NGS
National Grid Service

The NGS Applications Repository / Grid Portal

Slides selected from a more detailed presentation by
David Meredith
NGS + Grid Technology Group, e-Science Centre, Daresbury Laboratory, UK

d.j.meredith@dl.ac.uk

NGS Portal

The NGS portal can be used to access and interact with the HPC and Data resources available on the NGS Grid via (Certificates + myproxy):

- Browse for different applications available on a Grid this includes your own personal applications and pre-configured applications (e.g. the NGS is currently publishing applications within the NGS portal to be made easily available for its users).
- Create descriptions of applications and their associated resources (using standard Job Submission Description Language).
- Submit/monitor compute jobs/applications.
- Access and move data.

Portal and Portlets

- The NGS Grid portal extends a JSR-168 compliant portal container which hosts a selection of ‘portlets.’
- Portlets are online-accessible applications that are hosted and managed within the portal container.
- The list of portlets that are deployed to the portal make up the portal’s overall functionality (users may be interested only in a selection of portlets)
- Main Benefit: Portlets facilitate the sharing and re-use of applications (168-compliant 3rd party portlets can be used within a portal as required).
- NGS and SciTech e-Science Centre are developing a collection of portlets designed for the Computational and Data Grid.
- The NGS portal is a current implementation of these Grid portlets.

NGS portal main functionality

1. A searchable database of personal and shared JSDL job profile documents (Job Submission Description Language, i.e. ‘job recipes/templates’).
   - JSDL can be browsed for, selected, loaded and saved in order to run applications on the Grid (loaded either ‘out-of-the-box’ or, more usually loaded, modified/tweaked as required and saved).
   - JSDL can be searched for by category of interest in the portal (e.g bioinformatics, chemistry, tutorials/examples).
   - JSDL documents can be pre-configured and published by the portlet administrator(s) to be made available to all other users.
2. A JSDL GUI editor for graphically authoring, validating, sharing, uploading jobs described in JSDL.
3. An extensible Grid job submission and monitoring application for heterogonous Grids (currently, only Globus but more Grid middleware providers are being added, e.g. JDL/gLite).

JSDL – Job Submission Description Language

NGS National Grid Service

Ali Anjomshoaa, Fred Brisard, Michel Drescher, Donal K. Fellows, William Lee, An Ly, Steve McGough, Darren Pulsipher, Andreas Savva, Chris Smith

- **JSDL 1.0 is an OGF recommendation**
- **JSDL 1.0 is published as GFD-R-P.56** –
  - Description of JSDL elements and XML Schema,
  - Available at: **http://www.ggf.org/gf/docs/?final**

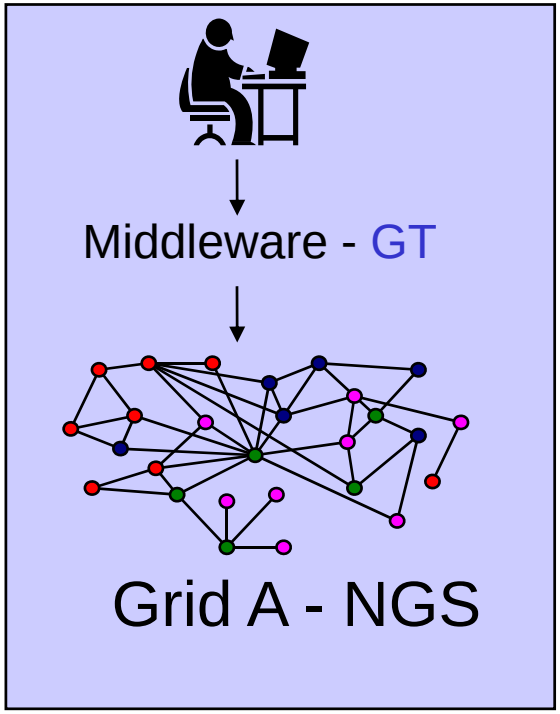
A language for *describing the requirements of computational jobs for submission* to Grids and other systems.

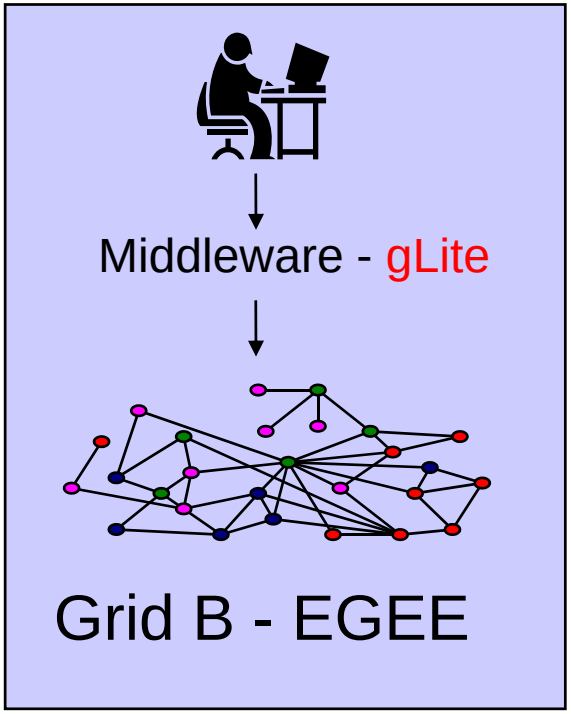
```
<jsdl:Application>
    <jsdl:ApplicationName>gnuplot</jsdl:ApplicationName>
    <jsdl-posix:POSIXApplication>
      <jsdl-posix:Executable>
        /usr/local/bin/gnuplot
      </jsdl-posix:Executable>
      <jsdl-posix:Argument>control.txt</jsdl-posix:Argument>
      <jsdl-posix:Argument>DavesControlFile.txt</jsdl-posix:Argument>
      <jsdl-posix:Input>input.dat</jsdl-posix:Input>
      <jsdl-posix:Output>output1.png</jsdl-posix:Output>
    </jsdl-posix:POSIXApplication>
  </jsdl:Application>
  <jsdl:Resources>
   ....
```

Example fragment

Why JSDL

Grids are currently heterogeneous:





- Different middleware implementations adopt different methods for the description of applications and their associated resources, and for their subsequent execution to a Grid.

