



THE WHITE ROSE GRID e-Science Centre

Getting started with the P-Grade portal on the White Rose Grid systems at the Universities of Leeds, Sheffield and York

Author: Dr Joanna Schmidt
Date: April 2010
Edition 0.4 © Copyright 2010, J G Schmidt



Contents

1. Introduction	3
2. About the P-Grade portal	3
3. WRG computers	3
4. Becoming a P-Grade user.....	3
4.1 Obtaining a personal X.509 certificate	3
4.1.1 Personal certificates with a local authentication system	4
4.1.2 Obtaining a personal X.509 certificate from the UK e-Science Certificate Authority.....	6
4.2 Preparing your X.509 certificate for use within the P-Grade.....	7
4.3 Getting the P-Grade account.....	7
4.4 Login to the portal	7
5. Job submission	8
5.1 Setting user credentials.....	8
5.2 Resources known to the portal.....	10
5.3 Creating a job to back-end computers	11
5.4 Submitting a workflow to back-end computers	12
6. Job output	13
7. Job examples	13
8. Help and support	13
9. References	13
Appendix A	15

1. Introduction

This document provides a short introduction to the P-Grade portal which enables users from the Universities of Leeds, Sheffield and York to access the White Rose Grid (WRG) high performance computing (HPC) facilities from their desktops. This is a getting started guide to the P-Grade portal. It explains how to get access to the P-Grade portal, describes the necessary steps for job submission to WRG systems, and provides information about retrieval of output. It points users to examples of simple and more complex jobs (constructed as workflows) available for download as well as comprehensive documentation about the P-Grade portal.

The P-Grade portal enables users to submit and run a variety of jobs (e.g. a simple serial job, complex serial jobs executing on distributed resources as well as parallel jobs) on computational resources known to the portal. So far P-Grade has been proven to be extremely stable. It is anticipated that it could potentially be extremely useful to new WRG users who do not like to login to UNIX/Linux systems and would prefer to submit their jobs to WRG systems directly from their desktop.

2. About the P-Grade portal

The core of the P-Grade portal has been developed by the Laboratory of Parallel and Distributed Systems at MTA SZTAKI [1] in Budapest, Hungary. Further developments are undertaken by The P-Grade Portal Developer Alliance, which currently also includes the Middle East Technical University in Ankara, Turkey, and the University of Westminster [2] in London, UK.

The WRG has installed version 2.7 of the P-Grade portal [3] which enables the three Universities' users to use a web browser from their PC desktops for job submission and transfer of input files to high performance computing (HPC) systems and retrieval of output files from remote computers (Leeds, Sheffield and York). The portal has been customised specifically for White Rose users to offer them a simplified graphical interface.

Additionally portal computational resources include a node of the National Grid Service (NGS) hosted at the University of Leeds. Users may submit jobs to this node from the portal.

3. WRG computers

The White Rose Grid comprises four large computational systems of which two are located at the University of Leeds, one at the University of Sheffield and one at the University of York. It offers a heterogeneous computing environment for running computationally intensive codes requiring HPC systems. The back-end machines run a UNIX/Linux operating system.

Users from the Universities of Leeds, Sheffield and York may register on HPC computers at their own University or may apply for accounts on the White Rose Grid HPC computers at all three White Rose Universities. Information on how to register is available via email from the relevant Help Desks - at Leeds from helpdesk@leeds.ac.uk, at Sheffield from helpdesk@sheff.ac.uk, and at York from systems@wrg.york.ac.uk.

4. Becoming a P-Grade user

4.1 Obtaining a personal X.509 certificate

Users need to have X.509 digital certificates for job submission from within the P-Grade portal to back-end (remote) WRG computers. These are used to authenticate (identify) users on remote computers.

Two types of certificates might be used within the portal:

- 1) personal certificates using a local authentication system and the sarongs software (Shibboleth approach); this is the recommended approach;

or

- 2) personal certificates issued by the UK e-Science Certification Authority (CA). A short introduction to obtaining digital certificates is at: <http://www.ngs.ac.uk/certoverview>.

Users wishing to obtain and use their personal certificates issued directly by the UK e-Science CA need to refer to paragraph 4.1.2 and, if needed, approach the relevant helpdesks for further information (refer to paragraph 8).

4.1.1 Personal certificates with a local authentication system

Further information on personal certificates using the Shibboleth approach is available at: <https://cts.ngs.ac.uk/>

The following steps need to be executed:

1. Specify the following URL in the browser: <https://cts.ngs.ac.uk/> Select on this page the link to the NGS portal at <https://portal.ngs.ac.uk/>
2. On the initial NGS portal page click on the link [Authenticate](#) which is available on the left hand side
3. The next page (refer to Figure 1) will inform you that you have no Grid proxy credentials loaded. Click on the link [Login to UK Federation using Shibboleth](#)

