



ICAT Job Portal

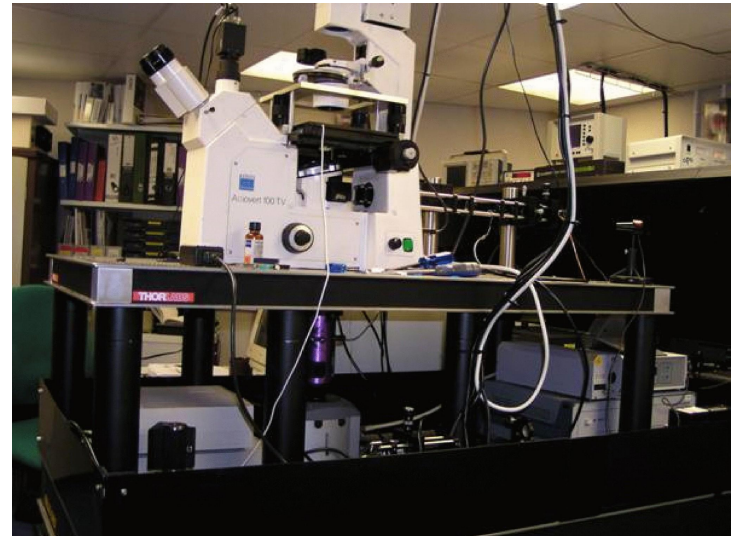
**a generic job submission system
built on a scientific data catalog**

24-25 March 2014

Steve Fisher, Kevin Phipps and Dan Rolfe
Rutherford Appleton Laboratory - STFC

Use Case

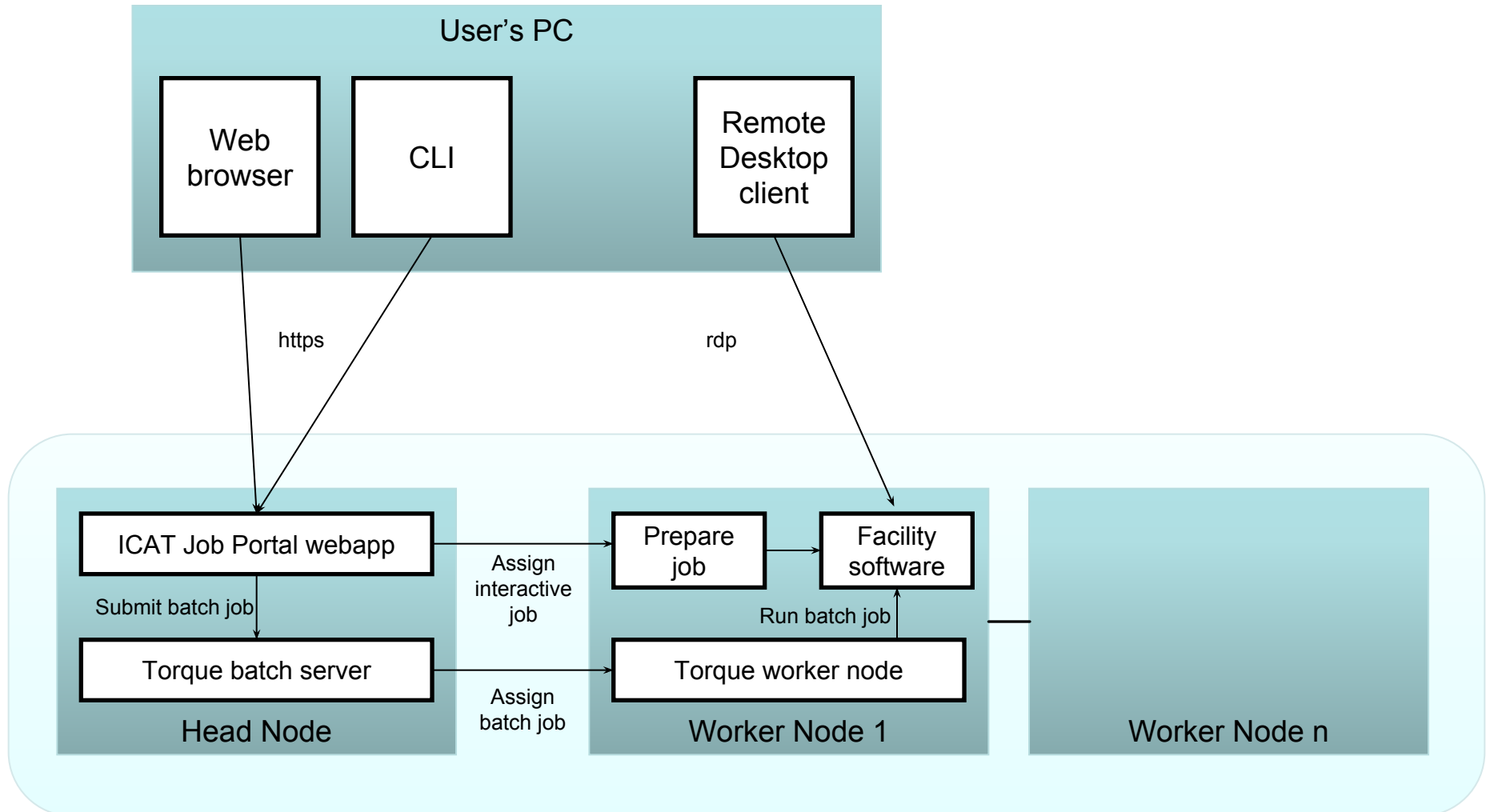
- LSF operate OCTOPUS imaging cluster: lasers coupled to interconnected microscopy stations.
 - a large number of data files
 - applications to process and visualise them
 - interactive program with an easy to use GUI to offer lists of raw and processed datasets and offer the ability to process those datasets
- Some requirements
 - GUI and command line from on and off site.+
 - Consult the metadata to locate the data.
 - Submit batch jobs to Linux nodes; listing, cancelling and retrieving output.
 - Interactive GUI based analysis/visualisation jobs able to access data.
 - Select and submit multiple datasets for processing through applications.
 - No facility dependence: configurable menus, datasets types, jobs and associated job parameters.