- A Number of different data storage resources are also relevant for the management and transfer of data, e.g. GsiFTP, SRB, SRM, WebDav and (S)FTP.

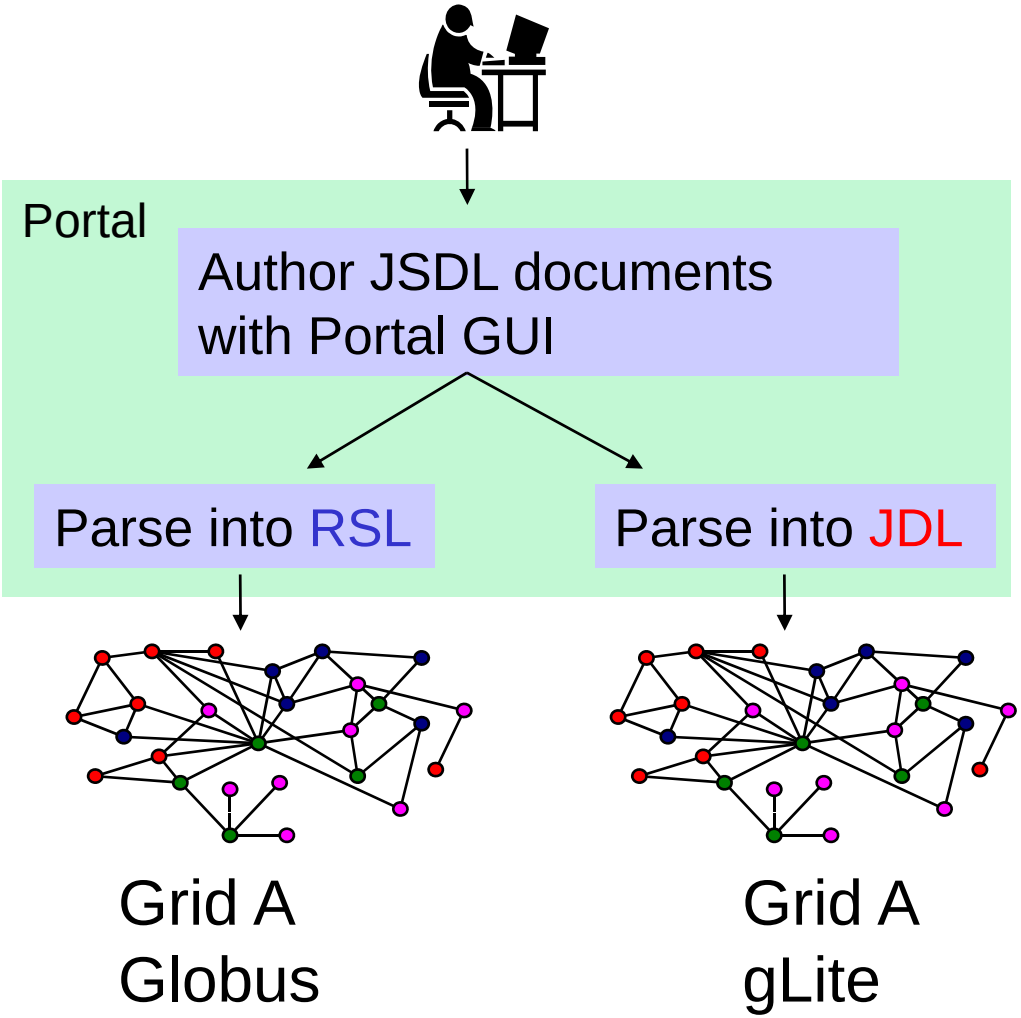
Grid A Globus RSL (Resource Specification Language)

```
&(executable=$(GLOBUSRUN_GASS_URL)/home/ngs0153/cpi) (arguments= 30 fileA)
(jobType=mpi) (environment = (NGSMODULES mpich-gm/1.2.5..10-
intel8.1:intel/fce/9.1.032) (TMP /tmp)) (count = 4) (hostCount = 8) (minMemory = 512)
(maxWallTime = 3)  (directory=/home/ngs0153) (stdin=/home/ngs0153/cpi.in)
(stdout=/home/ngs0153/cpi.out) (stderr=/home/ngs0153/cpi.err)
```

Grid B gLite JDL (Job Description Language)

```
Type = "Job";
JobType = "Normal";
RetryCount = 3;
Executable ="/home/ngs0153/cpi";
Arguments = "30 fileA";
VirtualOrganisation = "myGridVOproject";
StdInput = "cpi.in";
StdOutput = "cpi.out";
StdError = "cpi.err";
InputSandbox = { "gsiftp://grid-data.rl.ac.uk:2811/home/ngs0153/cpi", "gsiftp://grid-
data2.dl.ac.uk:2811/myhome/fileA" };
InputSandboxDestFileName = { "cpi", "fileA" };
OutputSandbox = { "cpi.out" };
OutputSandboxDestURI = { "gsiftp://mygridhome.dl.ac.uk:2811/myhome" };
DeleteOnTermination = { "fileA" };
Environment = { "NGSMODULES=mpich-gm/1.2.5..10-intel8.1:intel/fce/9.1.032", "TMP=/tmp" };
Requirements = ( other.GlueCEInfoLRMSType == "PBS" ) && ( member( GlueCEInfoHostName,
{"grid-data.rl.ac.uk:2119" , "mygrid-resource.dl.ac.uk:2119" } ) ) && ( GlueHostProcessorModel ==
"Intel" );
Rank = -other.GlueCEStateEstimatedResponseTime;
```

