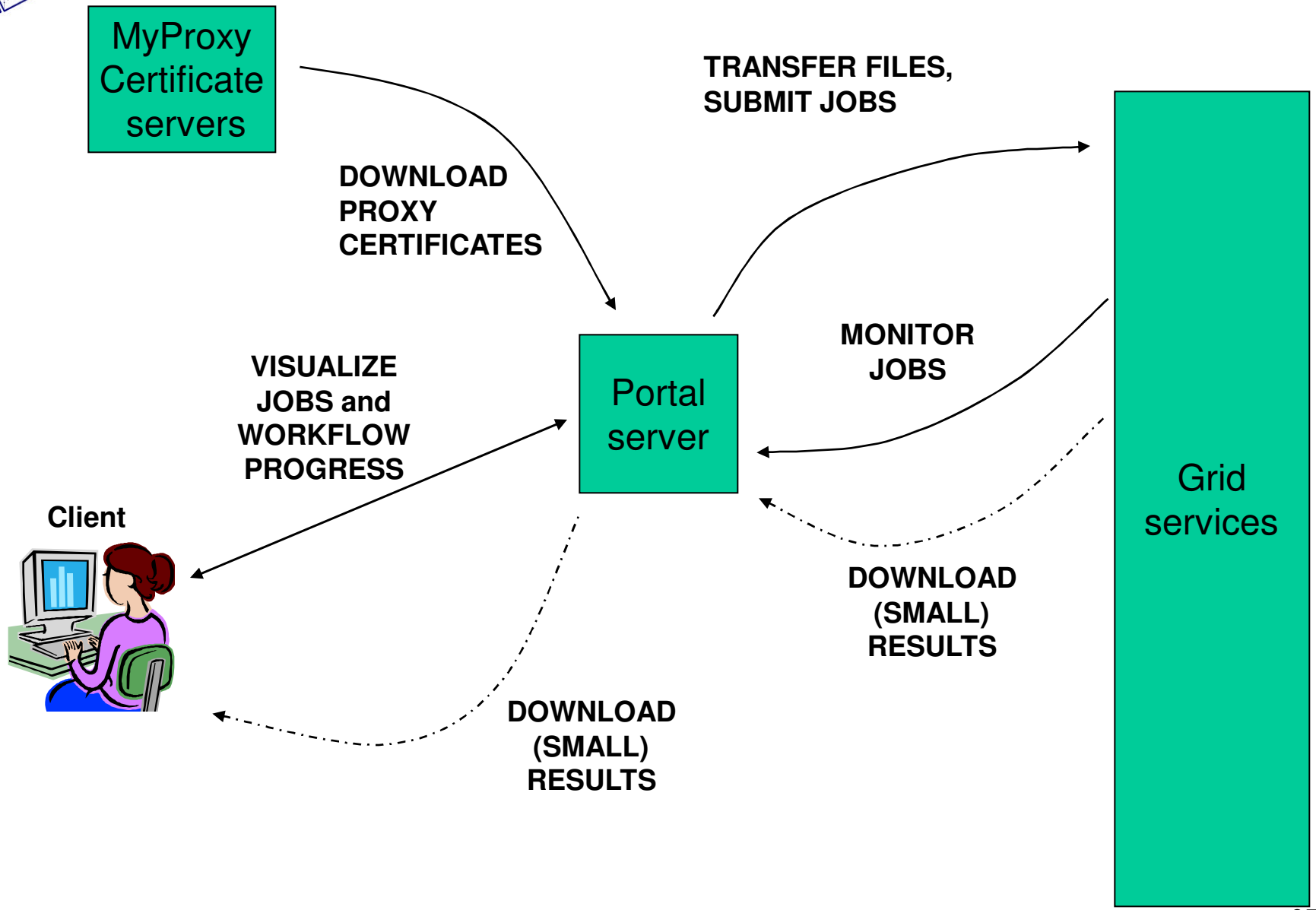


Grid services

Typical user scenario
Workflow development phase

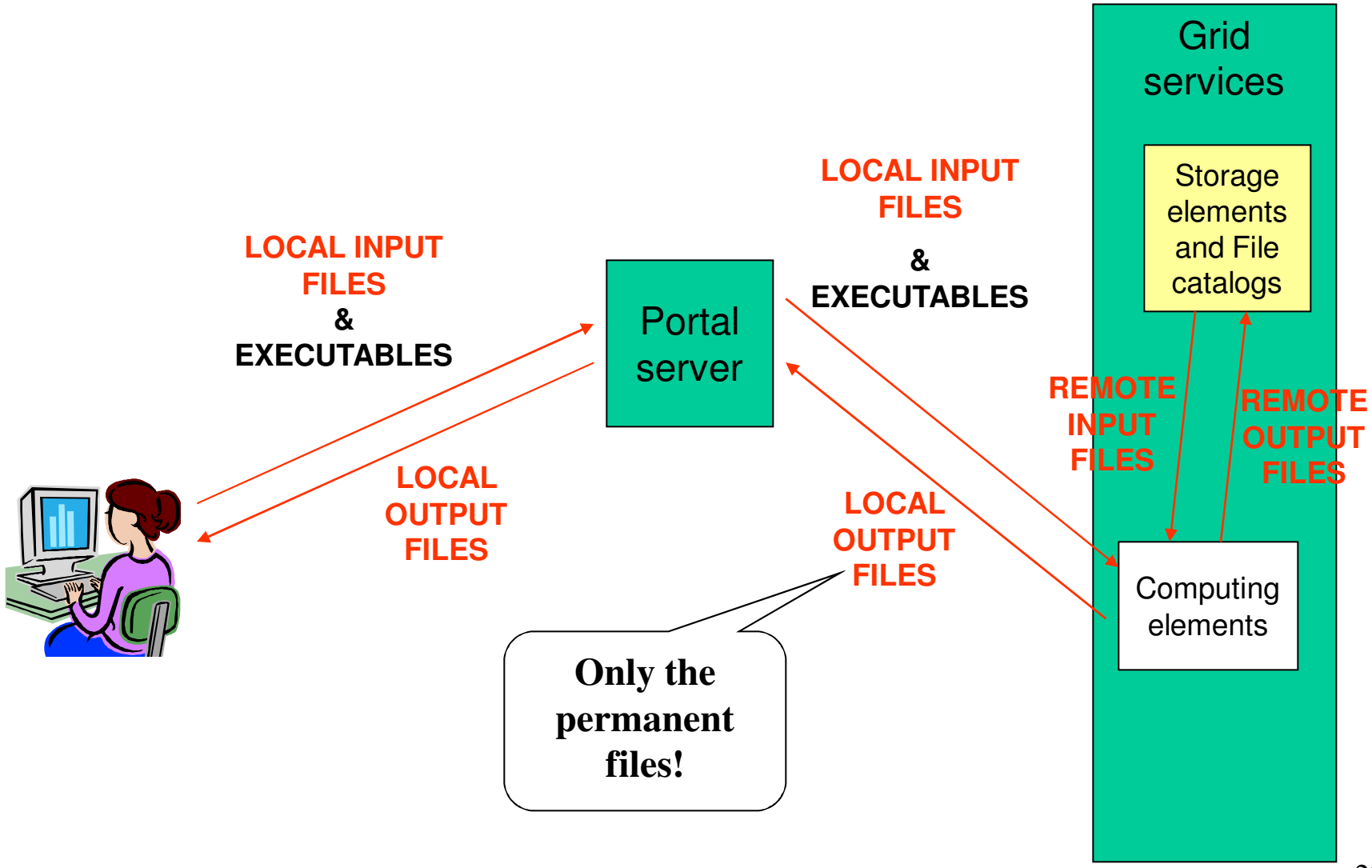


Typical user scenarios
Workflow execution phase

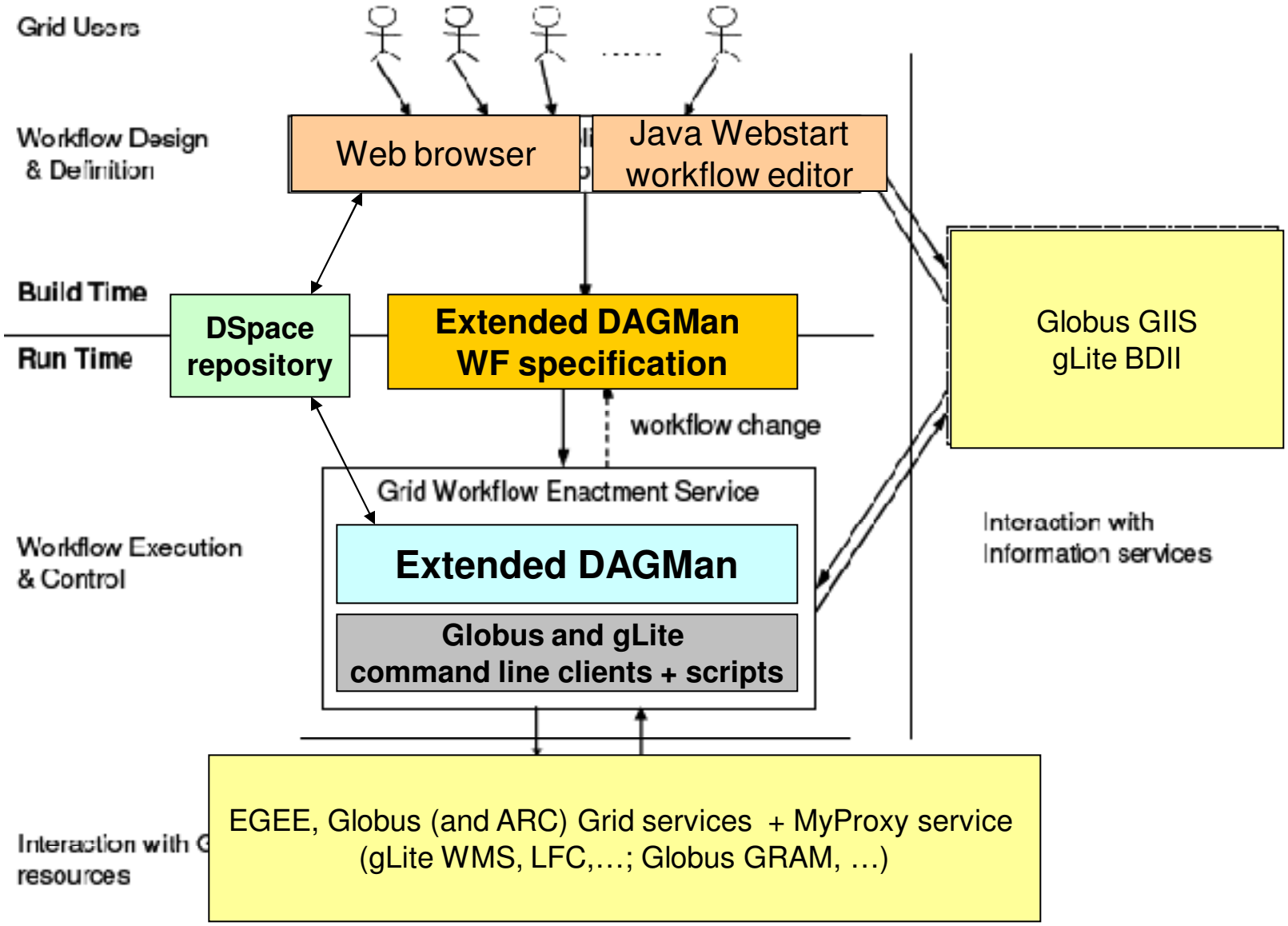


Accessing local and remote files

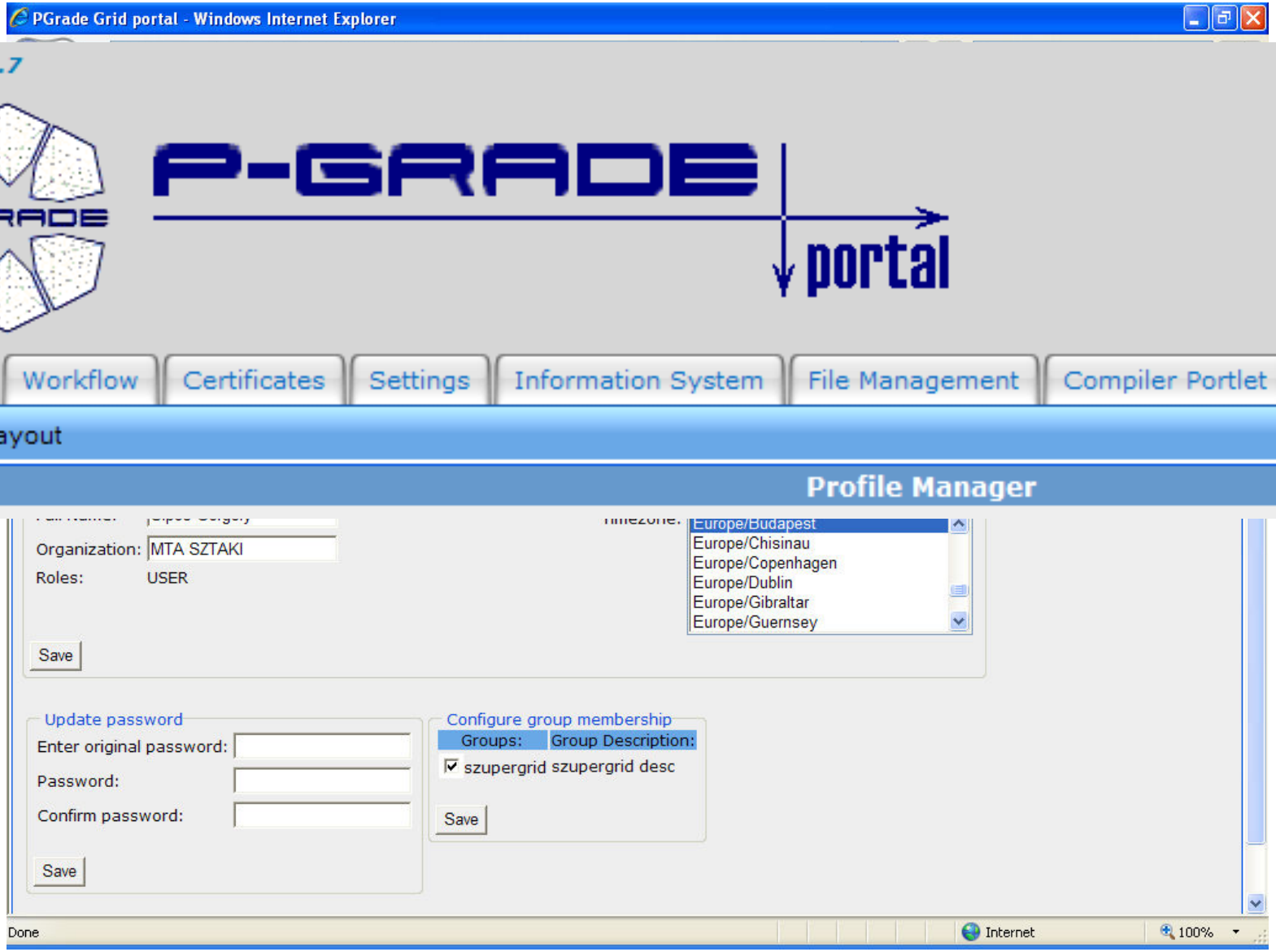
Use legacy executables with Grid files without touching the code