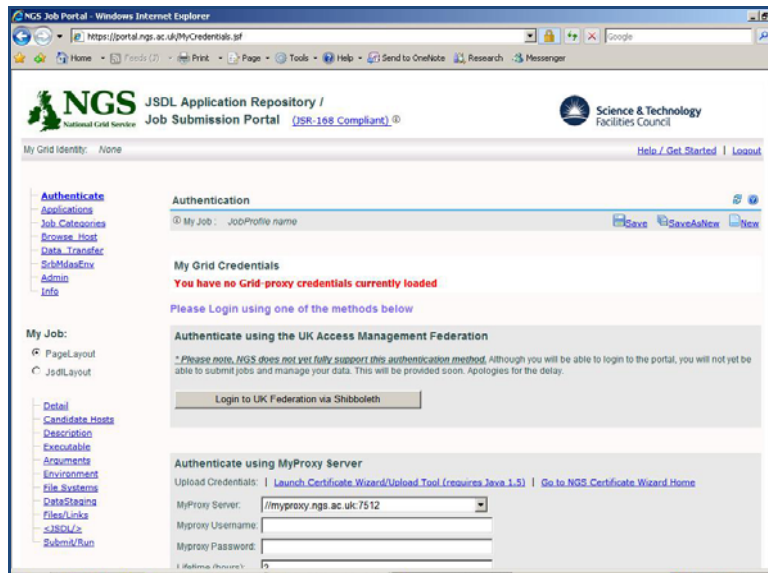


Figure 1: The NGS portal authentication page

4. You will be presented with the UK Access Management Federation page (Figure 2). Specify your institute from the drop down menu, and press the [Select](#) tab.

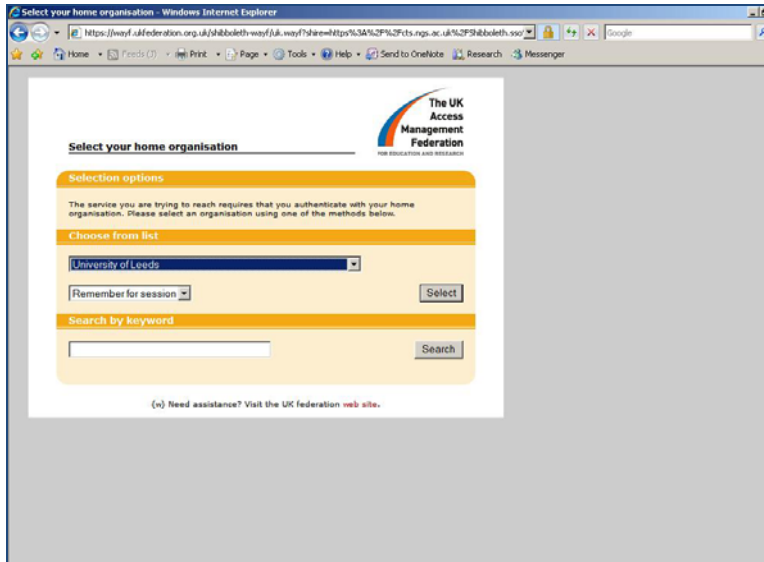


Figure 2: The UK Access Management Federation page

5. Now your local authentication system page will be shown. Enter your correct username and password for your local authentication system and you will be referred to your local user administration software to confirm your identity.
6. After your identity has been verified successfully you will be presented with the page confirming your identity (Figure 3).

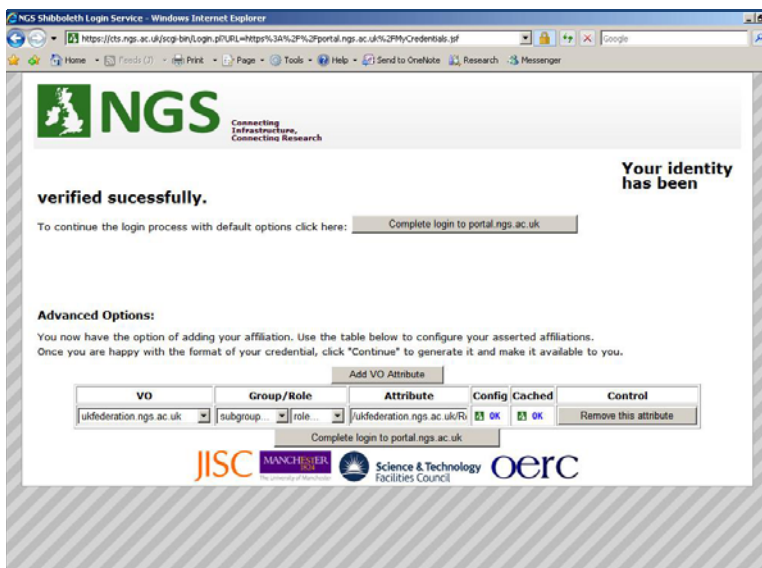


Figure 3: Identity verification page

7. Press Complete login to portal.ngs.ac.uk and you will be presented with a page providing you with details of your grid credentials. The example of these are below:

My Grid Credentials

You have Grid credentials (proxy) loaded

Grid Identity:/DC=uk/DC=ac/DC=ngs/DC=sarongs/CN=x1cryx469f060cdefr906546892gh3fbc8902hj2

The long string next to the Grid Identity field is your proxy certificate.

8. Email your system administrator your proxy certificate (specify the whole string) and request that is put into the grid-map file on the system you want to use. Leeds NGS node users will need to complete an on-line form for renewing resources. Further details are available from the relevant helpdesks (helpdesk@leeds.ac.uk , helpdesk@sheff.ac.uk, systems@wrg.york.ac.uk, support@grid-support.ac.uk) .