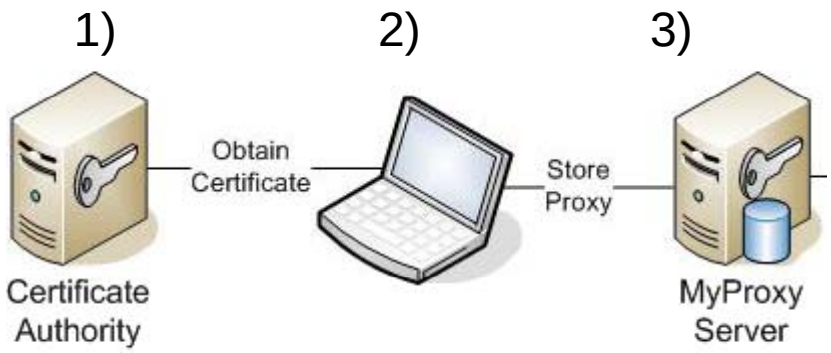
JSDL and the NGS Grid portal



- One interface definition for job submission
- One job description language
- JSDL is a standard job submission description language (middleware agnostic)

MyProxy Login Procedure

NGS National Grid Service



1) **Obtain a certificate** from CA - .p12 / .pfx certificate bundle (one time only)

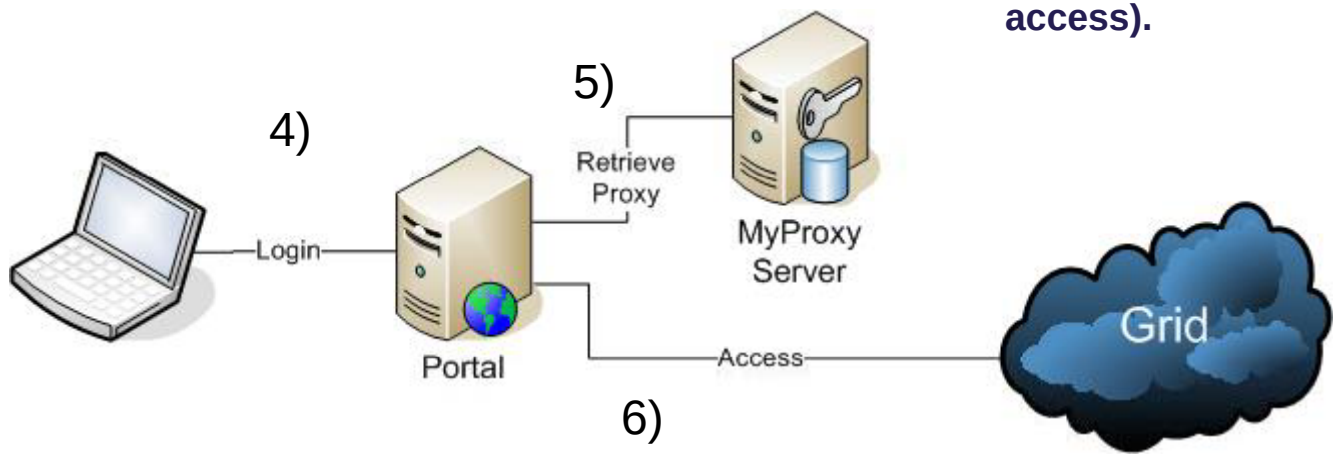
2) **Create a public/private key pair** - usercert.pem / userkey.pem (one time only)

3) **Store a proxy in UK eScience MyProxy** server with un/pw – delegation, private key never leaves own computer (done temporarily)

4) Can **login to portal from any computer/browser** using same un/pw.

5) **Portal retrieves proxy** using same un/pw.

6) **Access Grid** (portal uses proxy to authenticate to all Grid resources giving SSO access).



JSDL v 1.0

1. **XML Schema language for describing compute jobs in a platform independent language (XML).**
2. **Is agnostic of middleware - no dependencies on Globus, WSRF, gLite (portal that is generic and not tied to any particular set of Grid technologies).**
3. **OGF (once GGF) recommendation.**
4. **JSDL documents can be validated against the JSDL and JSDL POSIX XSD Schema documents to ensure correctness.**
5. **A JSDL XML *document* describes the requirements of a compute job and its associated resources, e.g. description of required files / data**
   - ***What to do, not how to do it***
   - **It is an extensible template language (elements defined in another XML schema with a different namespace can be added to a JSDL document). This facilitates customisation and nesting of extra information.**
   - **Is loosely typed – does not impose tight controls on the values of elements using advanced XML schema restriction techniques such as xsd:patterns**
6. **Not included in JSDL version 1.0**
   Scheduling, Workflow, Security …

Edited highlights

- The following few slides show some of the functionality of the Repository – not all!