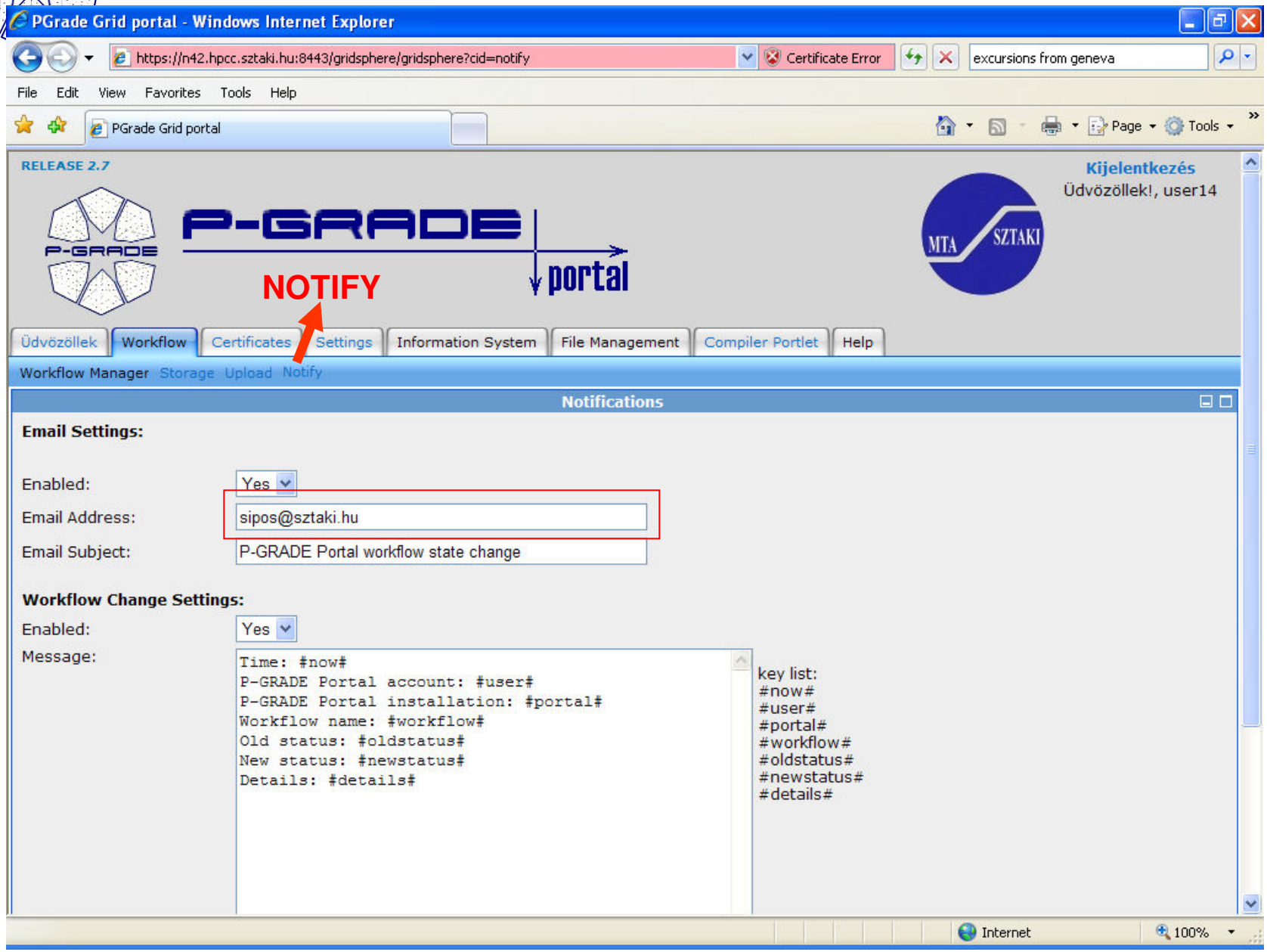
P-GRADE Portal structural overview



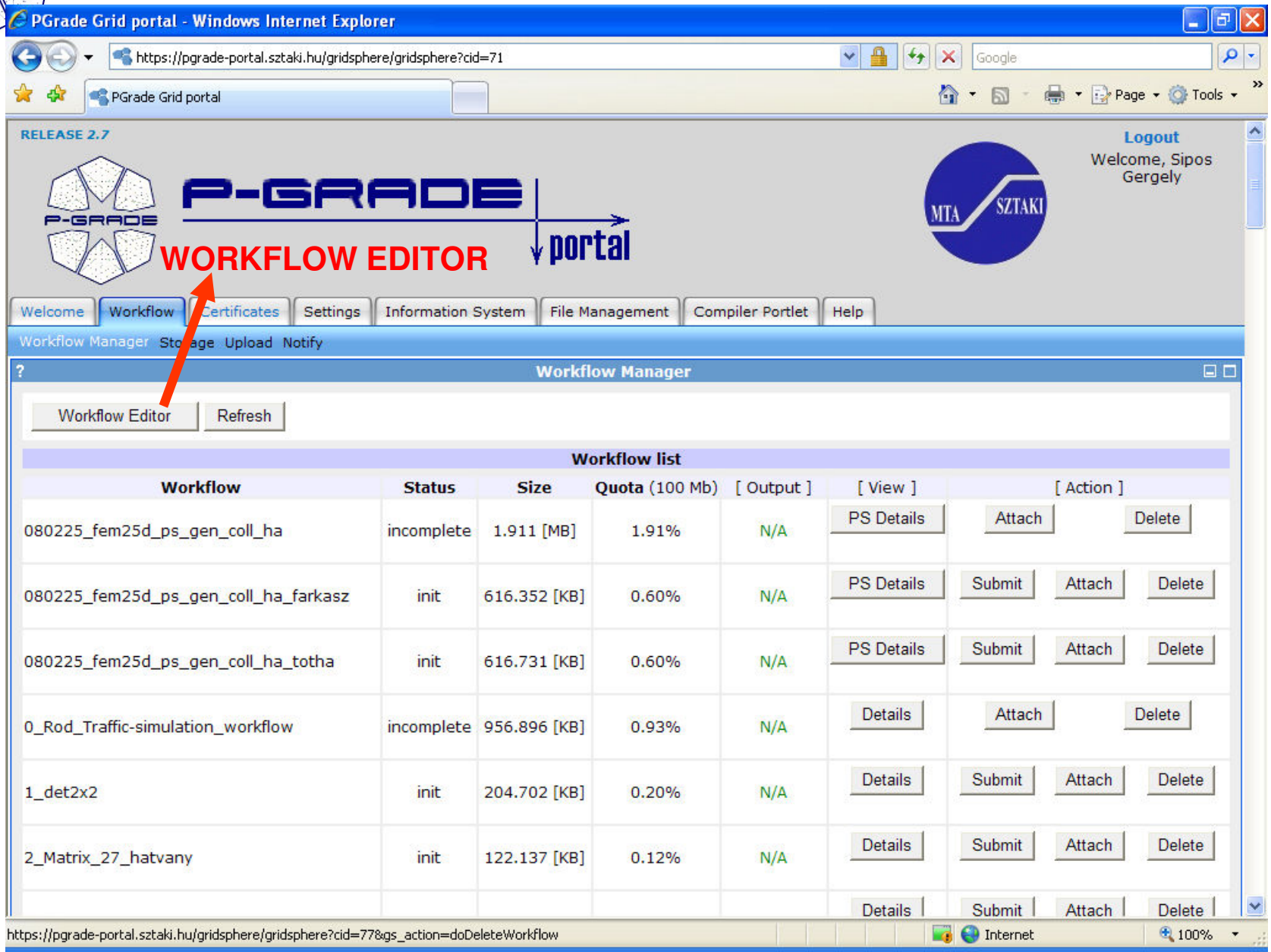
Web interface - Portlets



Email notifications



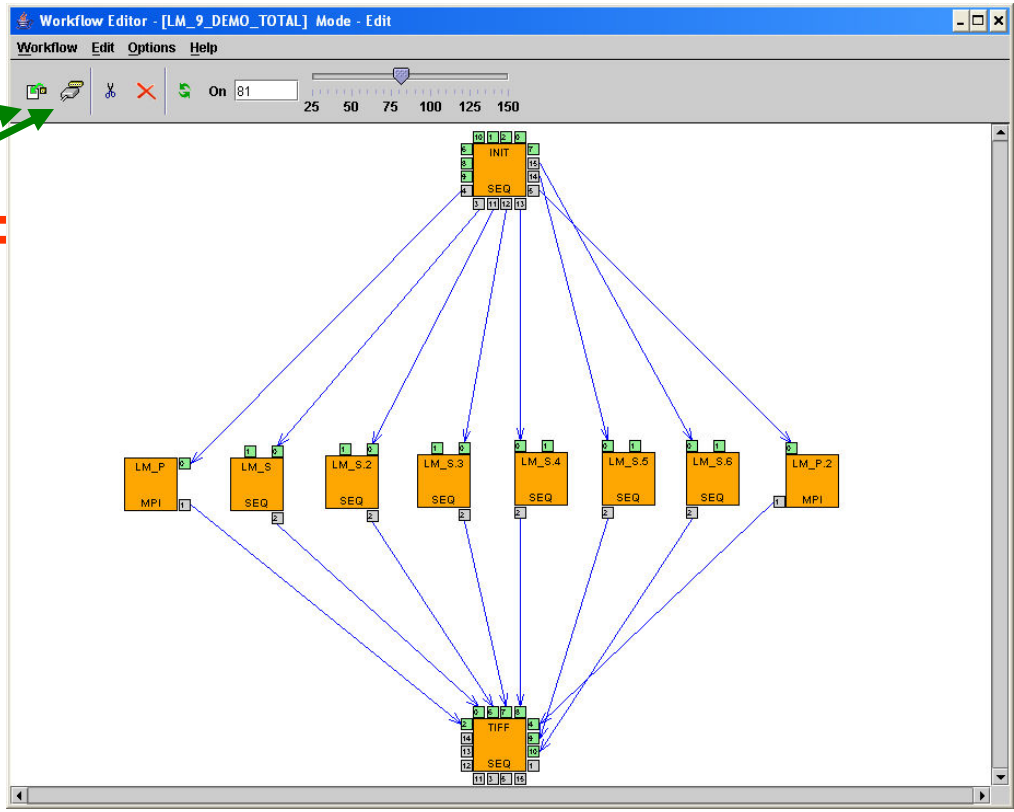
Workflow portlet



Graphical workflow editing

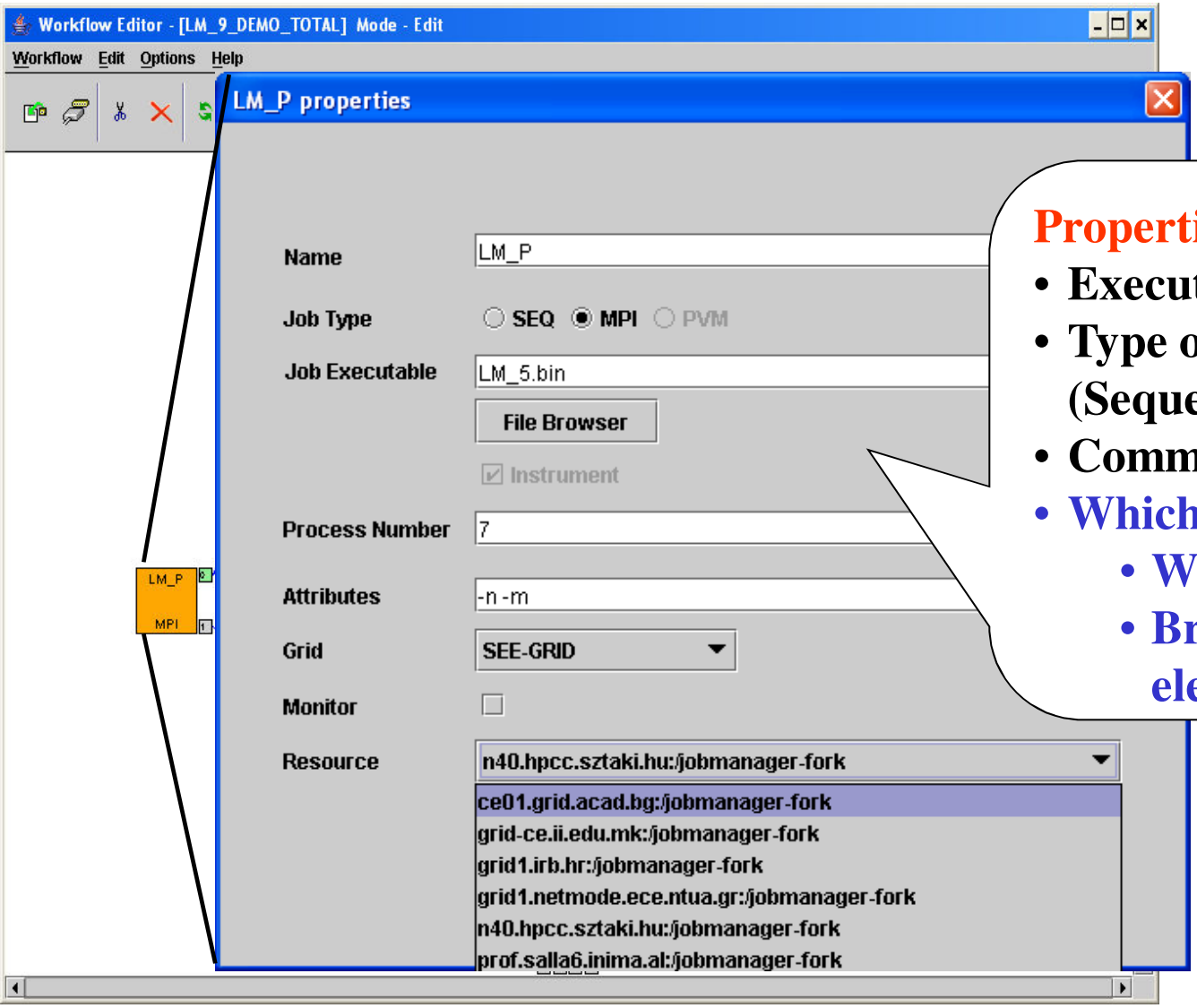
- To define a graph:
  1. **Drag & drop components:** jobs and ports
  2. **Define their properties**
  3. **Connect ports by channels** (no cycles, no loops)

  System generates JDL for each job automatically



Workflow Editor

Properties of a job

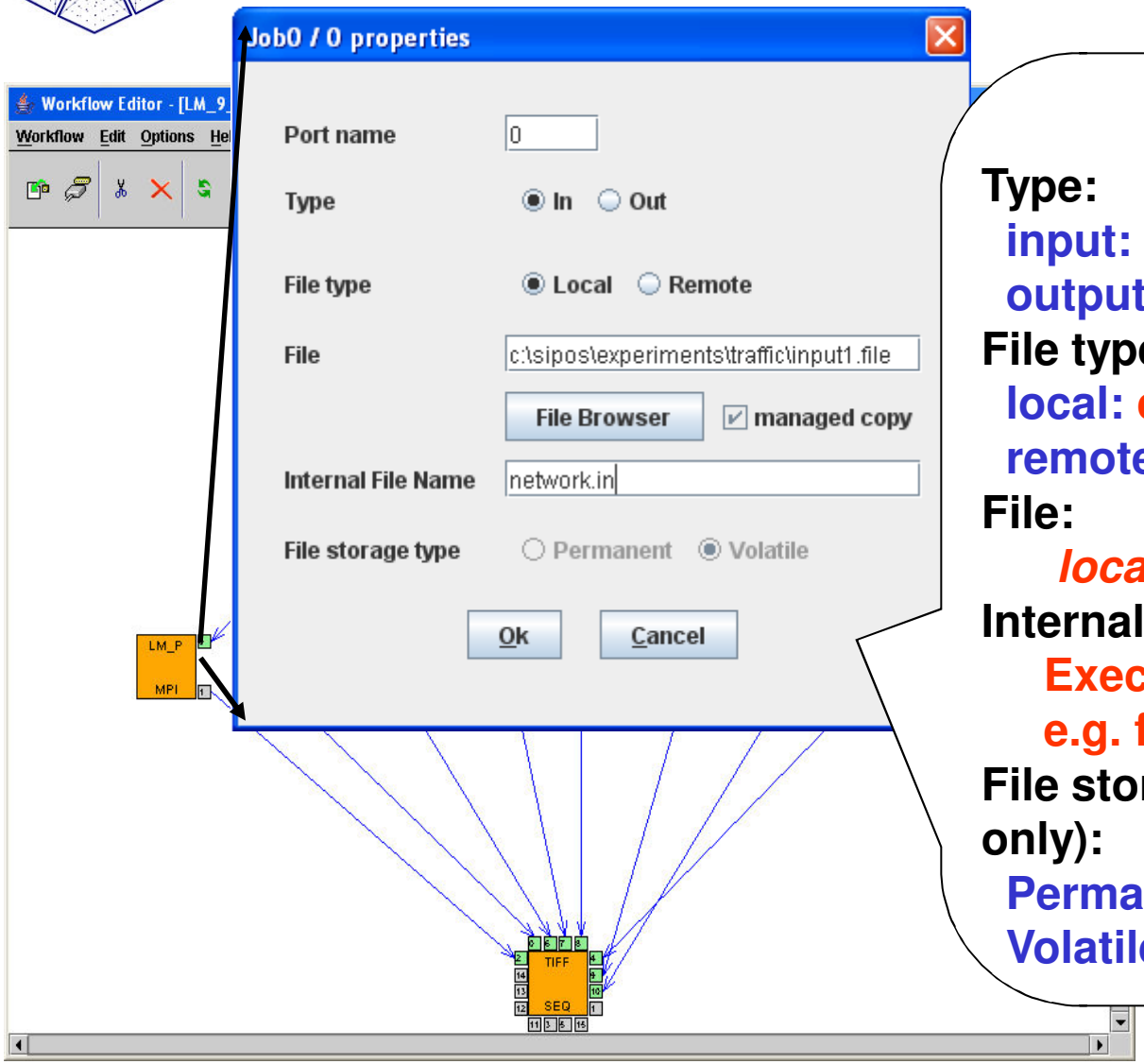


Properties of a job:

- **Executable file**
- **Type of executable (Sequential / Parallel)**
- **Command line parameters**
- **Which resource to use?**
  - **Which VO?**
  - **Broker or Computing element?**

Workflow Editor

Defining input-output files



File properties

Type:
- input: *the executable reads*
- output: *the executable generates*

File type:
- local: comes *from my desktop*
- remote: *comes from an SE*

File:
- *location of the file*

Internal file name:
- Executable uses this
- e.g. fopen("file.in", …)

File storage type (output files only):
- Permanent: *final result*
- Volatile: *temp. data channel*

How to refer to an I/O file?