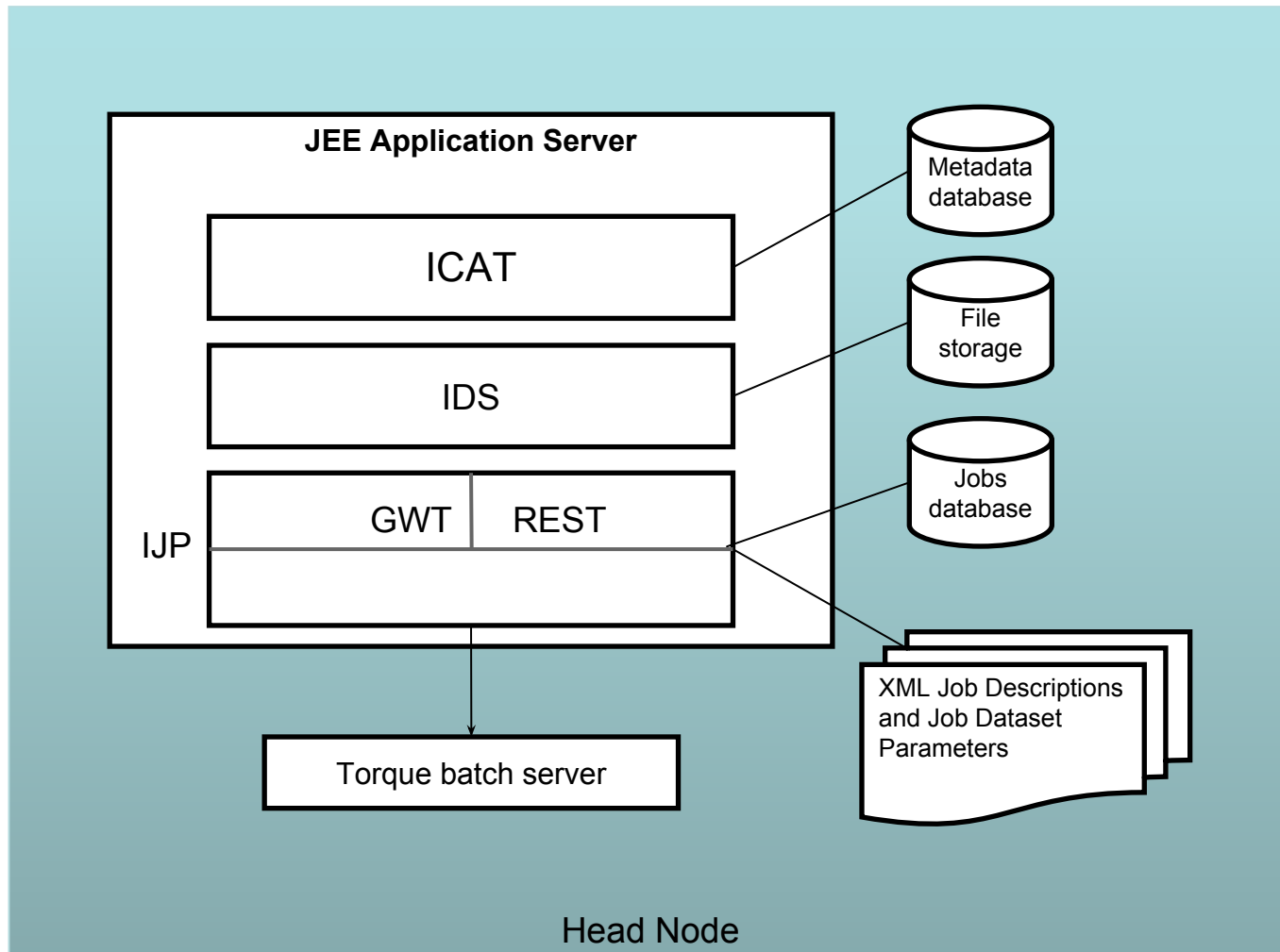
A solution

- Build a batch and interactive job portal on top of ICAT and IDS
- Implement GUI access via Google Web Toolkit
- Provide command line access via RESTful interface
- Use other tried, tested, scalable and preferably open source components

Architecture Overview



Head Node Architecture



Job Portal Main Panel (Datasets)

Firefox

ICAT Job Portal

Datasets Job Status

project

Any user
Unknown instrument
OctopusSM2
OctopusSM3

Any instrument
Unknown instrument
OctopusSM2
OctopusSM3

Any experiment type
Unknown experiment type
Colocalisation
Undefined

Any number of channels
1 channel
2 channels
3 channels

Search

X startDate BETWEEN 2012 Jan 1 12:00:00 2013 Jan 1 12:00:00

X nframes >= 500

7 datasets found.

Name	Description	Users
20120524_0002_0001_632c1ef9-9f32-4a39-a649-855ed5592c27	coloc 3 Affibodys 639 nm laser	
20120525_0004_0001_0bbb36de-dd79-4c13-84ca-72a6a86de334	coloc 3 Affibodys T47D	
20120524_0002_0001_e421cec3-d7eb-4e3f-baea-bf66fed31688	T47D 3 Affibodys 639 nm laser	
20120525_0004_0001_6e28e0b5-fe99-45a4-93d7-61a952a35912	coloc 3 Affibodys T47D	
20120524_0002_0001_c1b3dc55-0f05-4daf-be3f-e935291f812e	T47D 3 Affibodys 639 nm laser	
20120524_0002_0001_da8e9d70-b461-406f-9e06-b32678096d1d	T47D 3 Affibodys 639 nm laser	
20120524_0002_0001_aee07c8e-dc7d-4b6c-a599-6e62eb4f829e	T47D 3 Affibodys 639 nm laser	