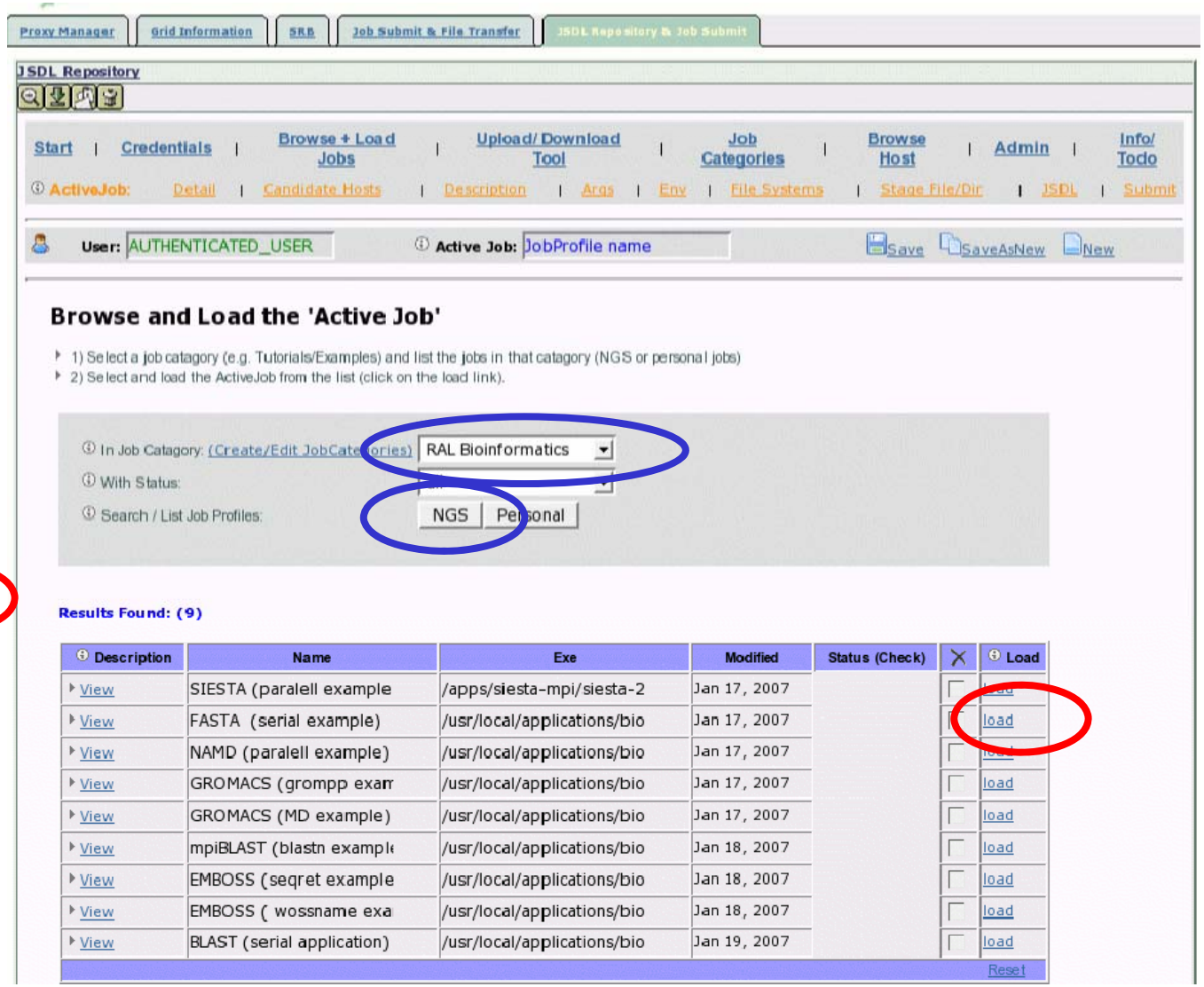
JSDL Repository / Database

Portal is 'open', are free to browse pubic JSDL documents without log-in and free to use portal as open JSDL editor.

Login to browse personal applications and submit jobs.

List jobs, read job descriptions and load a job to initialise the 'Active Job.'

Changes to the parameters in the GUI will update the generated JSDL automatically.