4.1.2 Personal X.509 certificate from the UK e-Science Certificate Authority

4.1.2.3 Obtaining Personal X.509 certificate from the UK e-Science Certificate Authority

The UK e-Science Certification Authority (CA) provides standard X.509 certificates for the UK community. A short introduction to obtaining digital certificates is at:

<http://www.ngs.ac.uk/certoverview>

All certificates issued will expire after one year, and if needed may be renewed.

WRG users need to identify their local Registration Authority (RA) before applying for digital certificates. They will have to visit the RA to have their identity vetted by the RA Operator. The following RAs are available for White Rose users:

University	Registration Authority	Name & email address
Leeds	ISS (visit the ISS Helpdesk, off the Edward Boyle Library corridor, level 10)	Stephen Corbett, Peter Corry, Barbara Edmondson, Jitesh Rathod (sysadmin@leeds.ac.uk)
Sheffield	CICS	Richard Gilbert, Paul Hodgson
York	York Computer Science	Aaron Turner (aaron@cs.york.ac.uk)

Users requiring digital certificates will need to:

- a) read the CA User Documentation available from: <https://ca.grid-support.ac.uk/cgi-bin/pub/pki?cmd=getStaticPage&name=index>
- b) check that there is the UK e-Science CA in your browser; otherwise download it to your browser from <http://www.ngs.ac.uk/cacerts>
- c) from the web browser access: <https://ca.grid-support.ac.uk/cgi-bin/pub/pki?cmd=getStaticPage&name=index> and select the tab Request a certificate; then click on User Certificate
- d) follow instructions on this page and complete the relevant fields; please note your PIN as you will need it later to identify yourself to your local Registration Authority
- e) visit your local RA (see the information above) with a photo identity document (the University ID card or your passport). You will be requested to provide the PIN that you specified in step d). Following your positive validation the RA Operator will authorise your request, which will be sent to the UK CA
- f) after receiving an email from the UK e-Science CA, usually 1 working day after issuing the request, you will be able to download your certificate from <https://ca.grid-support.ac.uk/cgi-bin/pub/pki?cmd=getStaticPage&name=index> by clicking on the tab Retrieve New Certificate or following email instructions
- g) install the latest Certificate Revocation List and test your certificate

4.1.2.4 Preparing your X.509 certificate for use within the P-Grade

P-Grade users will need to supply their credentials (digital certificates proxies) to remote computers so they can run their jobs on the back-end computers.

Currently it is recommended that users import their certificate from the web browser, where was originally stored, to users' backup areas, and then download its proxy to the MyProxy server. The MyProxy server will store time limited versions of users' credentials (proxy credentials) so that users can access them from within the P-Grade.

Further information on how to implement these two steps can be found at:

- Backing up your certificate <http://www.ngs.ac.uk/certbackup>
- MyProxy Upload <http://www.ngs.ac.uk/tools/certwizard>

4.2 Getting the P-Grade account

The White Rose Grid P-Grade portal can be accessed from each White Rose University campus from any browser by specifying the following URL: <http://portal.wrg.york.ac.uk/>

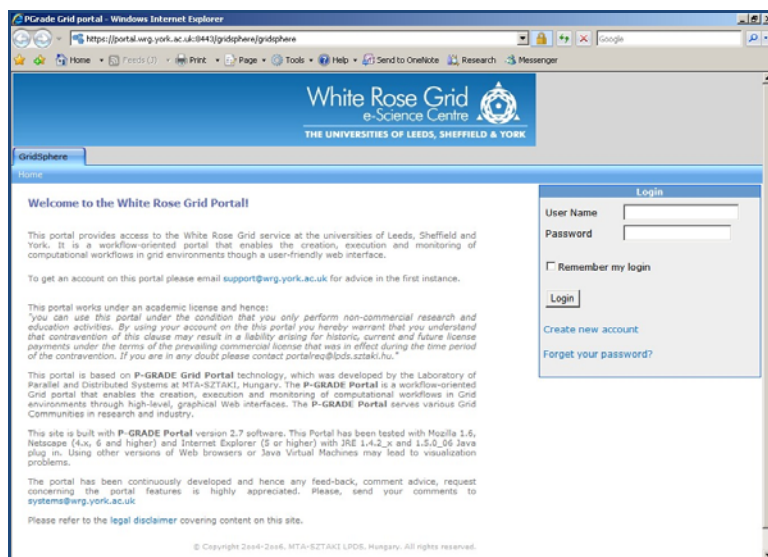


Figure 4: The P-Grade portal entry page

The entry page of the P-Grade portal has a box on the right hand side entitled `login` (refer to Figure 4). Click on the tab `Create new account`, and complete the required user information fields and press the save button. Following that an e-mail will be sent to you providing your account activation details. Now you should be able to access the WRG P-Grade portal.

4.3 Login to the portal

To login into the portal put the following URL: <http://portal.wrg.york.ac.uk/> into your browser and specify your username and password. Currently the P-Grade portal may be only accessed from the three Universities' campuses.

After receiving an email from the WRG portal administrator that states your username and password have been activated you will be able to login to the portal. You will be presented with the initial view which shows your profile settings (refer to Figure 5).