Input file

Output file

Local file

- Client side location:
  c:\experiments\11-04.dat

- Client side location:
  result.dat

- LFC logical file name
  (LFC file catalog is required – EGEE VOs)
  lfn:/grid/gilda/sipos/11-04.dat

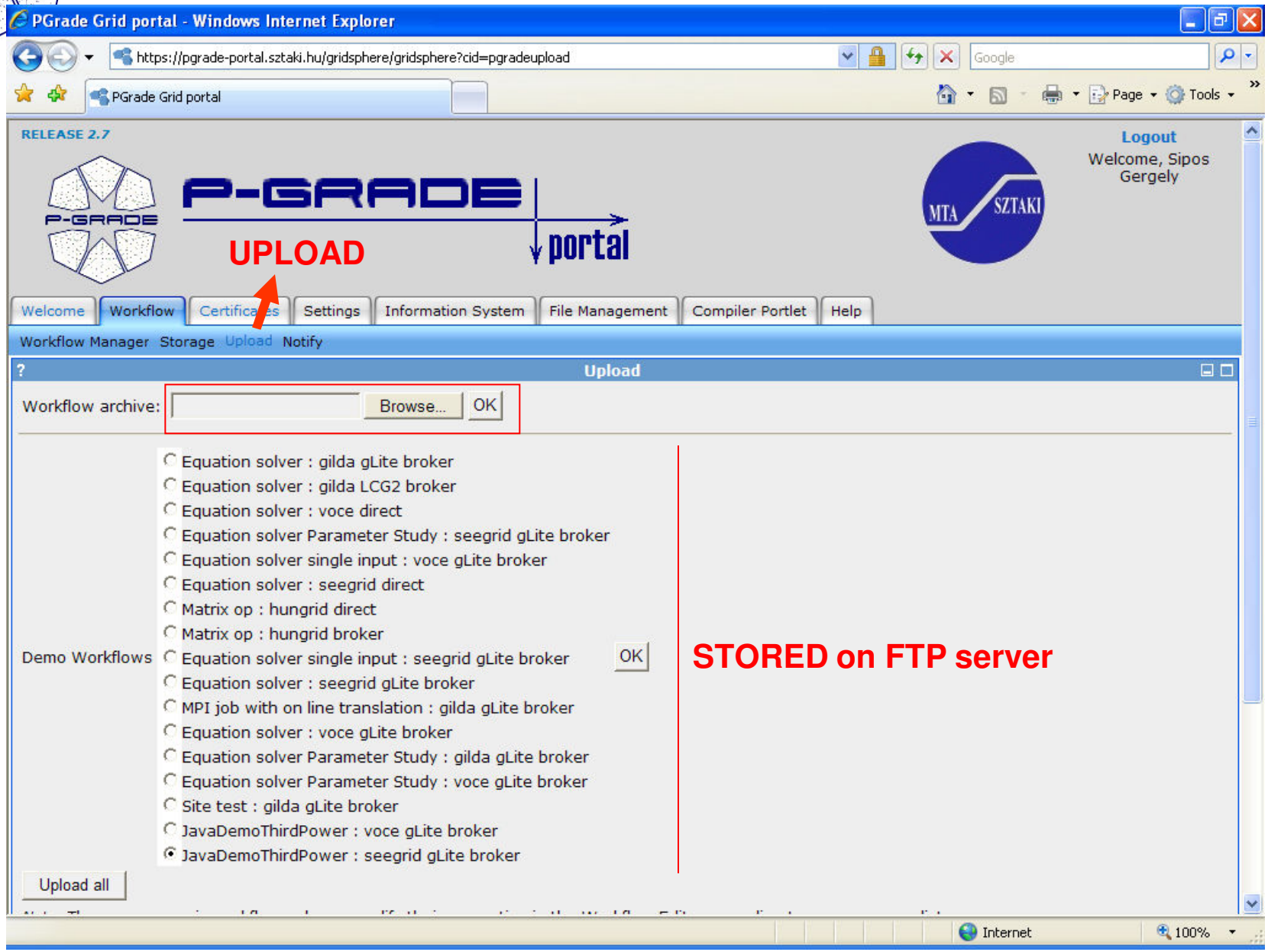
- GridFTP address (in Globus Grids):
  gsiftp://somengshost.ac.uk/mydir/11-04.dat

- LFC logical file name
  (LFC file catalog is required – EGEE VOs)
  lfn:/grid/gilda/sipos/11-04_-_result.dat

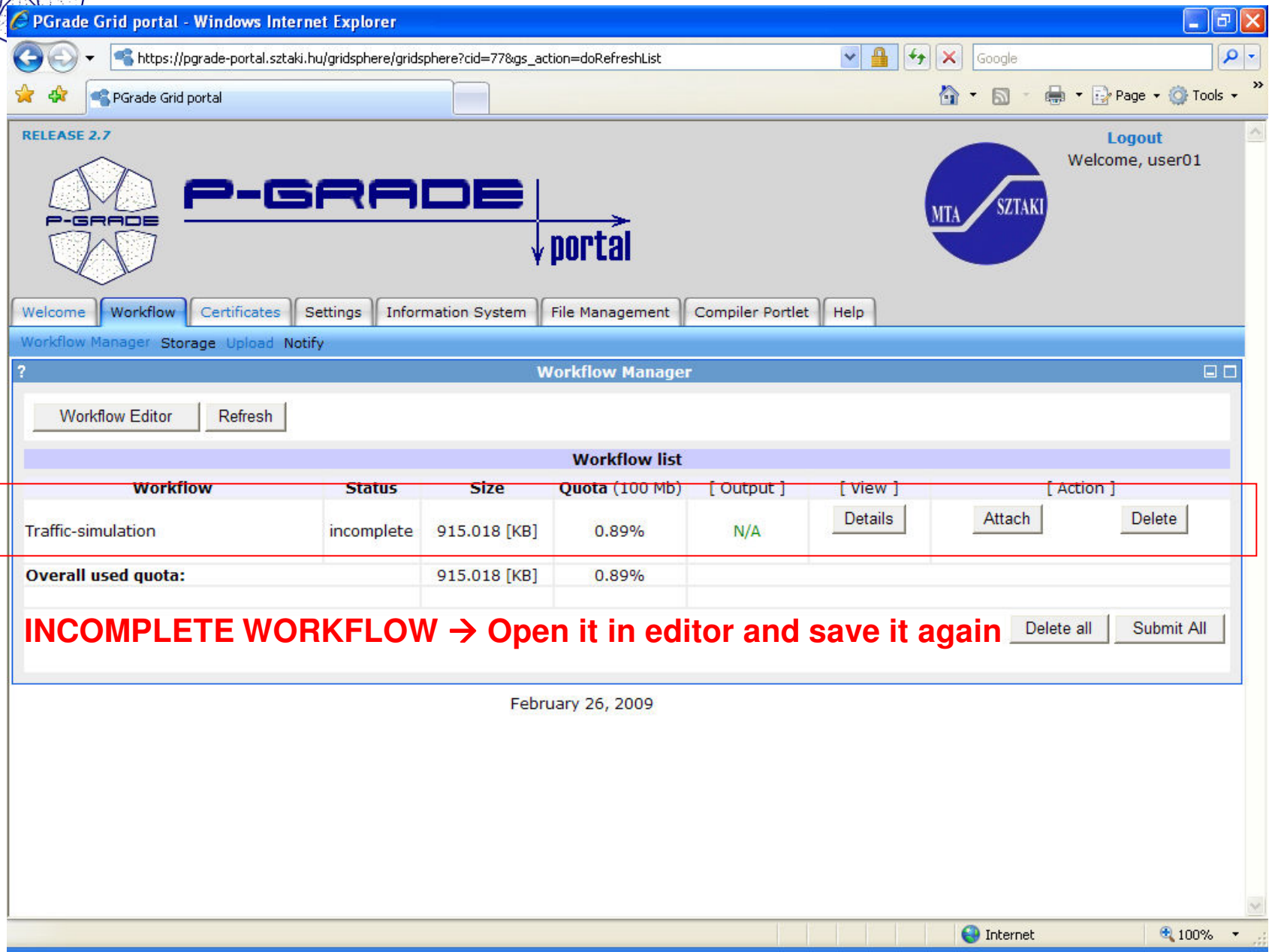
- GridFTP address (in Globus Grids):
  gsiftp://somengshost.ac.uk/mydir/result.dat

Remote file

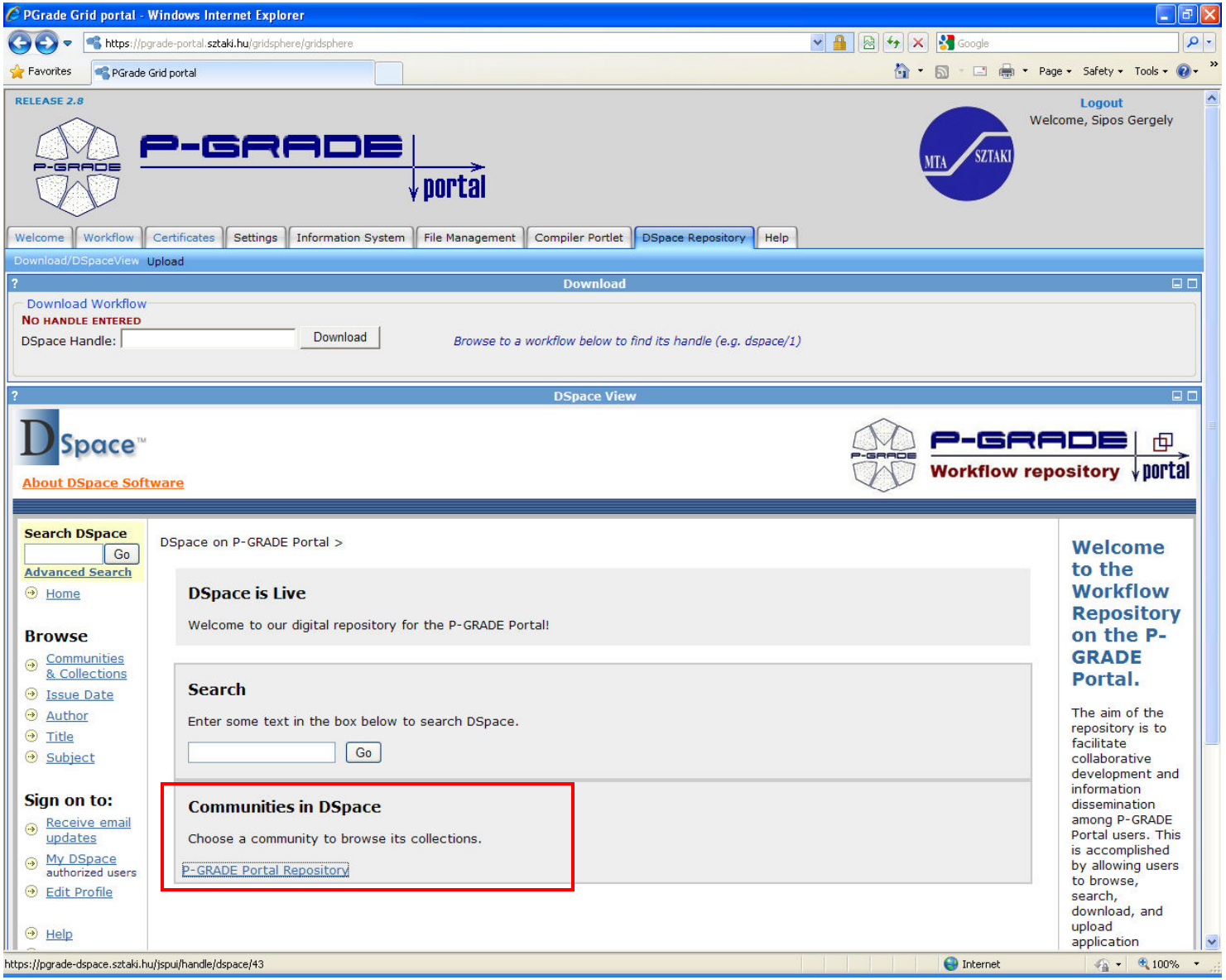
Upload a workflow from client side or from FTP server