Options ...
Options ...
Download
Show Download URL
MSMM Viewer Project

endDate	2012-11-27T14:18:17Z
experiment_type	Undefined
id	7201
instrument	OctopusSM3
location	Dummy Investigation 1/20120524_0002_0001_aee07c8e-dc7d-4b6c-a599-6e62eb4f829e
name	20120524_0002_0001_aee07c8e-dc7d-4b6c-a599-6e62eb4f829e
nchannels	1
nframes	571
sampledescription	T47D 3 Affibodys 639 nm laser
startDate	2012-11-27T14:16:21Z

Job Options

MSMM Viewer Project Options

View type View View beads View whitelights View reg residual frames View reg model frames

Track method

Show variance image instead of image

Do not load traces

Read features/tracks from hdf5 files (slow)

Set min,max for colour scale

Regular expression for images in directory

Do not clean levels/stats (default=0) (min=0) (max=10)

Min number of detected features per frame range of a level/state (default=2)

Threshold for the Chauvenet's outlier test (default=2) (min=1) (max=5)

Set the (real) EM gain by hand

Quantum efficiency (default=0.910000026) (min=-1.0) (max=1.0)

Set the (real) electron/ADU by hand

A unique identifier of the EMCCD (default=Command:Line)

Quit immediately after initialisation completes

Add a string to the view window title

Job Status Panel

Firefox

ICAT Job Portal x

Datasets Job Status

Refresh Job Status Display Job Output Display Job Error

Job ID	Worker Node	Batch Filename	Submitted	Status
81.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	qmybdzrphr.sh	01-03-2013 14:41:54	COMPLETED
78.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	icfhlkvhjf.sh	12-02-2013 13:51:57	COMPLETED
77.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	agefhjfwf.sh	12-02-2013 13:51:51	COMPLETED
76.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	xezhuyccms.sh	12-02-2013 13:40:39	COMPLETED
75.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	fcebrhyxvp.sh	12-02-2013 13:40:29	COMPLETED
74.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	iqlkkvabkk.sh	12-02-2013 10:50:48	COMPLETED
73.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	dnfsmuakvy.sh	12-02-2013 10:48:21	COMPLETED
64.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	ahtpltkhzc.sh	11-02-2013 15:27:14	COMPLETED
65.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	phqzrbcki.sh	11-02-2013 15:27:14	COMPLETED

Multiple Dataset Handling

- Jobs can accept a single or multiple datasets (specified in XML Job Description)
- Multiple datasets can be submitted to a job specified as accepting multiple datasets as input
- A separate batch job can be submitted for each dataset (with a single click)
- With multiple datasets selected, Job Options Form offers only options common to all datasets

Interactive jobs

The screenshot shows a Windows desktop environment. On the left, the desktop icons include Recycle Bin, SM1.txt, SM3.txt, CLF_email.txt, MobaXter..., portal_scre..., InductionC... 2(2).docx, jd-guicfg, GIMP2, Mozilla Firefox, Oracle VM VirtualBox, and TechWriter for Web Se... removal.

The main window is a Firefox browser displaying the ICAT Job Portal. The address bar shows `81/jjp_portal-2.0.0-SNAPSHOT/`. The page content includes a search filter for 'project' and a list of datasets. A table below shows details for a specific dataset:

Name	Sample ID
20130208_0001_0001_25bb2b25-f194-49a0-ac39-8454ed7cf26c	1 nM HEF
20130208_0001_0001_f98868cd-0127-4562-907c-3c3bf8e06e71	1 nM HEF
20130214_0001_0004_d199a7e3-9a8b-46cd-aae7-f8e989593ce1	4 nM EGF
20130211_0001_0001_73e2c807-1a30-4058-9036-8ce8086b110f	1 nM HEF
20130301_0005_0010_e517e655-92fe-43b5-a41a-e6f66849f290	Images for