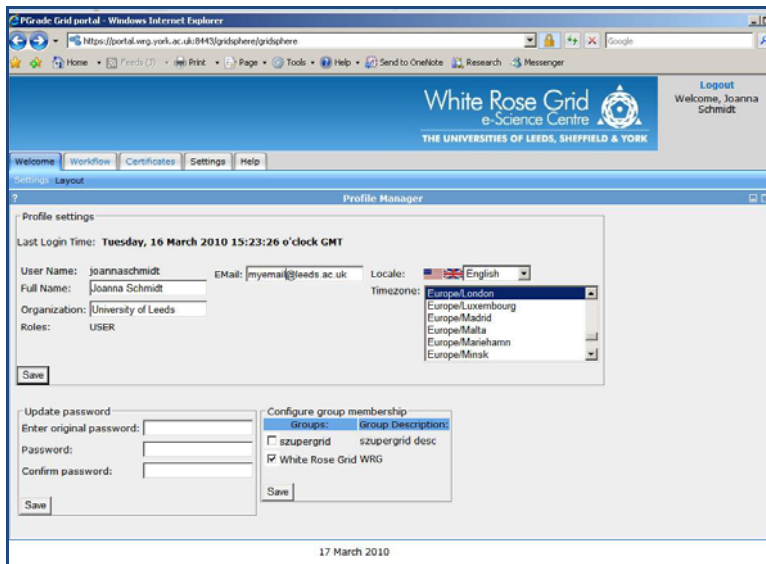


Figure 5: The P-Grade portal profile page

5. Job submission

Users need to have X.509 digital certificates for job submission to remote WRG computers from within the P-Grade portal. They are used to authenticate (identify) users on remote (back-end) computers.

5.1 Setting user credentials

Users with their personal certificates issued by the UK e-Science CA (refer to paragraph 4.1.2) need to approach the relevant helpdesks for further information.

Steps below should be followed by users who are using their personal certificates issued in conjunction with sarongs software and Shibboleth approach and using their local authentication system.

1. After login to the portal click on the **Certificates** tab at the top. This will bring up the **Certificates list** view (refer to Figure 6).

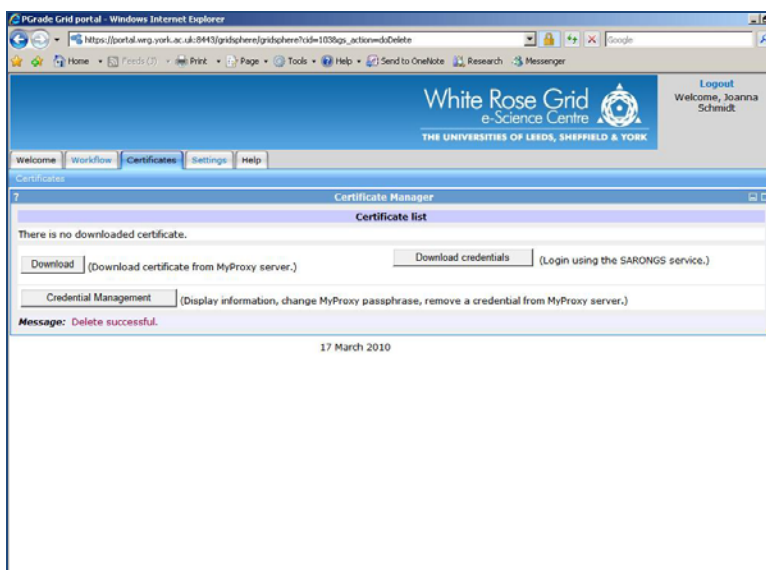


Figure 6: The Certificate list page

- Click on the Download credentials tab. The UK Access Management Federation page will appear (see Figure 7). Select your institute from the drop down menu, and press the Select tab. Now your local authentication system page will be shown. Enter your correct username and password for your local authentication system and you will be referred to your local user administration software to confirm your identity.

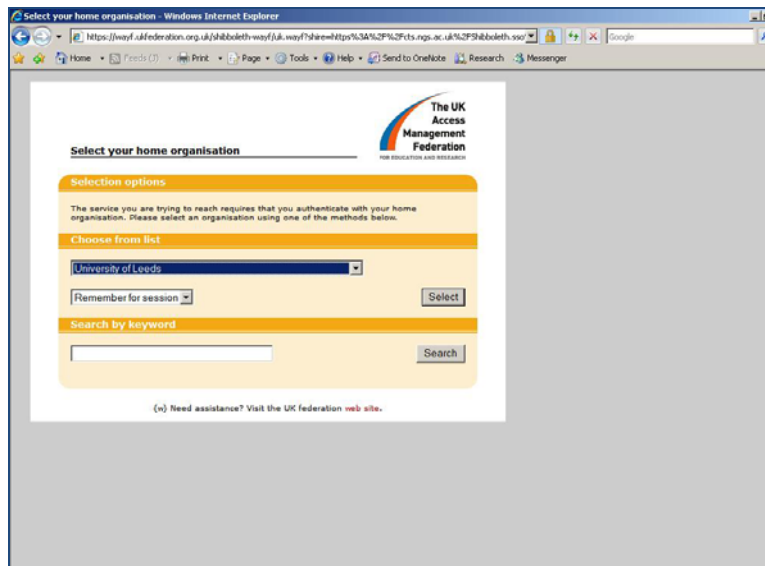


Figure 7: The UK Access Management Federation page

- Now click on the Complete login to portal.wrg.york.ac.uk tab at the top (refer to Figure 8).