Importing an application

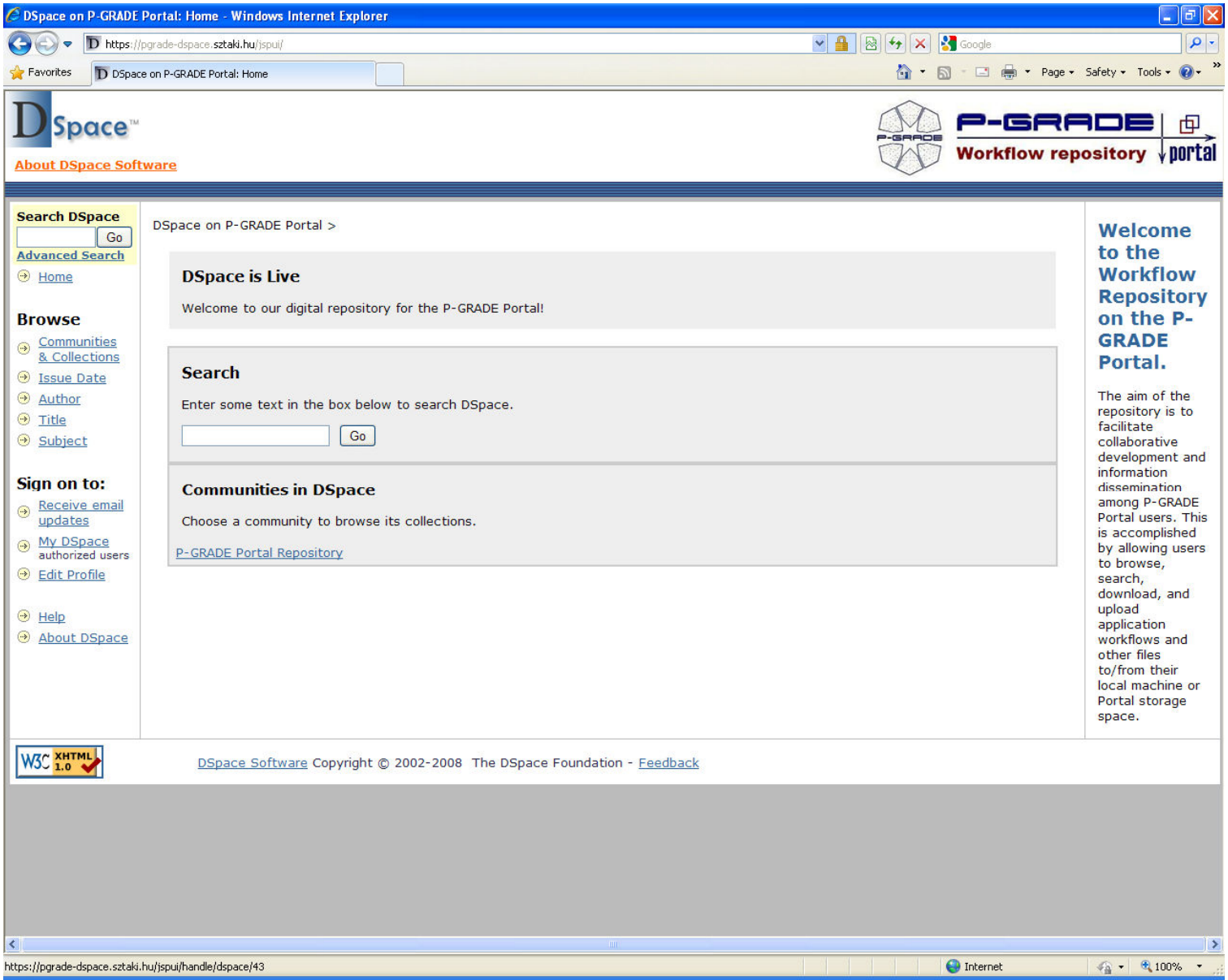


Import a workflow from DSpace repository

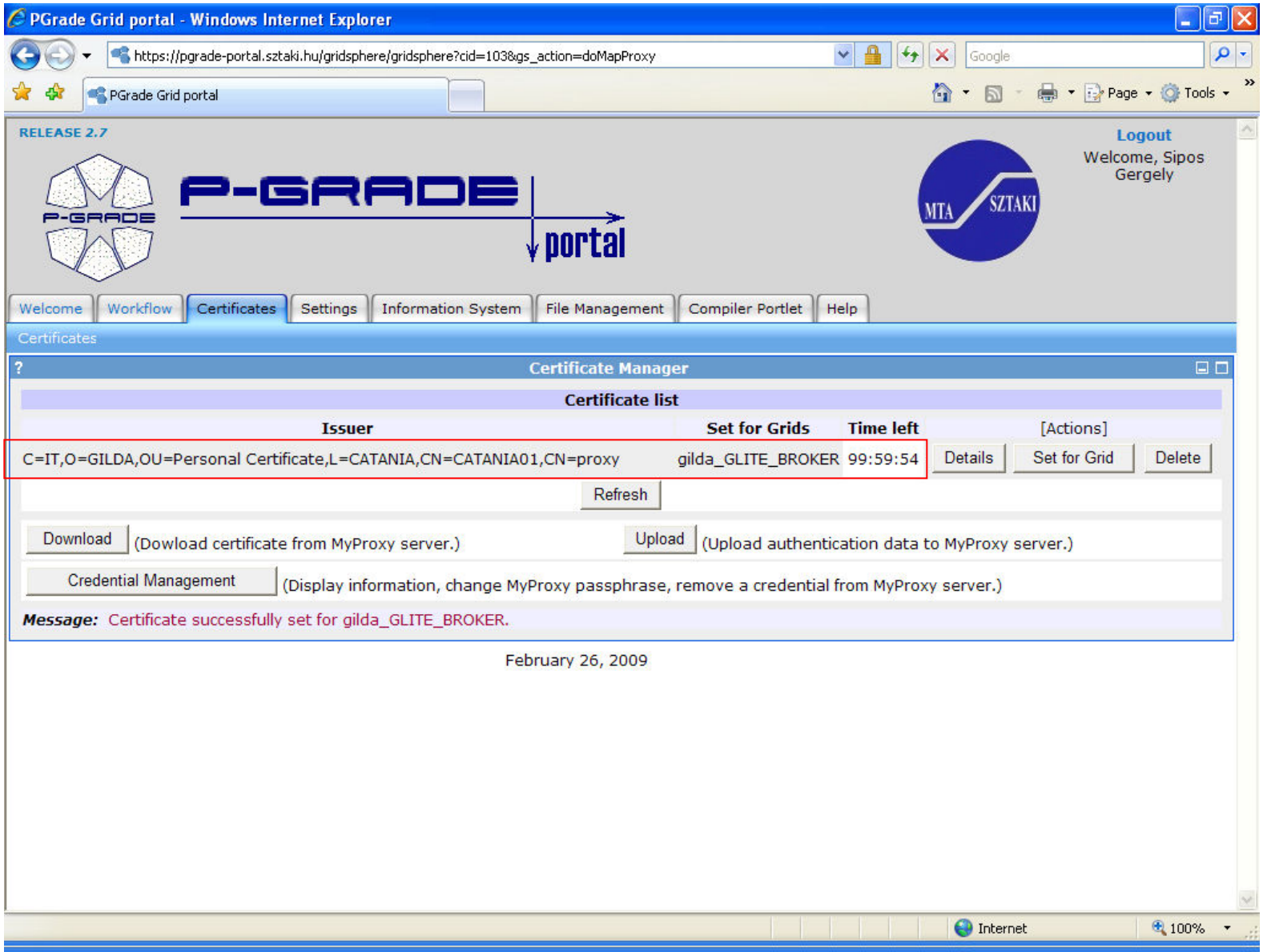


External access to DSpace

http://pgrade-dspace.sztaki.hu

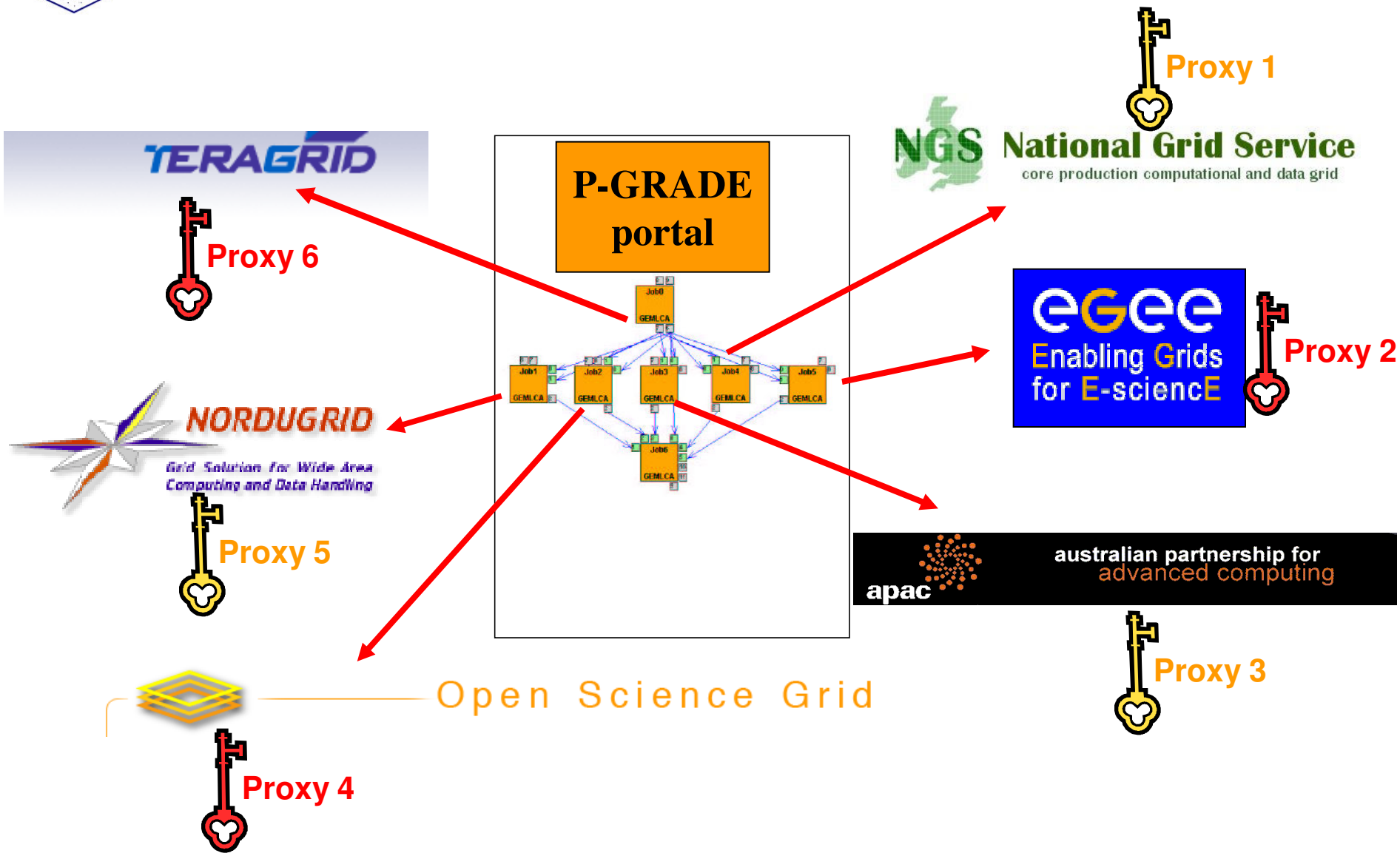


Certificate and proxy management Portlet

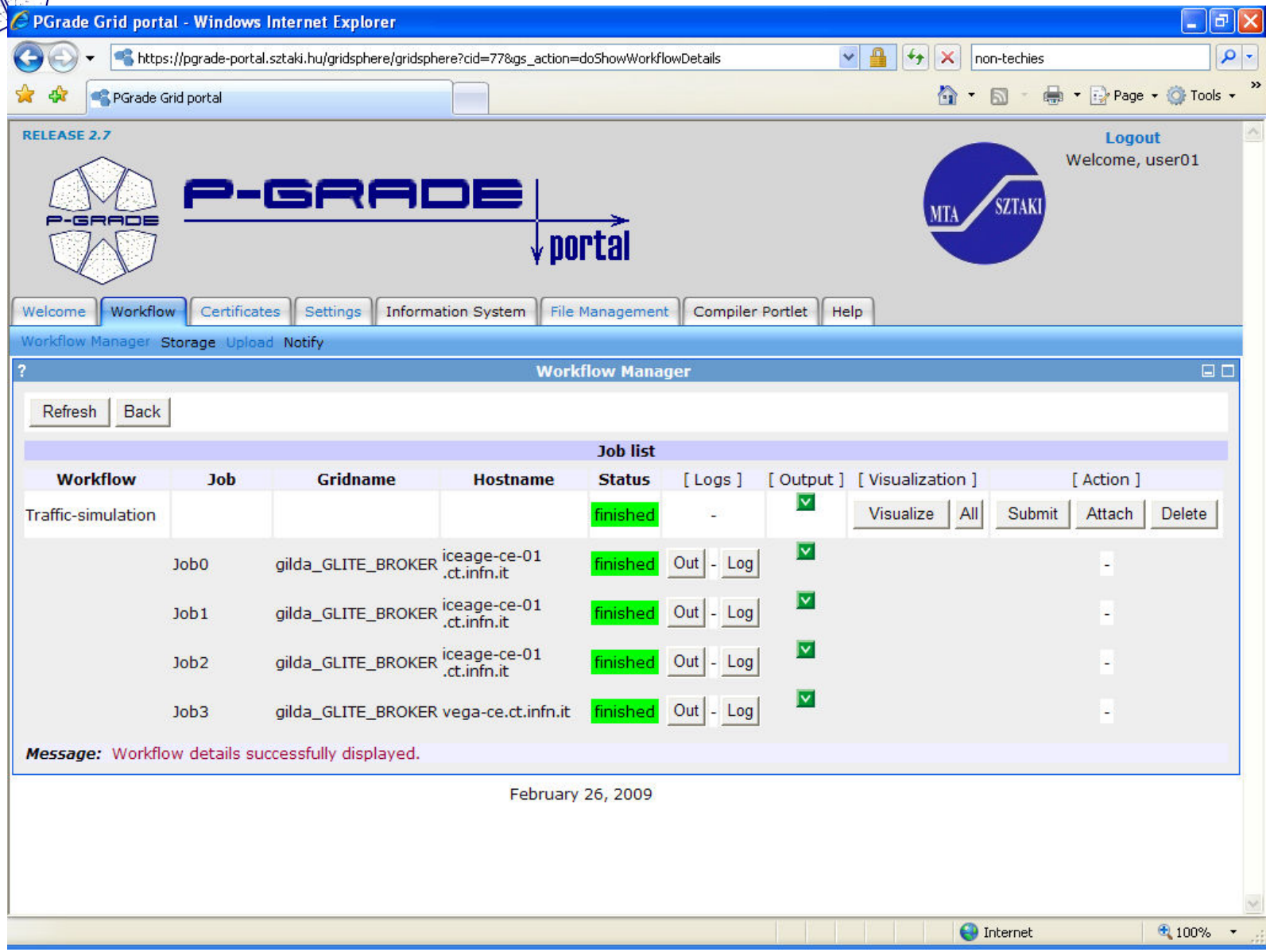


OGF GIN interoperability portal by P-GRADE

Acccessing Globus, gLite and ARC based grids/VOs simultaneously



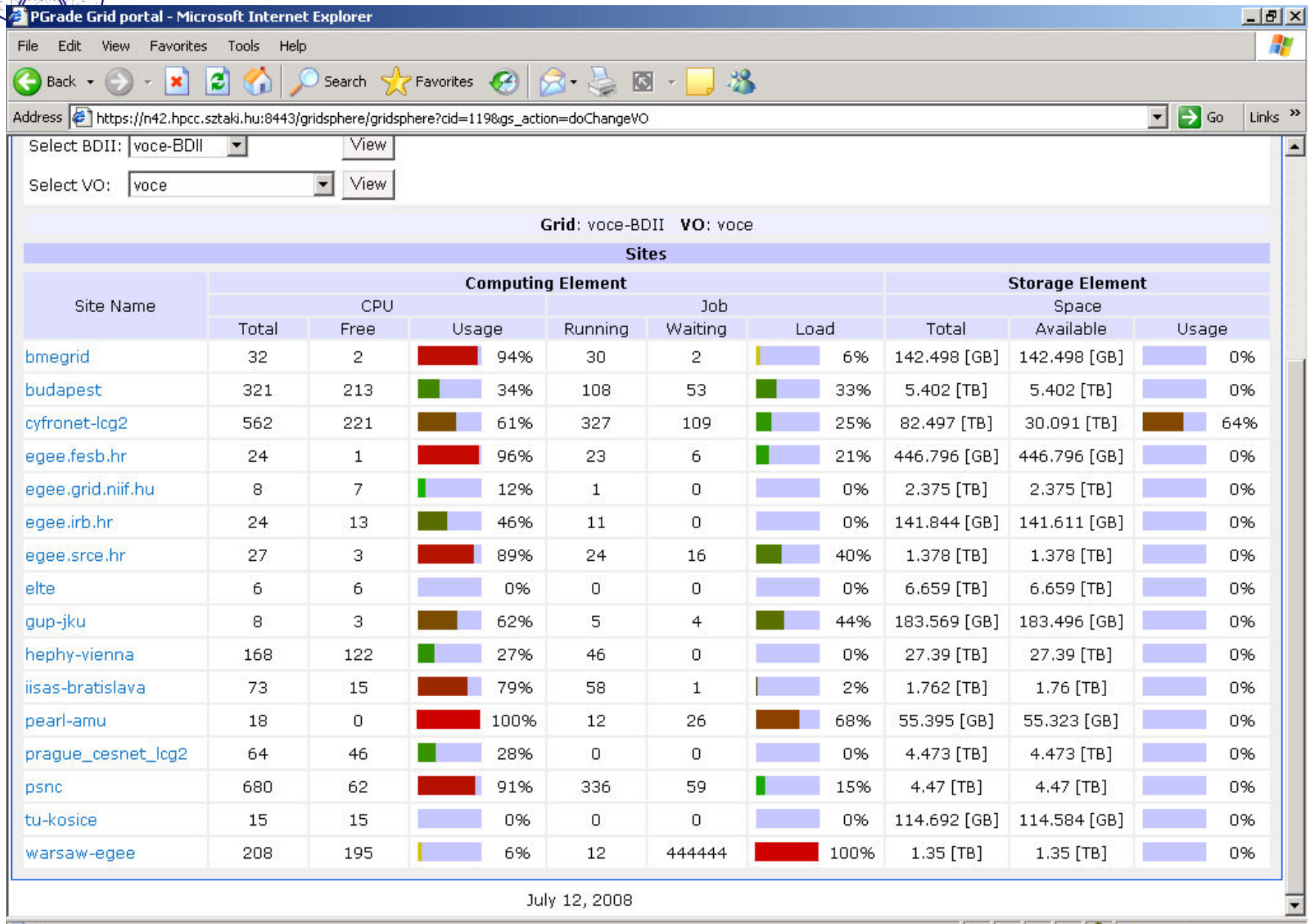
Application execution



Fault-tolerant execution

- Utilizing
  - Condor DAGMan's rescue mechanism
  - EGEE job resubmission mechanism of WMS
- If the EGEE broker leaves a job stuck in a CEs' queue, the portal automatically
  - kills the job on this site and
  - resubmits the job to the broker by prohibiting this site.
- As a result
  - the portal guarantees the correct submission of a job as long as there exists at least one matching resource
  - job submission is reliable even in an unreliable grid

Information system visualization



PGrade Grid portal - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites

Address https://n42.hpcc.sztaki.hu:8443/gridsphere/gridsphere?cid=119&gs_action=doChangeVO Go Links