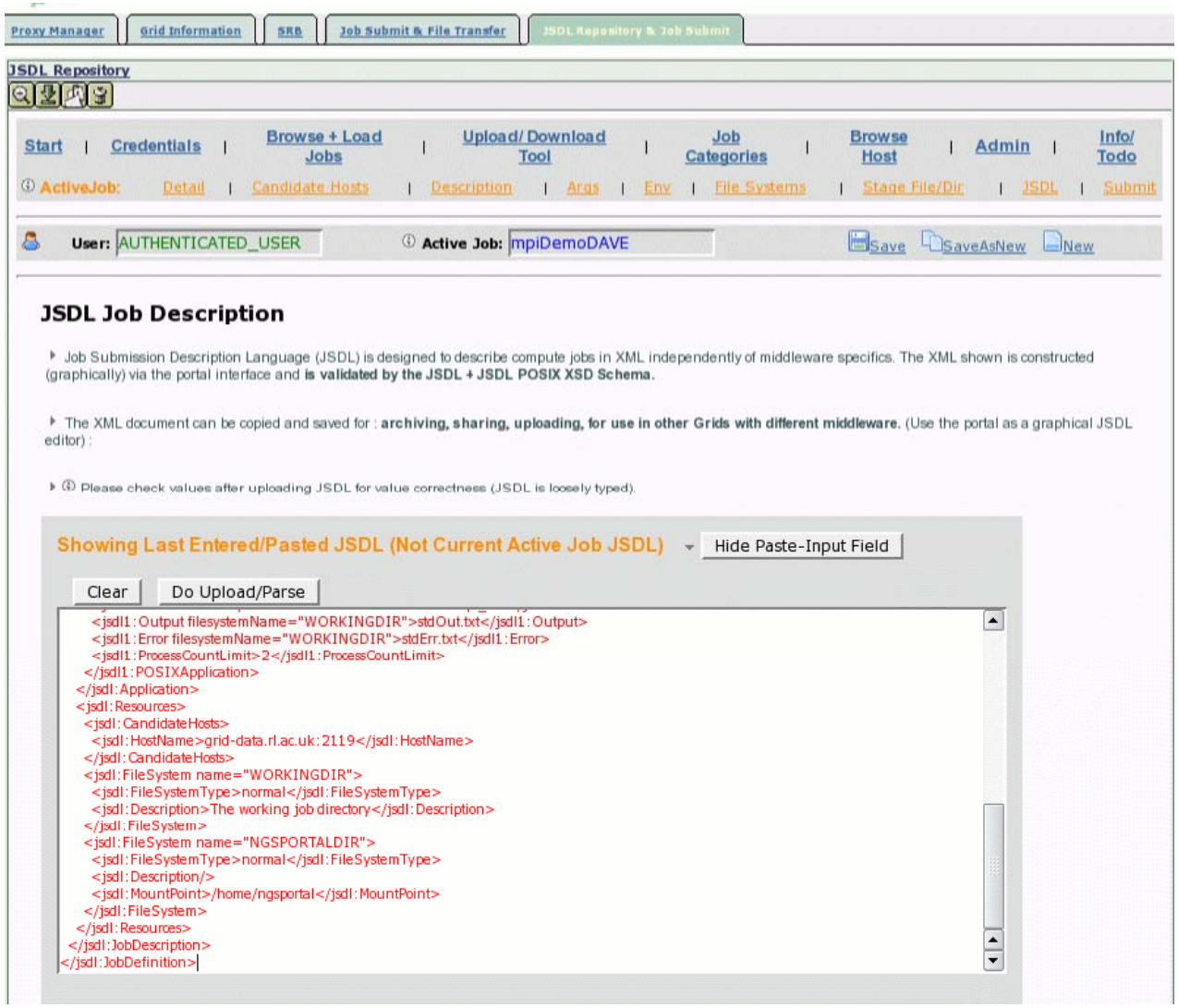


Upload / Share JSDL

JSDL documents can be uploaded to the portal in order to initialise the Active Job.

Validation errors and messages are displayed in the interface when uploading JSDL.

The portlet allows the sharing of job profiles and JSDL between users and user communities.



Stage Data

Compile a list of data from across the Grid that should be copied to the consuming system prior to job execution (source URI). Also after job completes (target URI).

JSDL does not mandate the protocol / URI format.

Data is staged relative to named file systems on the consuming system.

The source URI can be either specified manually or, more normally, browsed for in the 'Browse Host' page.


```
<jsdl:DataStaging>
  <jsdl:FileName>dave.txt</jsdl:FileName>
  <jsdl:FilesystemName>WORKINGDIR</jsdl:FilesystemName>
  <jsdl:CreationFlag>overwrite</jsdl:CreationFlag>
  <jsdl:DeleteOnTermination>true</jsdl:DeleteOnTermination>
  <jsdl:Source>
    <jsdl:URI>gsiftp://grid-data2.dl.ac.uk:2811/home/ngs0153/dave.txt</jsdl:URI>
  </jsdl:Source>
</jsdl:DataStaging>
```