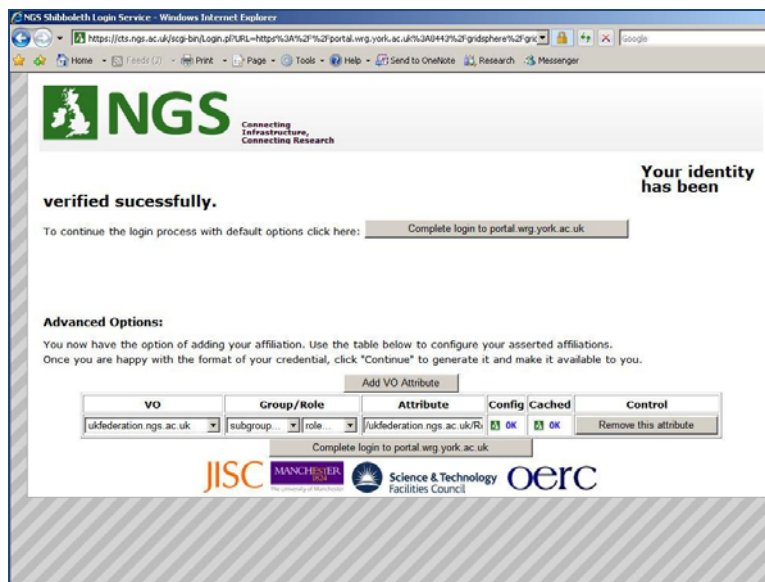


Figure 8: The sarongs page

- The next page will allow you to download your proxy credential from the MyProxy server. Your proxy credential has a 9 hour lifetime. Press the Download tab to get your proxy credential to the portal (Figure 9).

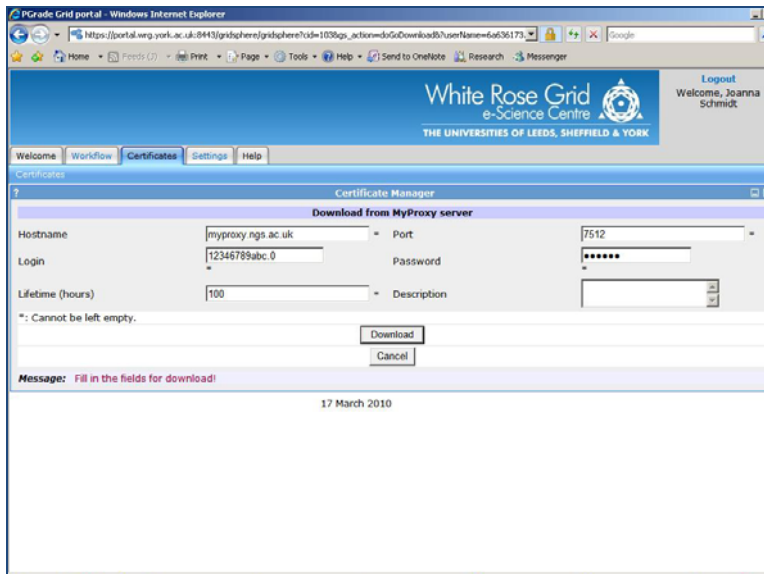


Figure 9: The screenshot of page to download a proxy certificate from the MyProxy server

5. You will be presented with the Set certificate for GRID page on which you need to click on the Set for Grid button (refer to Figure 10).

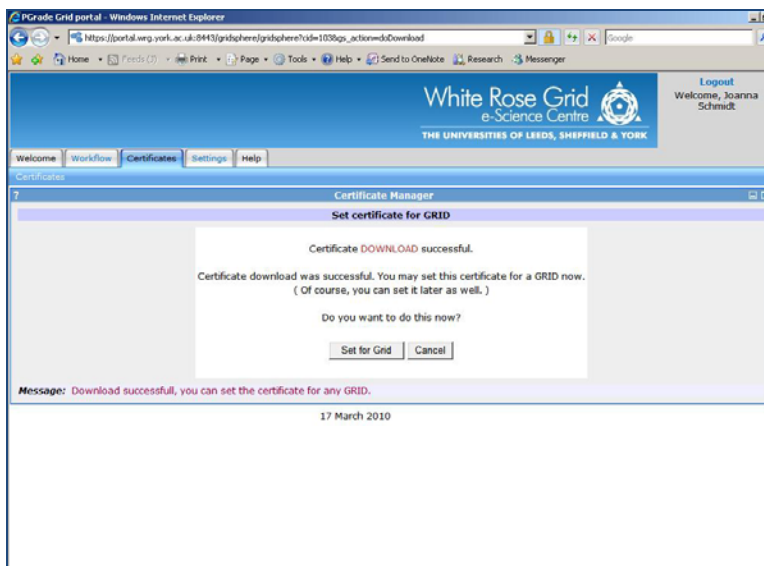


Figure 10: Screenshot of the page Set certificate for GRID

6. On the next page Setting certificate for the GRID you need to click on the ok button while the WRG is shown next to the Select from the list field.
7. The last page is the Certificate list and you should see your proxy certificate listed under the heading Certificate list. You are now ready to submit jobs to back-end computers.