Select BDII: voce-BDII View

Select VO: voce View

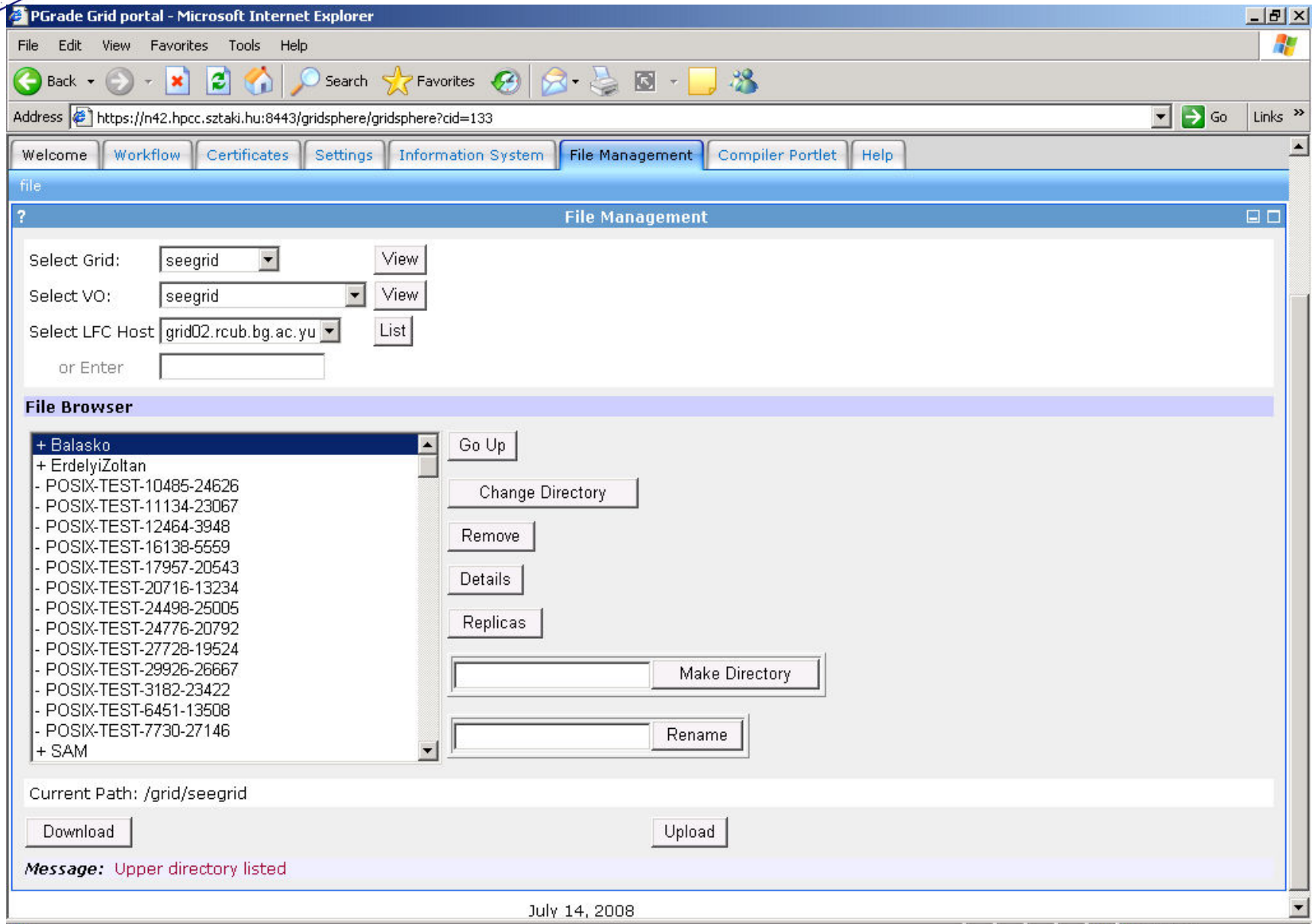
Grid: voce-BDII **VO**: voce

Sites

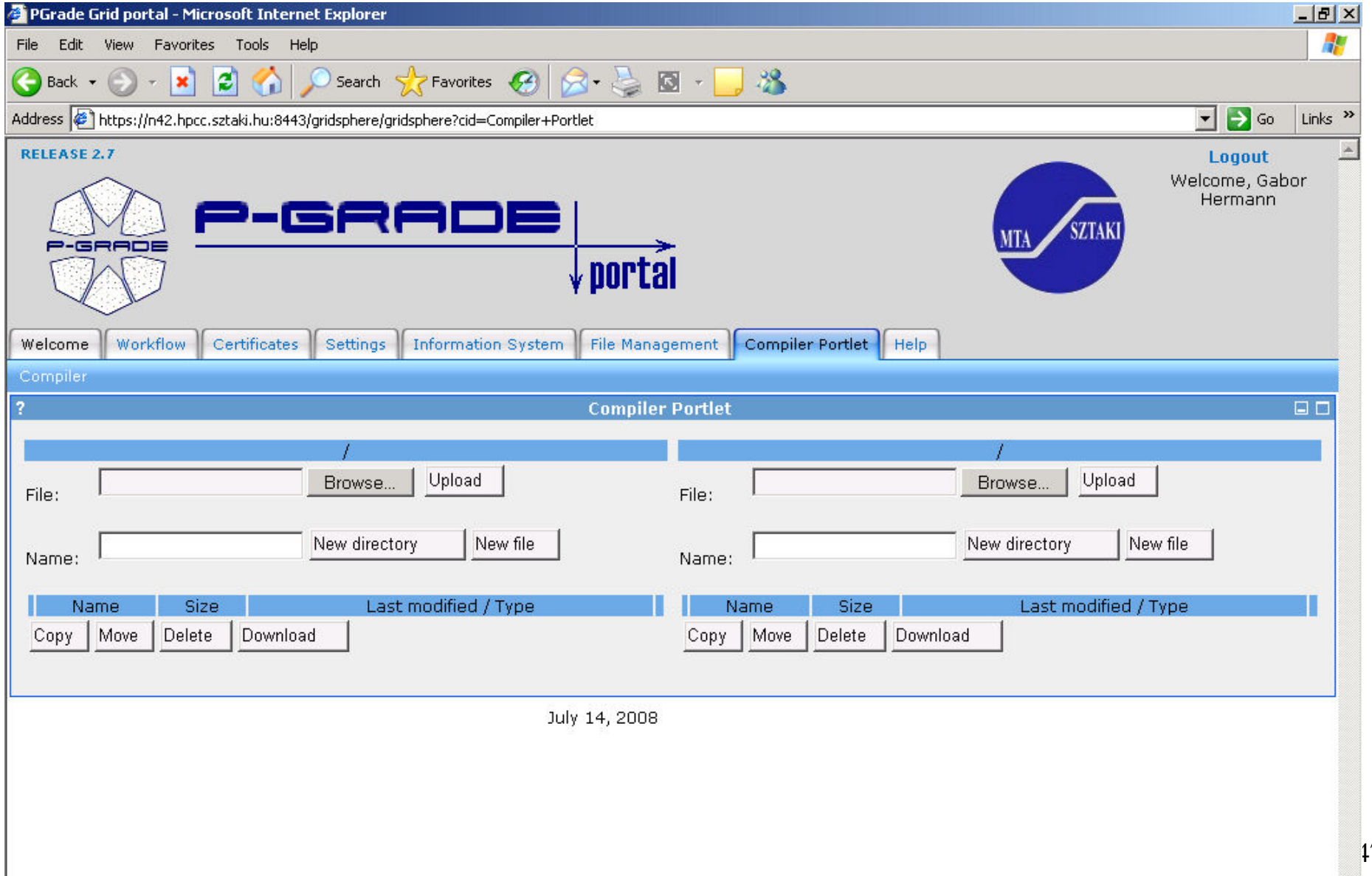
Site Name	Computing Element						Storage Element		
	CPU			Job			Space		
	Total	Free	Usage	Running	Waiting	Load	Total	Available	Usage
bmegrid	32	2	94%	30	2	6%	142.498 [GB]	142.498 [GB]	0%
budapest	321	213	34%	108	53	33%	5.402 [TB]	5.402 [TB]	0%
cyfronet-lcg2	562	221	61%	327	109	25%	82.497 [TB]	30.091 [TB]	64%
egee.fesb.hr	24	1	96%	23	6	21%	446.796 [GB]	446.796 [GB]	0%
egee.grid.niif.hu	8	7	12%	1	0	0%	2.375 [TB]	2.375 [TB]	0%
egee.irb.hr	24	13	46%	11	0	0%	141.844 [GB]	141.611 [GB]	0%
egee.srce.hr	27	3	89%	24	16	40%	1.378 [TB]	1.378 [TB]	0%
elte	6	6	0%	0	0	0%	6.659 [TB]	6.659 [TB]	0%
gup-jku	8	3	62%	5	4	44%	183.569 [GB]	183.496 [GB]	0%
hephy-vienna	168	122	27%	46	0	0%	27.39 [TB]	27.39 [TB]	0%
iisas-bratislava	73	15	79%	58	1	2%	1.762 [TB]	1.76 [TB]	0%
pearl-amu	18	0	100%	12	26	68%	55.395 [GB]	55.323 [GB]	0%
prague_cesnet_lcg2	64	46	28%	0	0	0%	4.473 [TB]	4.473 [TB]	0%
psnc	680	62	91%	336	59	15%	4.47 [TB]	4.47 [TB]	0%
tu-kosice	15	15	0%	0	0	0%	114.692 [GB]	114.584 [GB]	0%
warsaw-egee	208	195	6%	12	444444	100%	1.35 [TB]	1.35 [TB]	0%