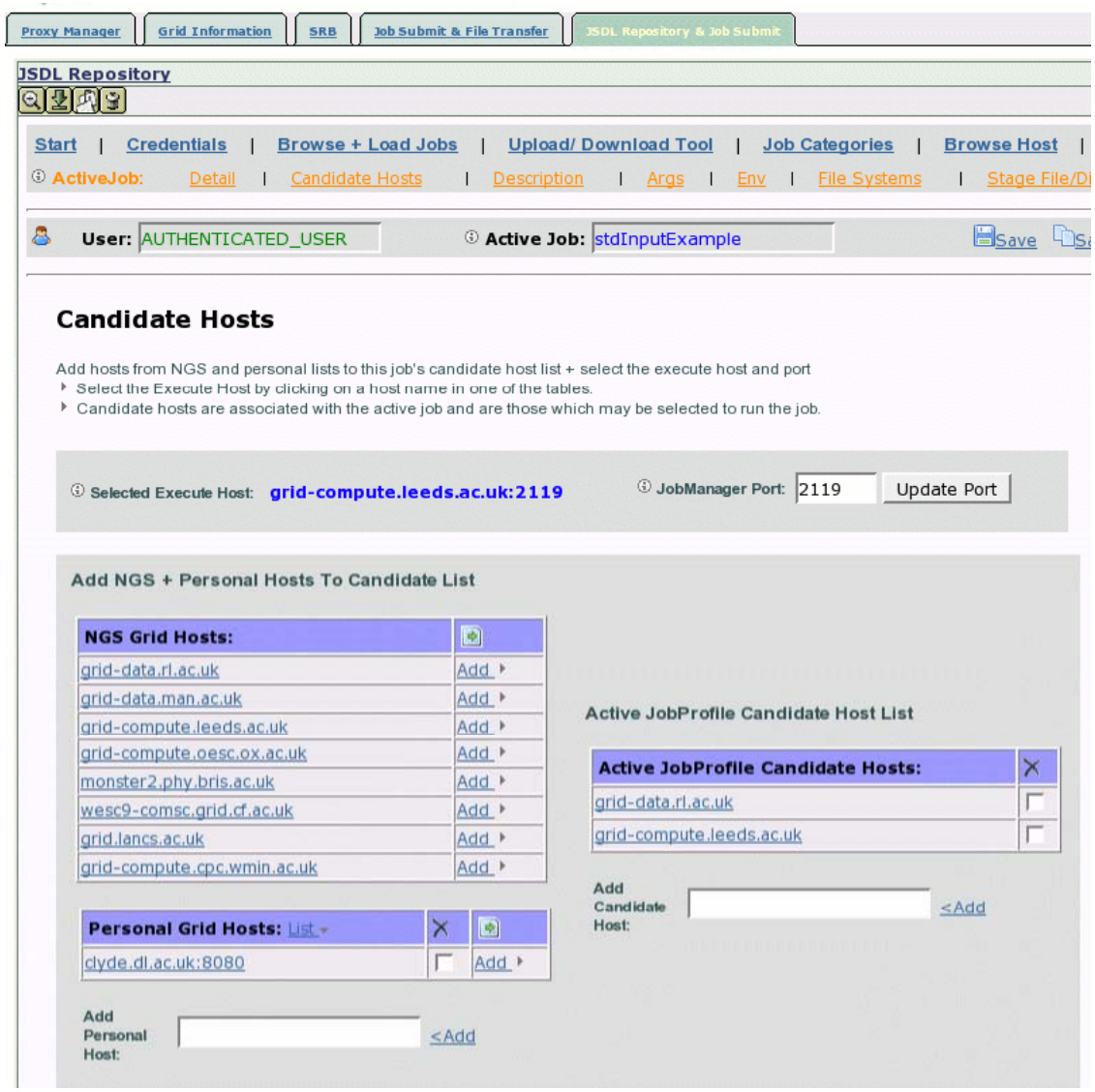
Candidate Hosts

Candidate hosts are consuming systems (execute hosts) that can be to run the Active Job.

The candidate host list can be compiled from a personal host list and from a default host list (available to all users).

In future, a RB matchmaking will be used to auto-compile the candidate host list.

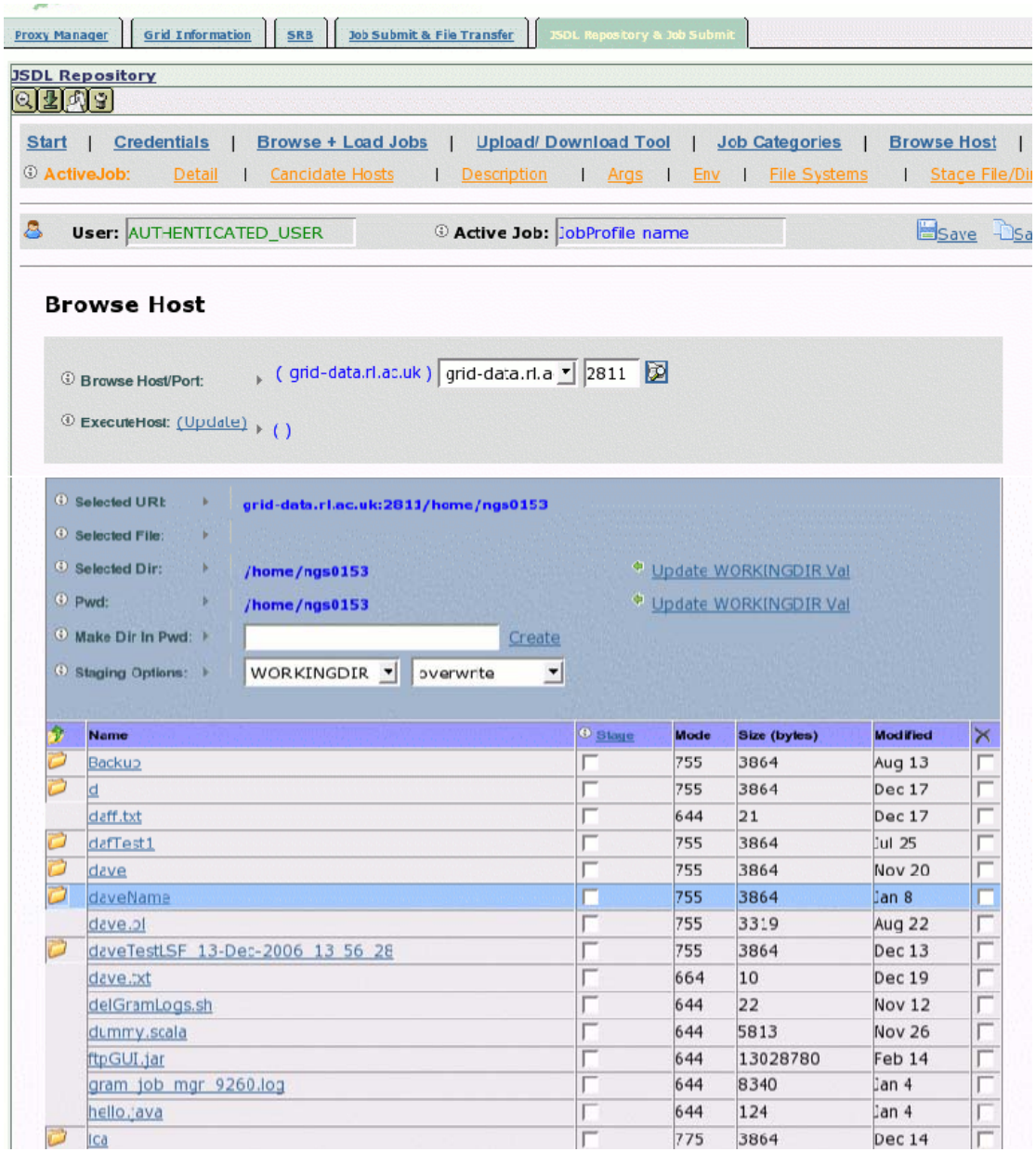
```
<jsdl:CandidateHosts>
   <jsdl:HostName>
      grid-data.rl.ac.uk:2119
   </jsdl:HostName>
   <jsdl:HostName>
      grid-compute.leeds.ac.uk:2119
   </jsdl:HostName>
</jsdl:CandidateHosts>
```