5.2 Resources known to the portal

Users might submit jobs from the portal to the WRG systems and a node of the National Grid Service (NGS) at Leeds. At present access to the following back-end systems is available from the portal:

everest.leeds.ac.uk:/jobmanager-wrg
 globus.wrg.york.ac.uk:/jobmanager-sge
 iceberg1.shef.ac.uk:/jobmanager-sge
 ngs.leeds.ac.uk:/jobmanager-pbs
 snowdon.leeds.ac.uk:/jobmanager-sge

Further systems might be added on requests from users.

5.3 Creating a job to back-end computers

A job in P-Grade is constructed from a set of building blocks which form a workflow. Each of the building blocks must specify the application/software to be run and the processing system (back-end computer) to be used. Many jobs will specify an input file (dataset) and/or an output file (dataset). Optionally, arguments may be specified for each of the jobs in the workflow.

Click on the `Workflow` tab at the top and then on the `Workflow Editor` tab. This will open a new window in which you can create a workflow describing your job. Click on the `Workflow` tab at the top and select `New job` from the drop down menu. A yellow widget (2 of these are in Figure 12) will be created in the workflow editor view.

Click twice on this yellow widget and you will be presented with a dialogue box (Figure 11), whose purpose is to describe a single step of your job. Complete the relevant fields. You need to supply the following information: job name, job type (serial/sequential or mpi), name of application/executable-file, number of processes for your mpi job, parameters to be passed onto your executable (supply these in the `Attributes` field), and resource name which your job will be run on (e.g. `everest.leeds.ac.uk/jobmanager-wrg`), and to finish click on the `ok` button.

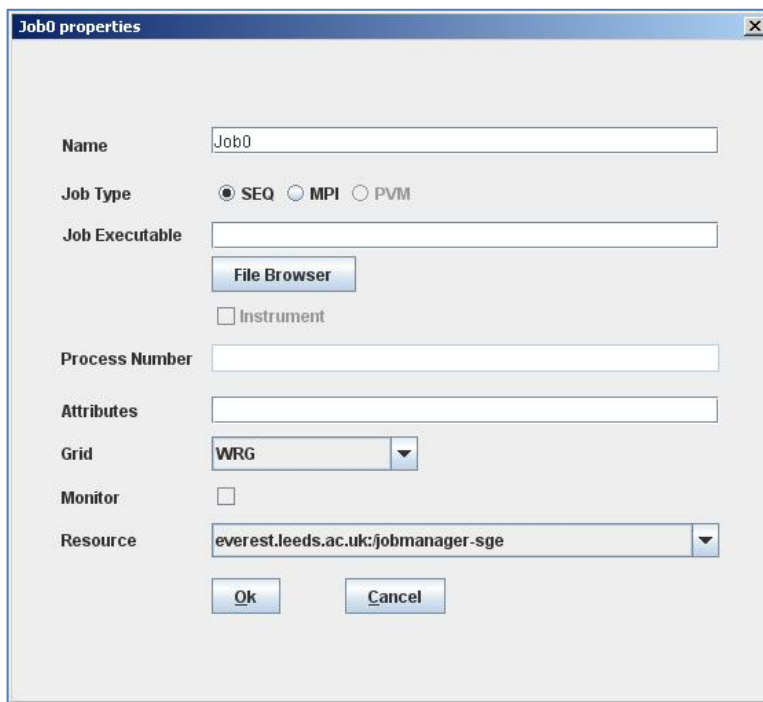




Figure 11: Job properties view

In the next step click on  and provide details of input and output files associated with your job (refer to Figure 13). Several input and output files might be specified. To add another job to  the workflow click on and follow the steps above to describe this job, and add input and output files. Join jobs by joining the output of the previous job with the input of the next one.