July 12, 2008

LFC-SE file browser portlet



Compilation support

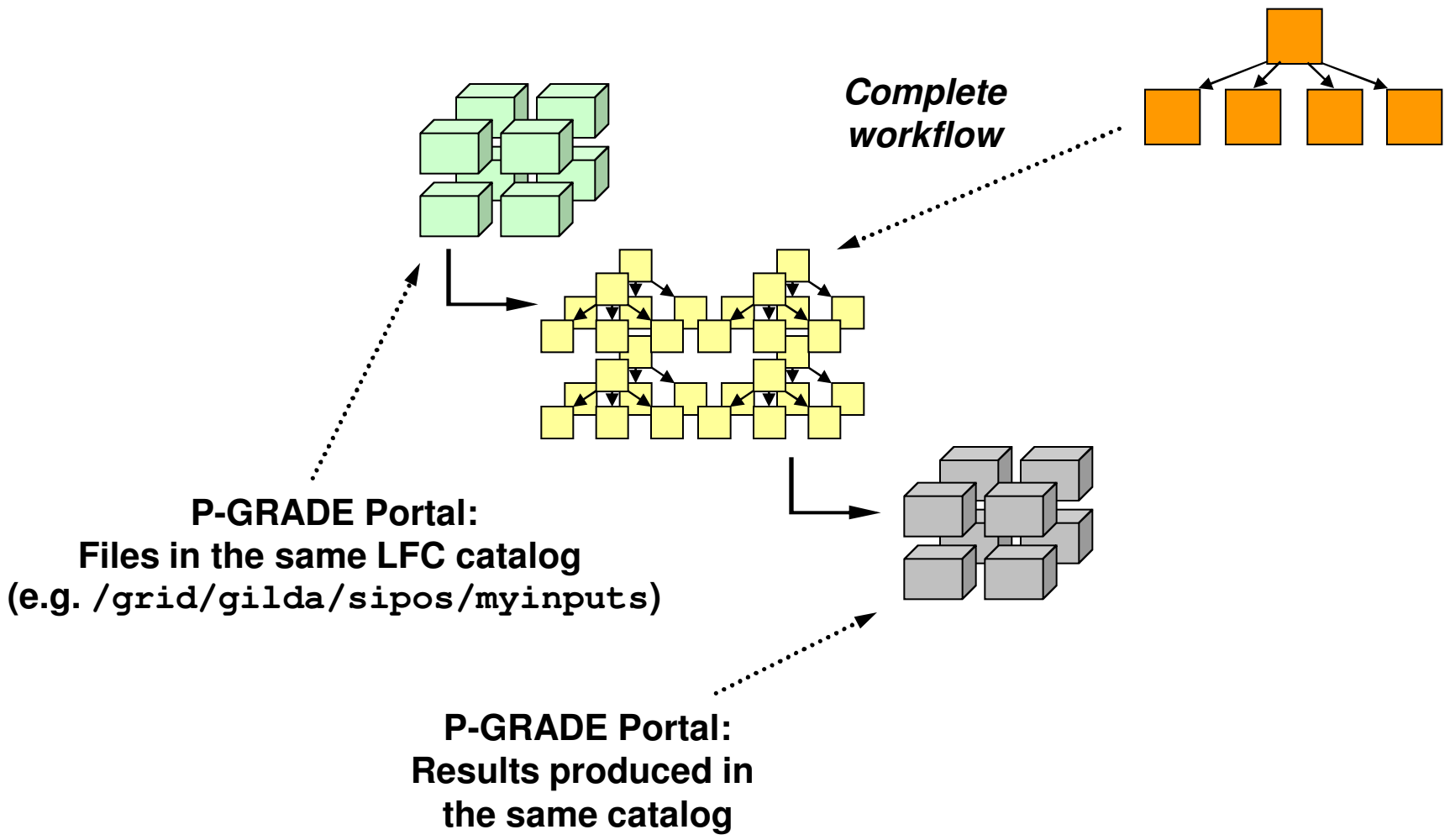


WORKFLOW HANDS-ON

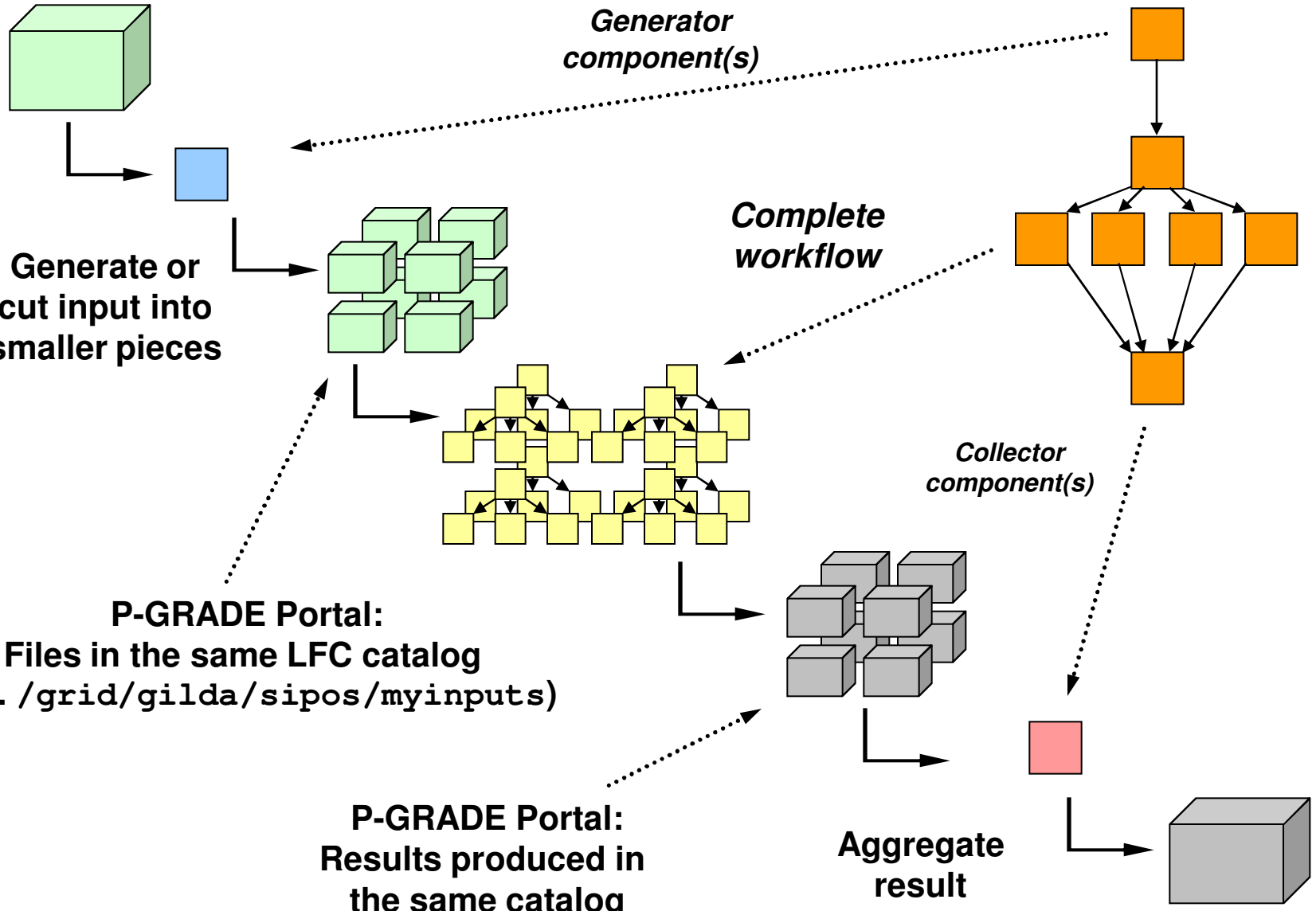
From workflows to parameter studies

Advanced execution patterns

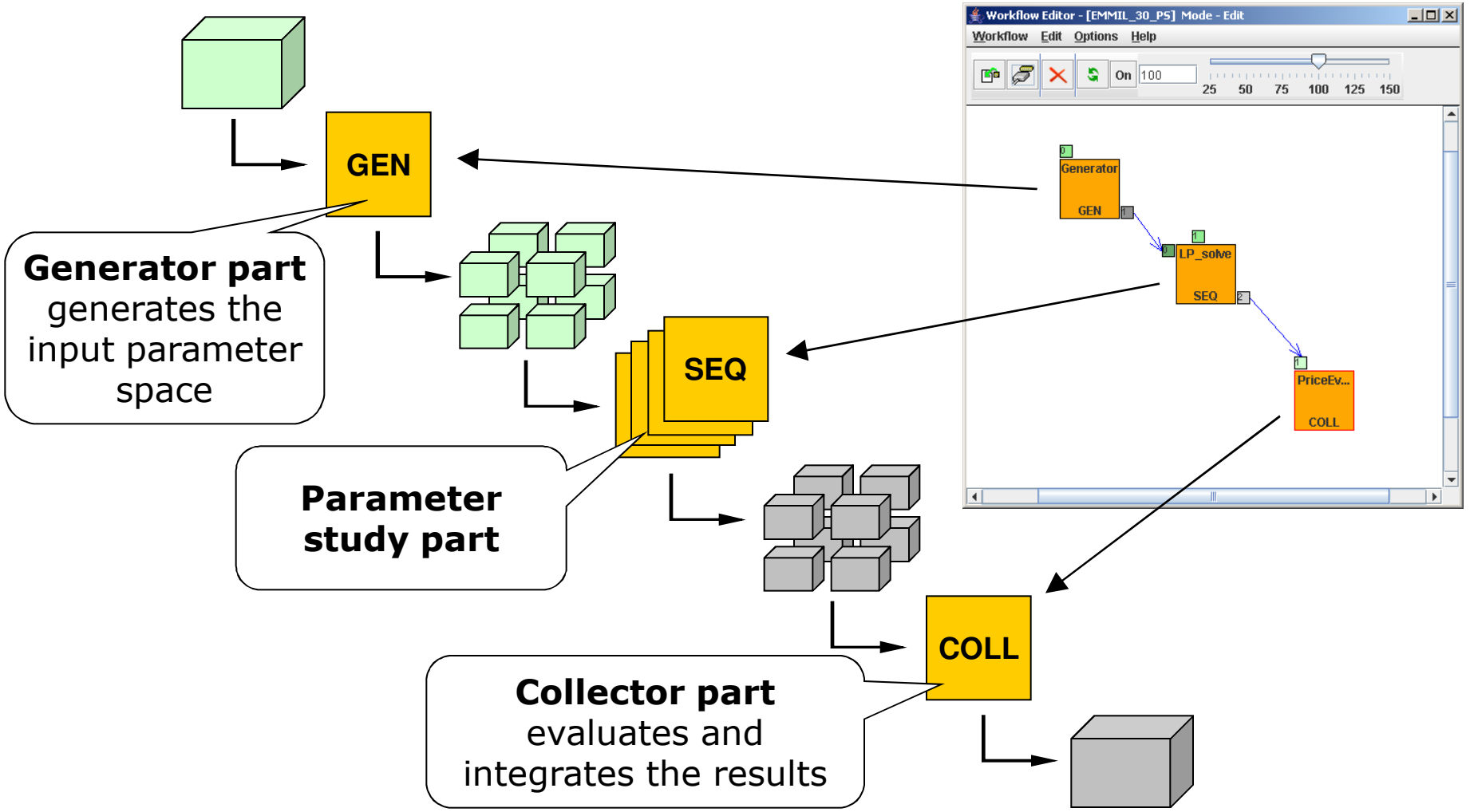
Scaling up a workflow to a parameter study