Browse Hosts

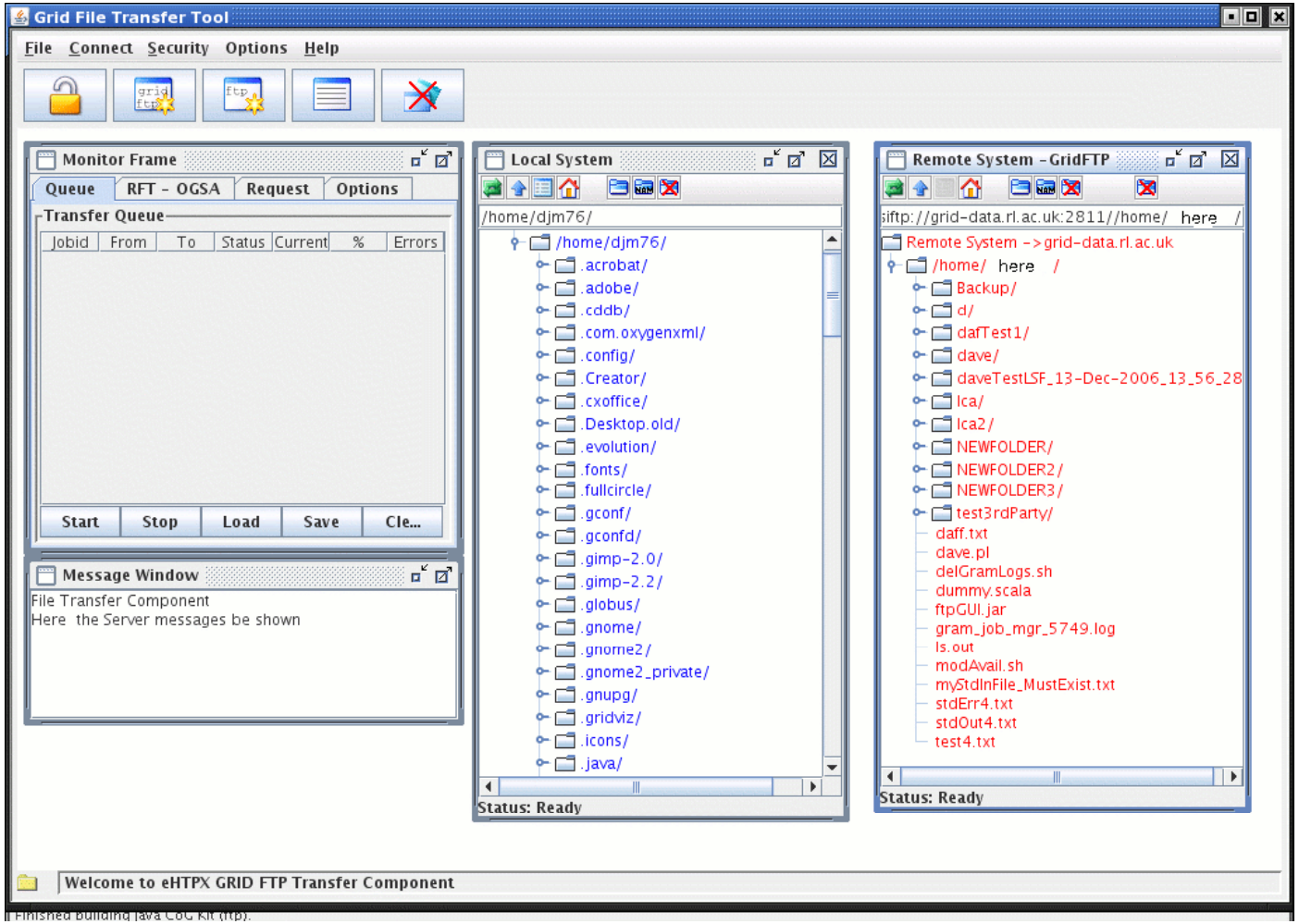
Browse remote Grid hosts for data required for staging.

Select files and directories that should be copied to the consuming system via GsiFtp (more protocols to be supported inc srb, webdav).



GridFtp Upload / Download / File Transfer Tool

Transfer files to/from your desktop and a remote Grid ftp server via ‘drag-n-drop’ – akin to Windows explorer for the Grid (in serious need of update – currently using depreciated libs).



Summary

1. Open / free to use JSDL GUI editor, browse the list of applications hosted to date. Please contact NGS to request more hosted applications (currently an ongoing task, more will appear soon).
2. The application is designed to be generic and not tied to any particular set of Grid technologies (facilitated using JSDL). Middleware dependencies added at job-submit/monitor time when the specifics of the middleware have to be accommodated (e.g. parsing the JSDL into RSL, adding mw-specific parameters, e.g. RSL JobType).
3. Application can be deployed as a JSR-168 portlet or as a standalone Web application. Portal can therefore be hosted in project specific portals.
4. Currently, application only supports Globus, but GridSAM and gLite JDL (NGS RB) are scheduled for addition (more on GridSAM next few days).