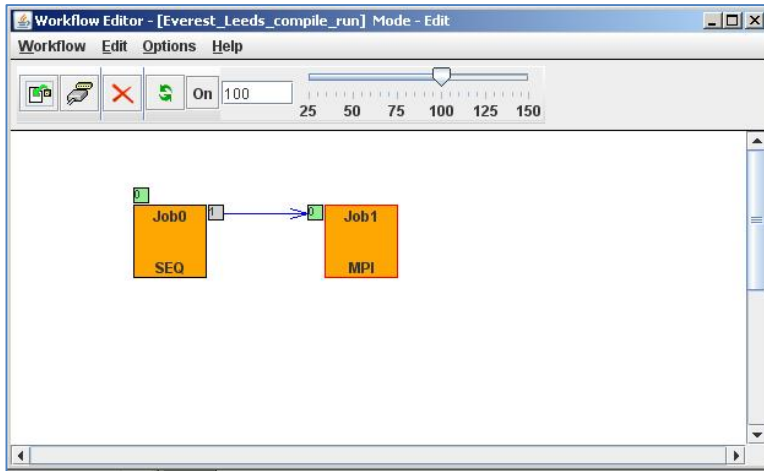


Figure 12: Workflow Editor view

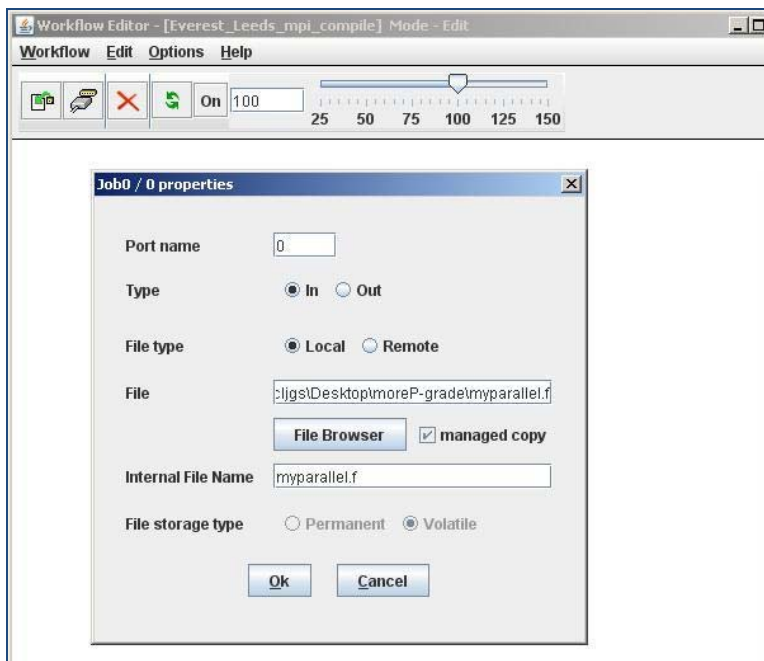


Figure 13: Screenshot of job input and output files page

A set of workflow examples (refer to section 7 and Appendix A) will help you to understand how to build more complex workflows reflecting the data flow in your computations. Figure 12 presents a workflow for the demo job Parallel 2 step compile and run Leeds.

5.4 Submitting a workflow to back-end computers

After you have created your own workflow or downloaded the example workflow (refer to section 7) you can submit your job to HPC back-end computers. This is simply done by pressing the `submit` button in the workflow view (Figure 14).

6. Job output

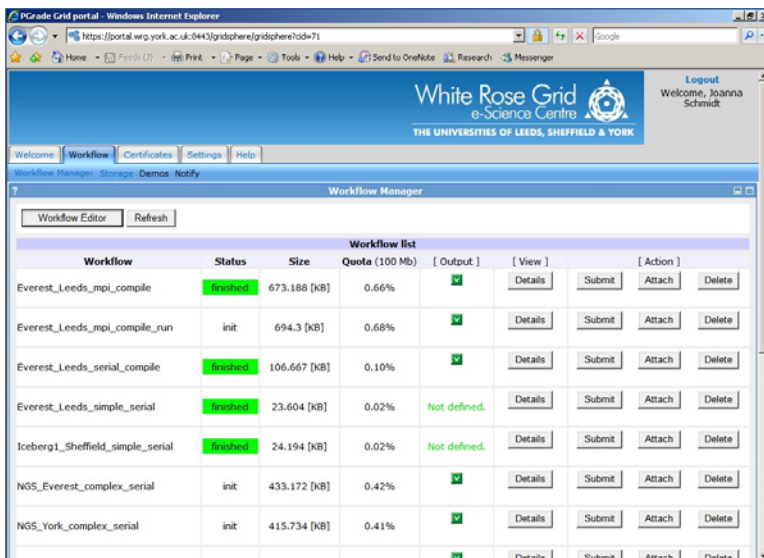
The status of jobs submitted to back-end computers is reported under the `status` tab within the workflow view (Figure 14). This view needs to be frequently refreshed to ensure that it is kept up-to-date.

7. Job examples

There are example workflows available for download. In the workflow view click on the `Demos` tab and then on the `Upload all` button. This will download all the example workflows to the user's P-Grade environment. Should you wish to download them to your desktop click on the `Storage` tab within the workflow view, select the workflow which you to download to your desktop, and then click on the `Download` button. This example workflow should now be saved on your desktop.

The example workflows will help you to build your own workflows describing your jobs. If you want to run these example workflows after downloading you will need to open the workflow editor view (Figure 12) and save them in this view prior to job submission as otherwise the example jobs will fail with the status `rescue`. Appendix A contains a list of examples and their description.

8. Help and support



Workflow	Status	Size	Quota (100 Mb)	[Output]	[View]	[Action]
Everest_Leeds_mpi_compile	finished	673.188 [kB]	0.66%	✓	Details	Submit Attach Delete
Everest_Leeds_mpi_compile_run	init	694.3 [kB]	0.68%	✓	Details	Submit Attach Delete
Everest_Leeds_serial_compile	finished	106.667 [kB]	0.10%	✓	Details	Submit Attach Delete
Everest_Leeds_simple_serial	finished	23.604 [kB]	0.02%	Not defined.	Details	Submit Attach Delete
Iceberg1_Sheffield_simple_serial	finished	24.194 [kB]	0.02%	Not defined.	Details	Submit Attach Delete
NGS_Everest_complex_serial	init	433.172 [kB]	0.42%	✓	Details	Submit Attach Delete
NGS_York_complex_serial	init	415.734 [kB]	0.41%	✓	Details	Submit Attach Delete

Figure 14: Workflow view

The P-Grade comes together with a built-in help file which contains comprehensive information on the portal. Users will need to click on the `Help` tab in the main view to access this documentation. Leeds users are encouraged to report any problems to the ISS Helpdesk (ext 33333; email: helpdesk@leeds.ac.uk); Sheffield users may email their Helpdesk: helpdesk@sheff.ac.uk, and York users need to report their problems to systems@wrg.york.ac.uk.