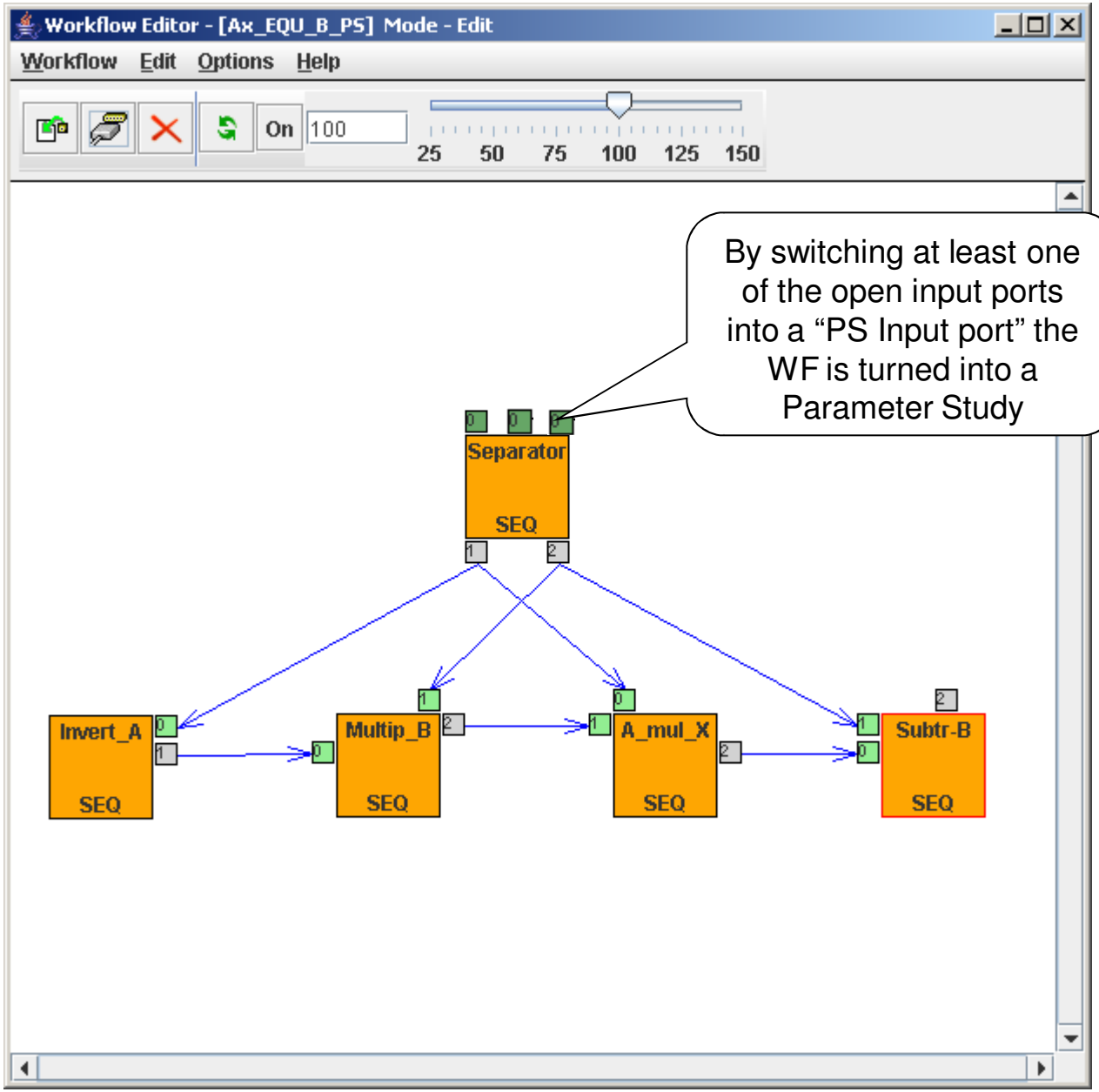
Advanced parameter studies



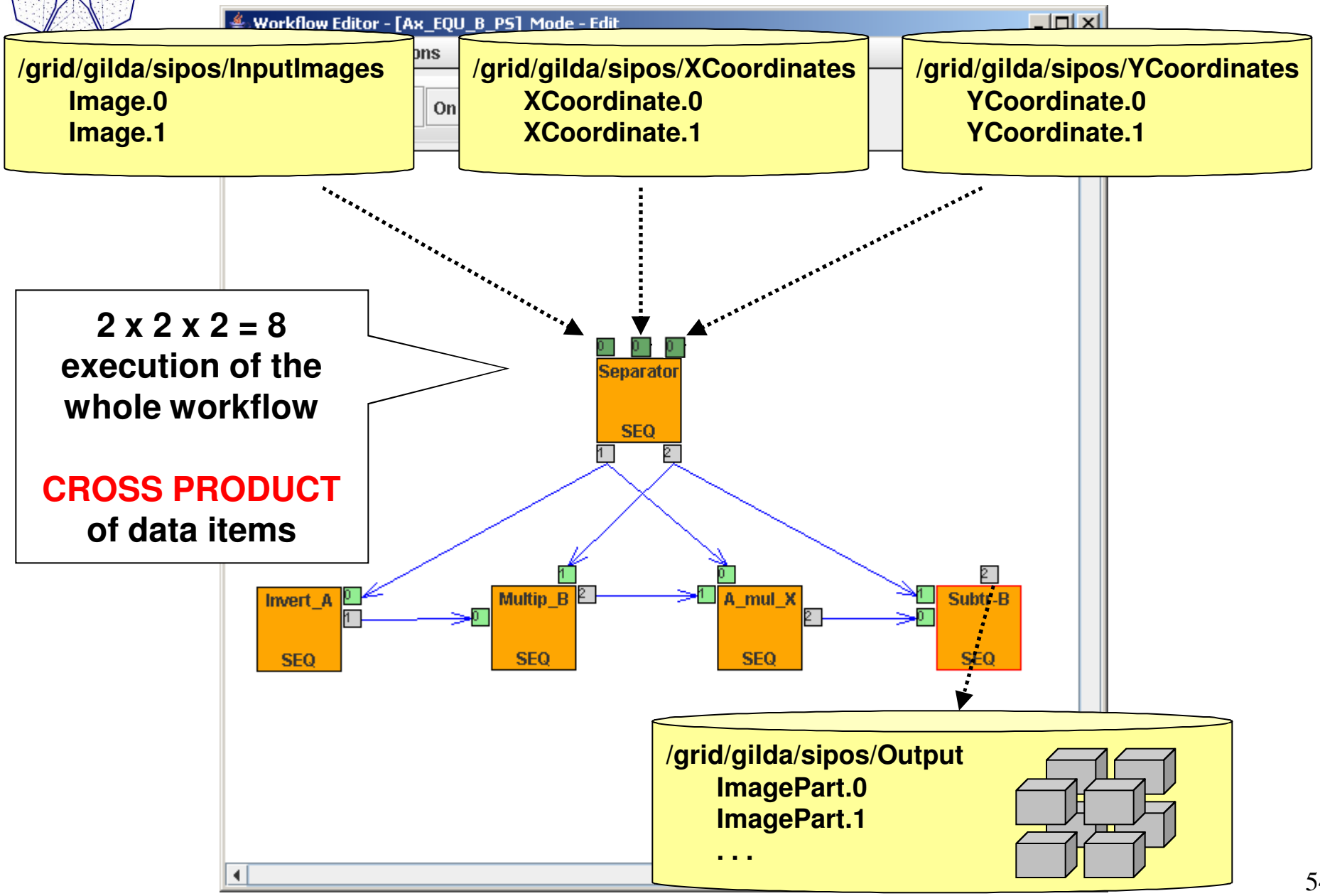
Concept of parameter study workflows



Turning a WF into a parameter study

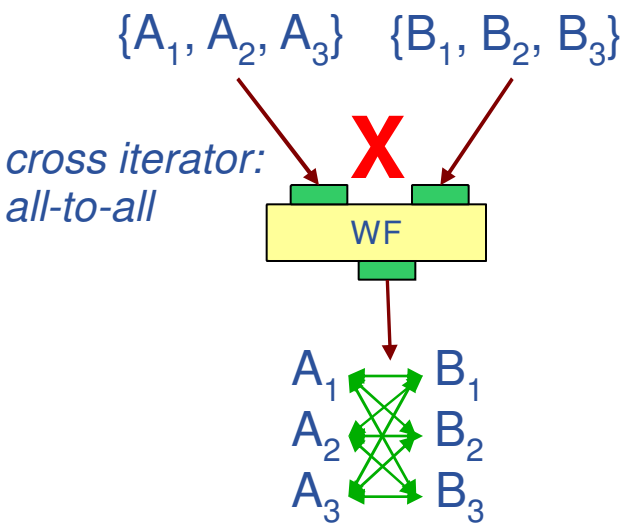


Input-output files are stored in SEs



Typical data-flow compositions

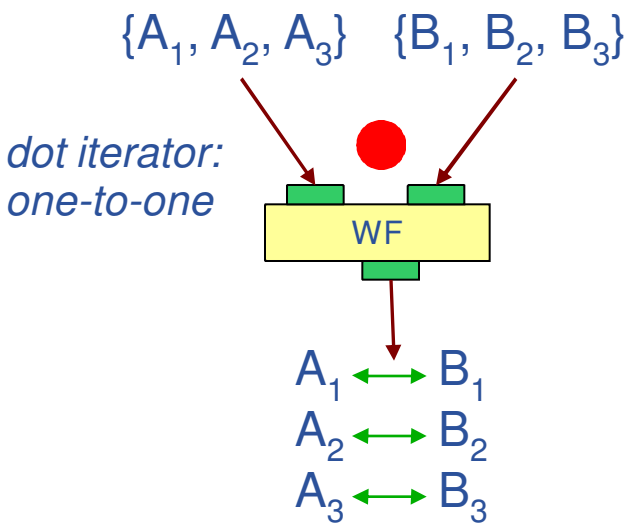
CROSS ITERATOR



A x B

P-GRADE Portal supports this

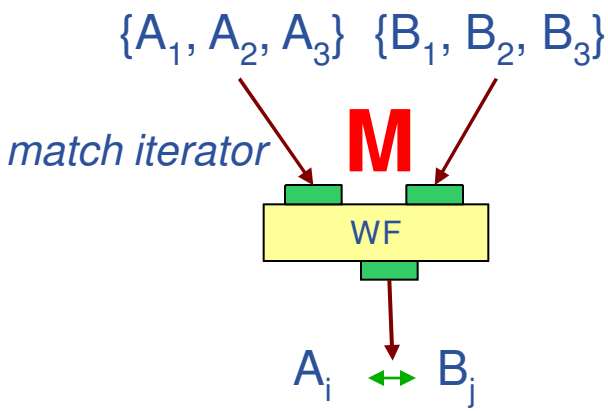
DOT ITERATOR



A • B

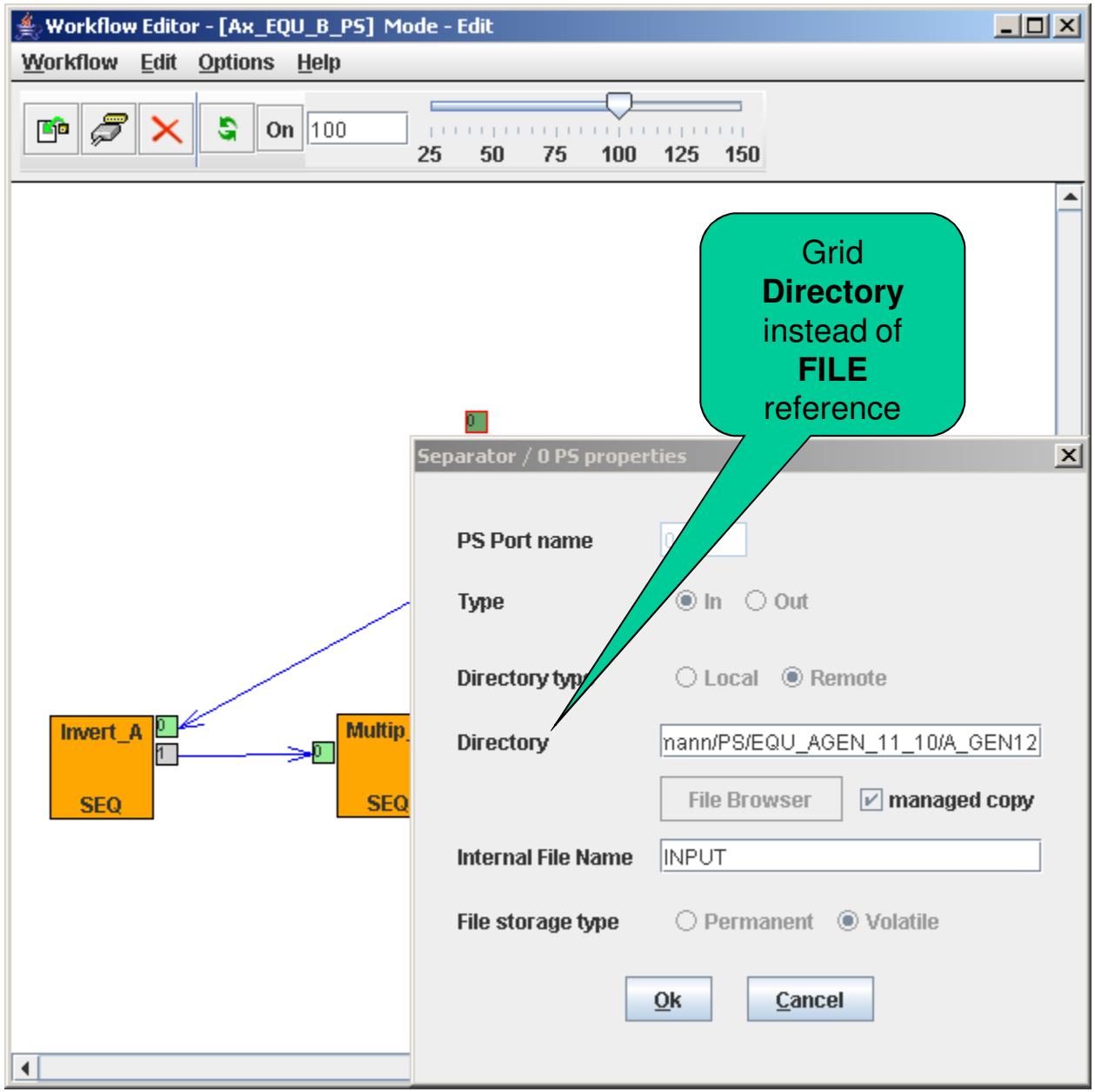
Find these in e.g. TAVERNA, MOTEUR

MATCH ITERATOR

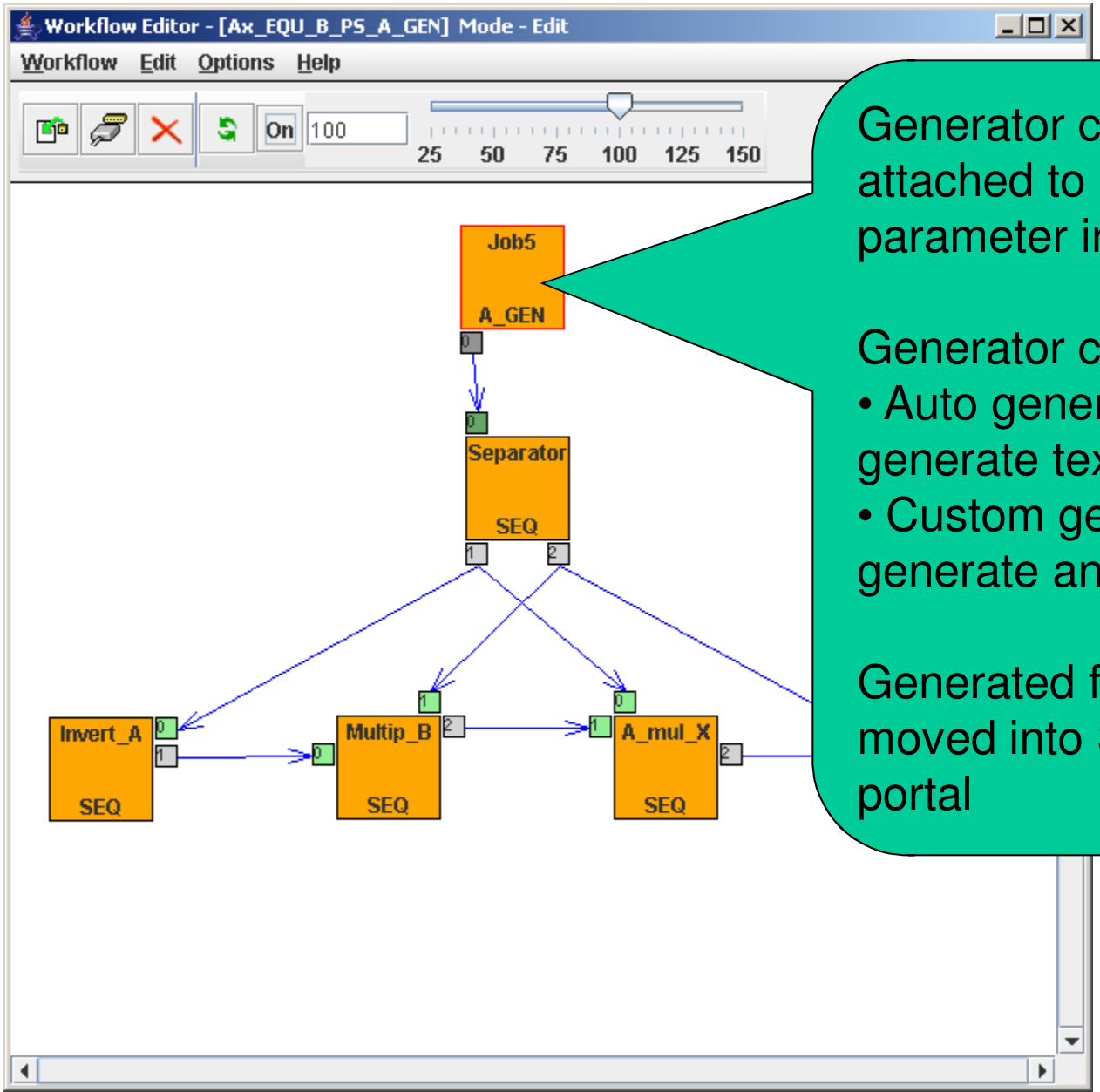


A M B

PS Input Port



Parameter generator

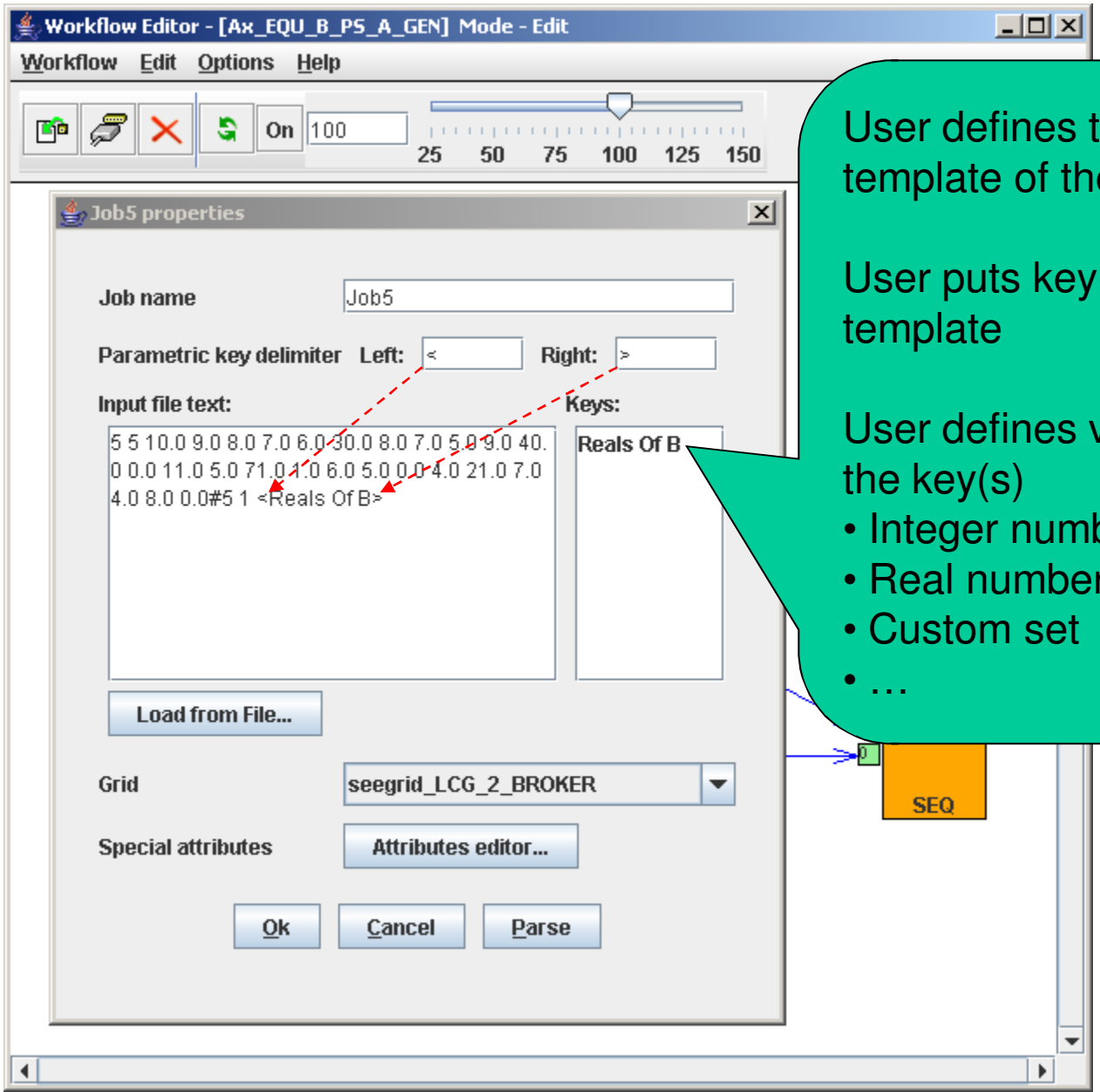


Generator can be attached to any parameter input port

Generator can be
- Auto generator: to generate text files
- Custom generator: to generate any content

Generated files are moved into SE by the portal

Definition Window of Auto Generator Job

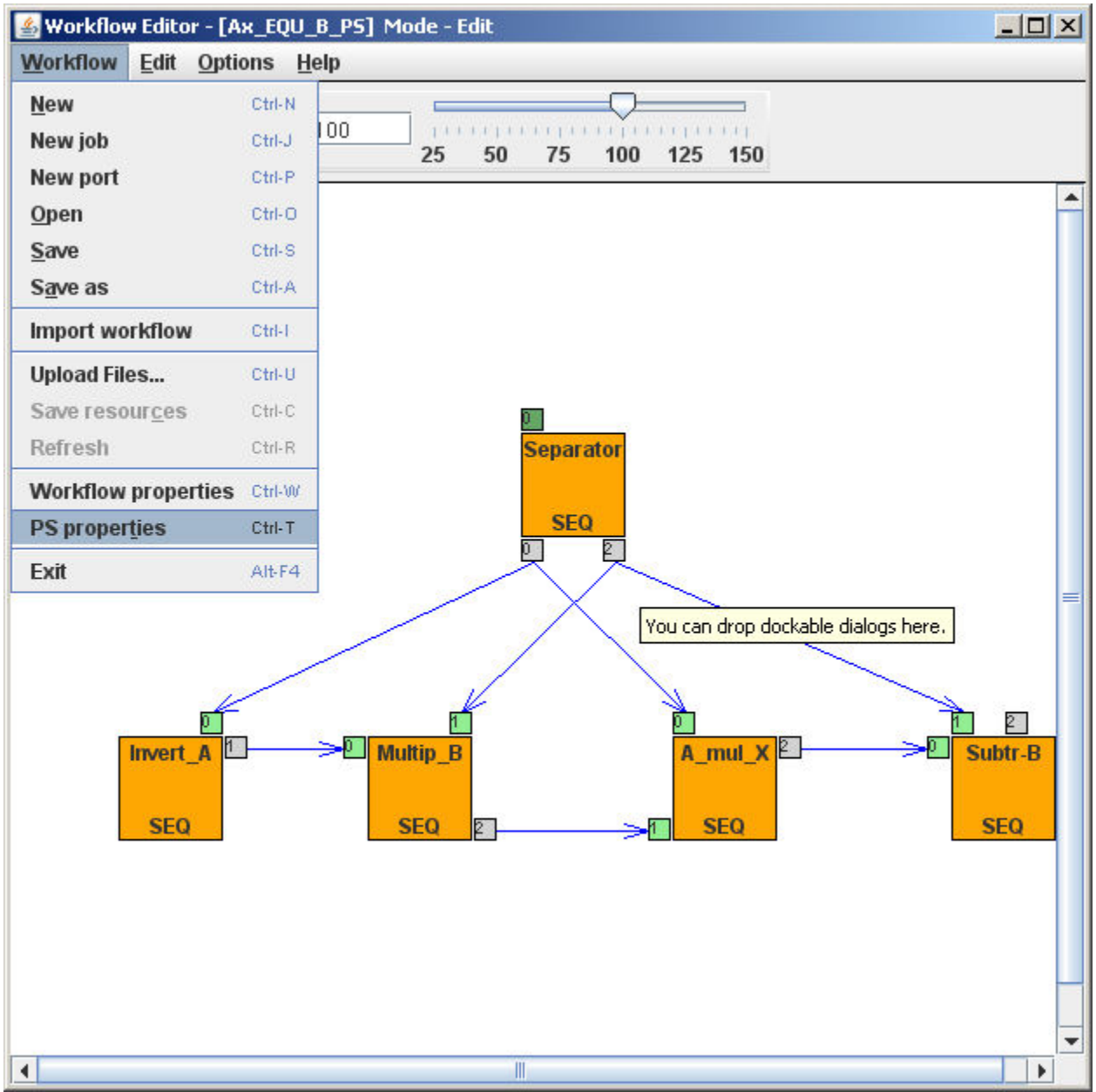


User defines the template of the text file

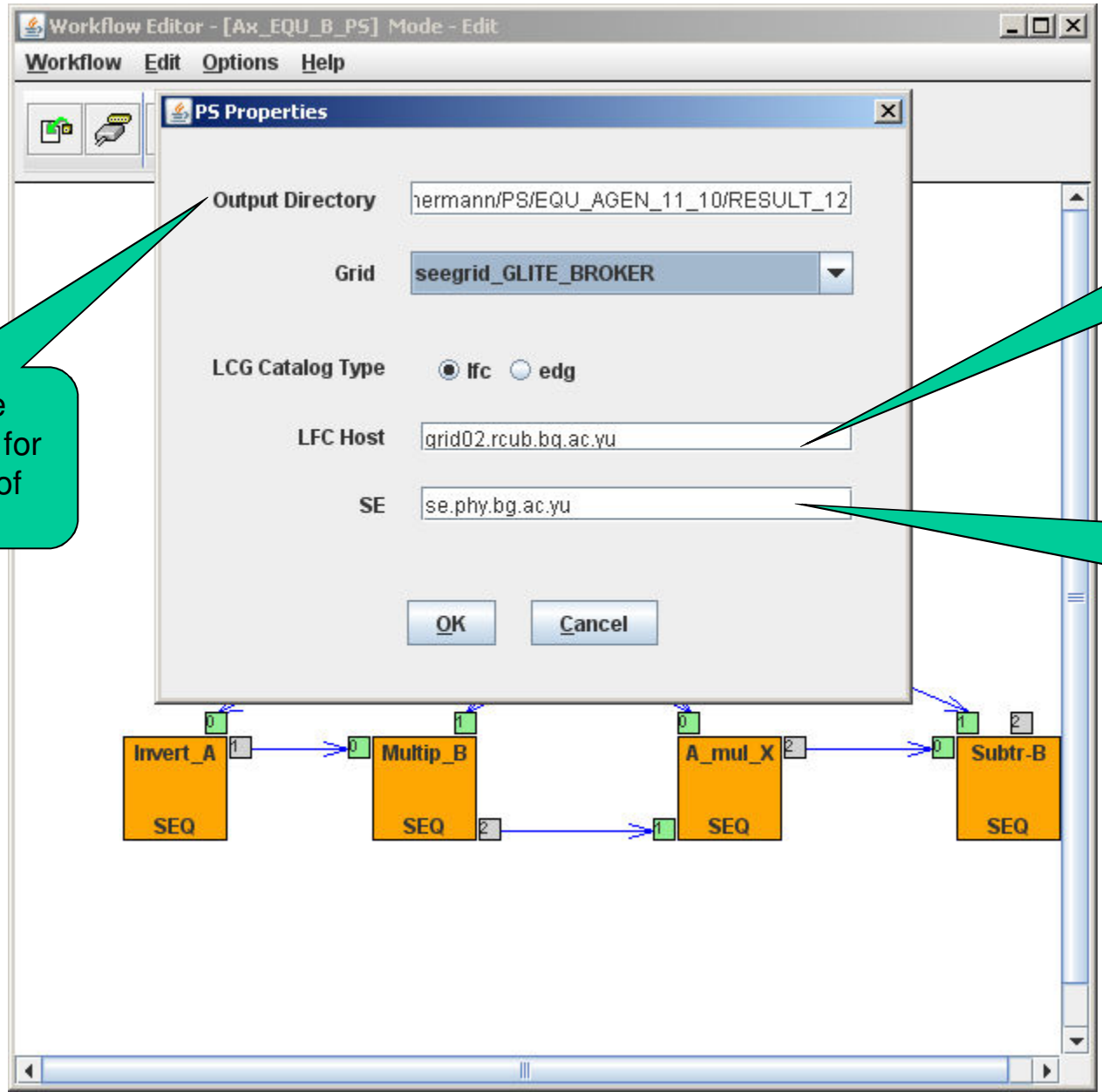
User puts key(s) into the template

User defines values for the key(s)
- Integer number
- Real number
- Custom set
- …