TODO / Future

NGS National Grid Service

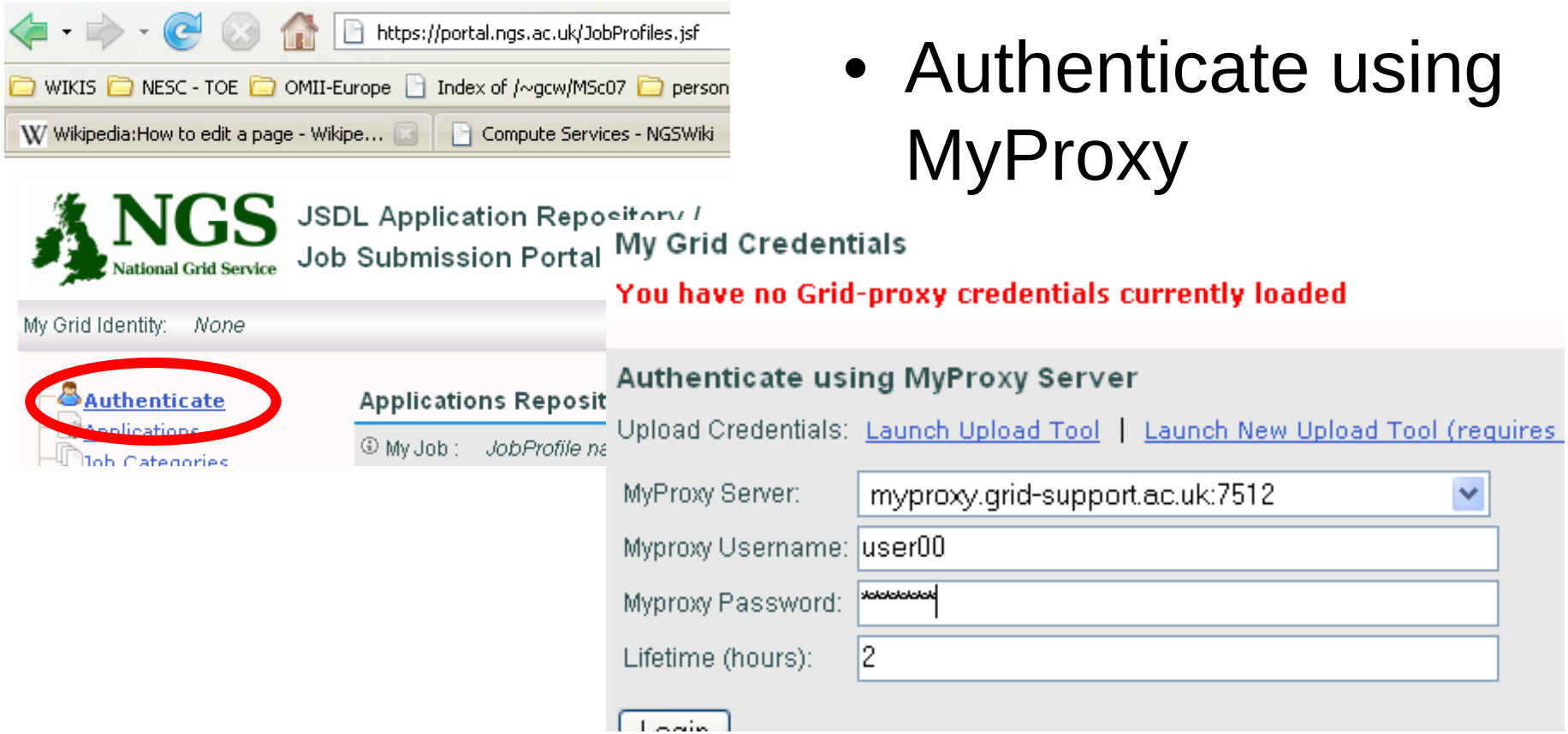
- **Extend portal to accommodate more/new middleware, e.g. NGS gLite resource broker and JSDL conversion to JDL.**
- **Parametric jobs (depends partly on parametric schema extensions for JSDL and subsequent middleware support.**
- **Extend application to support staging from more Data Grid + Web protocols (SRB). This will involve browsing / interacting with different data storage resources within the interface (e.g. browse SRB). Performing stating across different protocols adds some complexity.**
- **Growing list of improvements / suggestions to refine interface**
- **Release the application for use in other projects / Grids (scheduled for OMII open source release).**
- **Grid will probably move to a more SOA style architecture – more use of temporary 'sandboxes' for job execution rather than provision of permanent user accounts on compute resources (staging become more important).**

***** A portal is only as good as the underlying deployed infrastructure…. portal development often involves debugging the underlying consuming systems and middleware *****

Software Stack

1. **JSF (Java Server Faces) for server application interface development: GUI interface components + MVC controller (Http session and request scope data). Vanilla JSF is JSR-168 compliant so can host applications as Web app or portlet.**
2. **Spring v2.0 for managing objects in an n-tier server application (recommended in a lightweight non J2EE application server, e.g. Tomcat/Jetty), e.g. managing singleton + prototype object graphs, declarative transaction demarcation, data source management, propagation of persistence context across long running transactions.**
3. **C3p0 db connection pooling.**
4. **JPA (Java persistence API) for ORM (object relational mapping). Hibernate 3.2 for domain model (could use Kodo, Toplink, apache openJPA)**
5. **CogKit for Globus API from Globus Alliance.**
6. **Xml-to-object data binding framework – Apache XMLBeans (or JAX-B).**

Go to https://portal.ngs.ac.uk/



- Authenticate using MyProxy

- Then in another browser window, follow link on agenda page