9. Caveat

The P-Grade version (2.7) currently installed on the WRG contains a bug which causes that input files associated with your workflow are not uploaded and updated after their modification. This means that if after the first run you have changed the input file associated with your workflow you will need to

take additional steps in order to have your input files updated. You will need to delete the port which includes the name of the modified input file and then re-instate it. This action will force the upload of the new (modified) input file with your workflow.

10. References

[1] Laboratory of Parallel and Distributed Systems at MTA-SZTAKI, Hungary:

<http://www.lpds.sztaki.hu/>

[2]University of Westminster Centre for Parallel Computing

http://www.cpc.wmin.ac.uk/cpcsite/index.php/Main_Page

[3] The P-Grade Portal: <http://portal.pgrade.hu/>

Appendix A

Example jobs available from the within of the P-grade portal:

Hello world: A simple "Hello World!" example (script) which also outputs the hostname of the machine/resource where the job was executed. This job is executed on the computer at York.

Whoami: Similar to the Hello World example; this prints remote (back-end) username and lists the environment variables settings. This may be useful for debugging. The job is submitted to York's resource.

Cross-site: This workflow has two steps involved; the output from the first job is used as an input for the second job. The individual jobs can be run on separate hosts and P-Grade should move the files around as required. The first job is executed on a resource at York and the second job is run on a resource at Sheffield.

MPI: This is a workflow that executes a single MPI job (an executable file) on the Leeds NGS resource. Should you wish to change where this job will run, the executable needs to be replaced with one compiled for the target resource.

MPI-compile-run-Leeds: This is a two-stage workflow. The first step is a serial job, which will compile the input source code file (an MPI "Hello World", written in C) on the Leeds NGS resource. The compiled executable is then routed as the input to a second job which executes it in parallel on NGS Leeds.

Serial NGS job: a simple serial job to the NGS Leeds system that outputs a hostname.

Serial NGS compile: A simple serial job to compile a parallel code on the NGS system at Leeds. It loads intel compiler and uses mpif90 to compile the Fortran code. It issues the following commands: `source /etc/bashrc; module list; module add 64-bit intel-compilers; mpi module list; mpif90 myparallel.f; exit.`

Parallel NGS two steps (compile and run) home directory: This example compiles a parallel Fortran code and then runs the executable on 2 cores on the NGS system at Leeds. The workflow comprises 2 separate jobs. The first job compiles the Fortran code and saves the executable in the user's home directory. The second job sets the permissions of the executable code to "executable" and runs this code. To run this job the user needs to specify correctly the full path for `gsiftp` in the File field of channel 1 (refer to Figure 13; e.g.

`gsiftp://ngs.leeds.ac.uk//nfs/ift02_h01/home/ngs0013/a.out`). This job, if not modified, will not run successfully under a different user account as the home directory needs to point to the user's home directory.

Parallel NGS two steps (compile and run): This example compiles a parallel Fortran code and then runs the executable on 2 cores on the NGS system at Leeds. It contains a workflow comprising 2 separate jobs, one for compilation and one for running an executable code. This is a similar example to the previous example; the difference is that the executable file is not saved in the user's home directory.

Simple parallel NGS: A simple job to run a parallel code (mpi) on the NGS system at Leeds. This job includes an executable code (a.out).

Simple serial Sheffield: A simple serial job to Sheffield (Iceberg1) that executes the hostname command.

Simple York: A simple serial job to the `globus.wrg.york.ac.uk` system at York. It outputs the hostname.

Complex serial four steps NGS Leeds and York: This is a complex serial job. The workflow comprises 4 steps. It compiles 2 codes on the NGS node at Leeds and then transfers both executable files to the `globus.wrg.york.ac.uk` system at York. It then sets the "executable" permission on the first file and runs this executable. The output from this job is directed to the second executable which outputs the number "92".

Parallel compile Leeds: This example compiles an mpi program on Everest at Leeds; and returns the executable file.

Simple serial Leeds: This is a simple serial job example that prints the hostname of the Everest system at Leeds. The job is submitted to the *Everest* at Leeds.

Complex serial four steps NGS Leeds and Everest Leeds: This is a complex serial job that comprises 4 steps. Two programs are compiled on the NGS node; then their permissions are changed to "executable"; the output from the first job is read by the second executable which is run at Leeds (Everest). After a simple calculation is carried out the number 92 is printed. This job is very similar to the "Complex serial four steps NGS Leeds and York" job which is computed on NGS and the `globus.wrg.york.ac.uk` system at York.

Parallel two steps (compile and run) Leeds: This job comprises a two steps workflow; in the first step a parallel code is compiled and then in the second step it runs this code on 4 cores at Leeds (Everest).