Placement of result



Placement of result

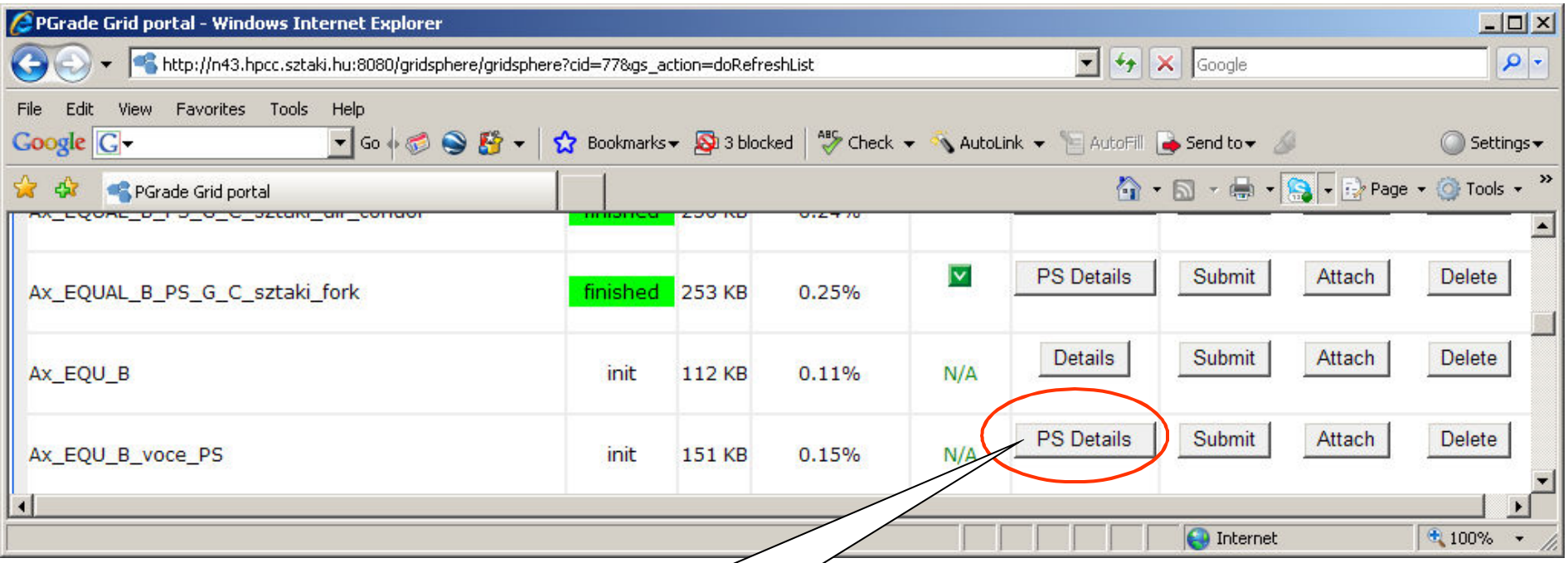


Will contain one compressed file for each execution of the workflow.

Use the default value!

Choose a „reliable" Storage Element

Executing PS workflows



PS Details for parameter sweep workflows applications

Detailed view of a PS workflow

Generator job(s)

Overall statistics of workflow instances

Workflow instances

Collector job(s)

PARAMETER STUDY HANDS-ON

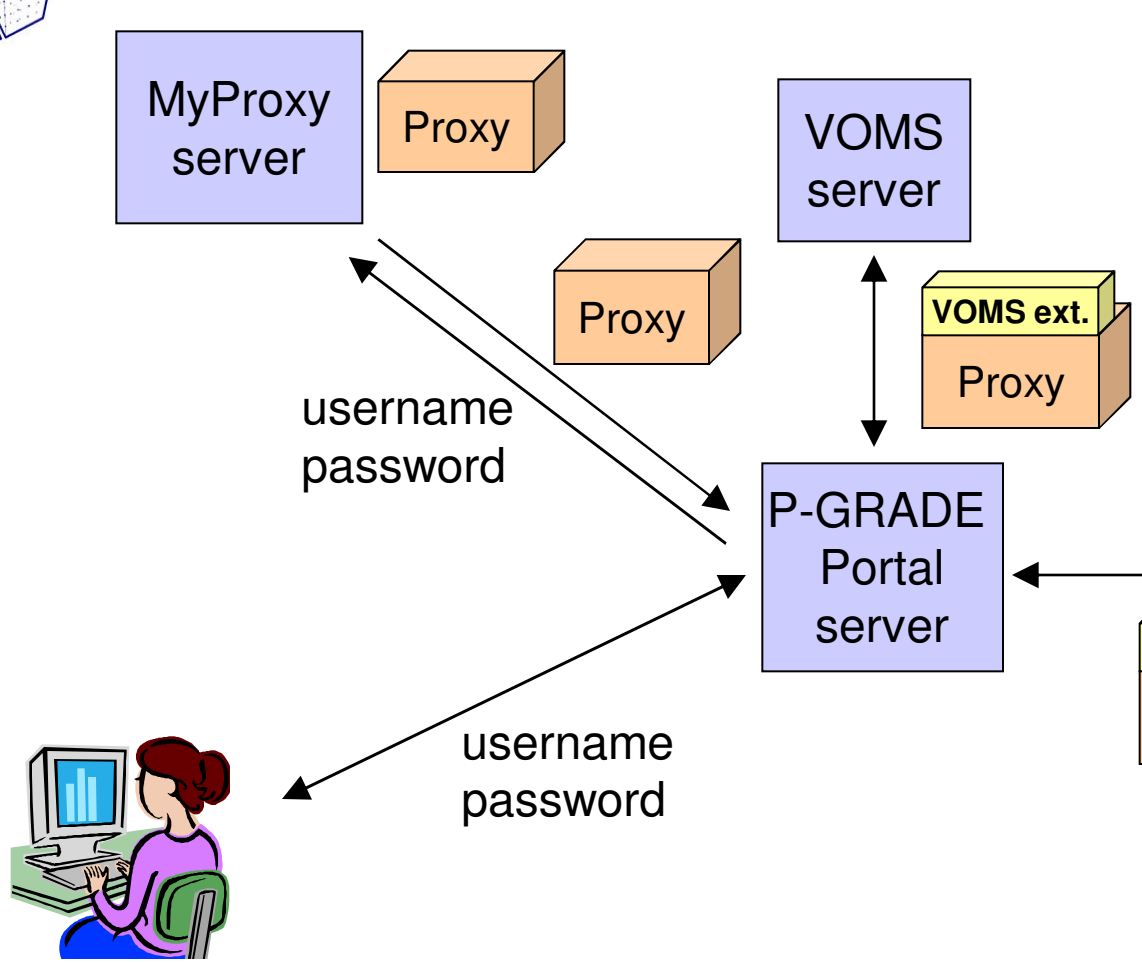
Learn once, use everywhere
Develop once, execute anywhere

Thank you!

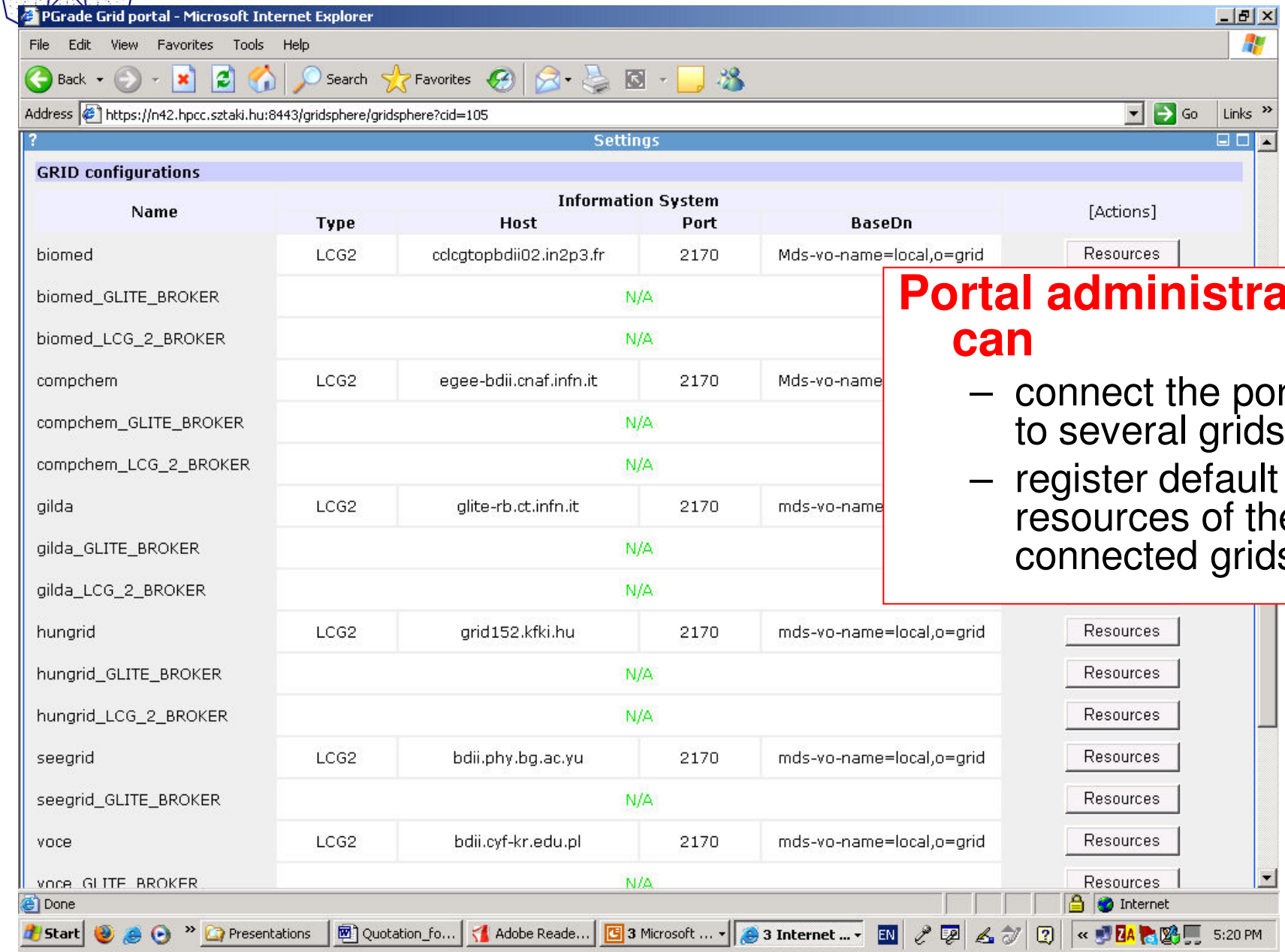
www.portal.p-grade.hu
pgportal@lpds.sztaki.hu

Backup slides to answer questions

Proxy delegations



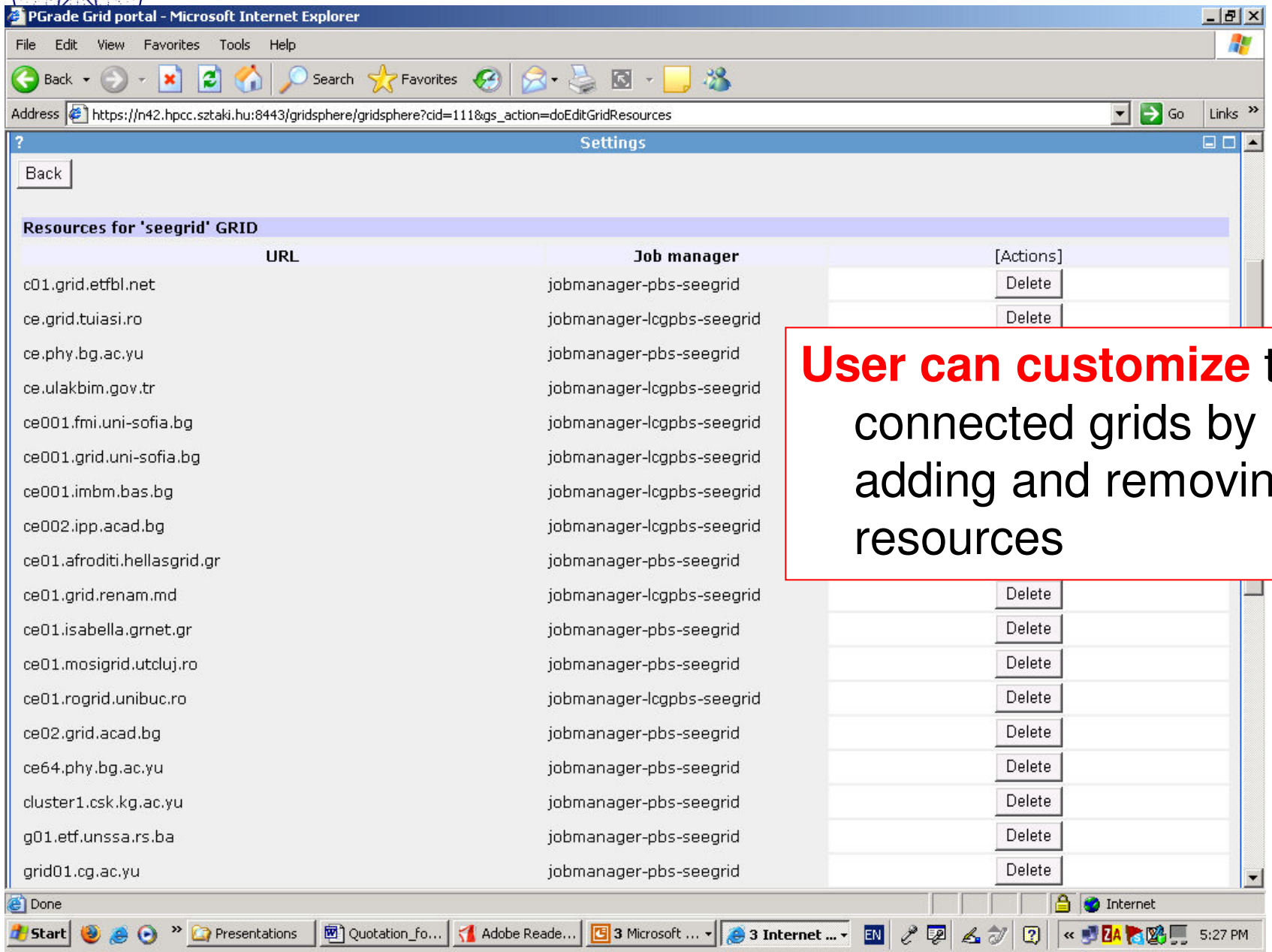
Settings



Portal administrator can

- connect the portal to several grids
- register default resources of the connected grids

Settings



User can customize the connected grids by adding and removing resources