Below the table, a 'comments' section lists metadata for the selected dataset:

endDate	2013-05-16T13:22:00+00:00
experiment_type	Undefined
id	18714
instrument	OctopusSM4
location	Dummy Investigation 1/
name	20130301_0005_0010_e517e655-92fe-43b5-a41a-e6f66849f290
nchannels	1
nframes	60
sampledescription	Images for ccd spec est
startDate	2013-05-16T13:22:00+00:00

The LSF Remote Desktop Connection window shows a 'FrameViewer' application. The title bar reads '20130301_0005_0010 - Images for ccd spec estimation'. The interface includes a 'Show channels' dropdown (set to 1), a 'Layout' dropdown (set to Vertical), and various menu options like 'Options', 'Load', 'Save', 'Save default', 'Reset', 'Annotation', 'Verbose', 'Font', 'Sans 7', 'Figure', 'Movie', and 'Help'. The main display area shows a dark red microscopy image with numerous white circular spots. A red label '(R) 1: Cy3 (Undefined)' is positioned above the image. Below the image, the text reads 'Image coords (x,y) = (263.15, 231.43) pix. Chid 1 value = 97'. At the bottom of the window, a 'Messages' pane displays the text: 'FrameId 2 (number 1 of 60) - 0.000sec - FrameRole:Sample - proframe_0002.h5 - Unregistered - NOT Bias-subtracted'.

The system tray at the bottom right shows the time as 09:54 and the date as 20/05/2013.

Configuration

1. Create XML files for each dataset type picking out dataset features relevant to Job Options
2. Create XML Job Descriptions
3. Write applications (or wrappers around existing applications) – loading and saving datasets from IDS and recording provenance in ICAT

```
<jobType>
  <name>MSMM Viewer Project</name>
  <executable>/usr/local/mamm/bin/run_mamm_viewer</executable>
  <multiple>>false</multiple>
  <type>interactive</type>
  <datasetTypes>project</datasetTypes>
  <jobOptions>
    <name>View</name>
    <groupName>View type</groupName>
    <type>boolean</type>
    <programParameter></programParameter>
    <condition></condition>
  </jobOptions>
  <jobOptions>
    <name>View reg beads</name>
    <groupName>View type</groupName>
    <type>boolean</type>
    <programParameter>--reg-beads</programParameter>
    <condition>numBeadFiles>0 && numChannels>1</condition>
  </jobOptions>
  <jobOptions>
    <name>Track method</name>
    <type>enumeration</type>
    <programParameter>--trackmethod</programParameter>
    <values></values>
    <values>Simple</values>
    <values>SLH</values>
    <values>Biggles</values>
    <values>Simulation</values>
  </jobOptions>
  <jobOptions>
    <name>Regular expression for images in directory</name>
    <type>string</type>
    <programParameter>--image-pattern</programParameter>
  </jobOptions>
  <jobOptions>
    <name>Do not clean levels/stats</name>
    <type>integer</type>
    <programParameter>--Levels.no-clean</programParameter>
    <defaultValue>0</defaultValue>
  </jobOptions>
</jobType>
```

Job Options from XML

XML Job Description on Head Node

```
<jobType>
  <name>MSMM Viewer Project</name>
  <executable>/usr/local/msmm/bin/run_msmm_viewer</executable>
  <multiple>false</multiple>
  <type>interactive</type>
  <datasetTypes>project</datasetTypes>
  <jobOptions>
    <name>View</name>
    <groupName>View type</groupName>
    <type>boolean</type>
    <programParameter></programParameter>
    <condition></condition>
  </jobOptions>
  <jobOptions>
    <name>View reg beads</name>
    <groupName>View type</groupName>
    <type>boolean</type>
    <programParameter>--reg-beads</programParameter>
    <condition>numBeadFiles>0 && numChannels>1</condition>
  </jobOptions>
  <jobOptions>
    <name>Track method</name>
    <type>enumeration</type>
    <programParameter>--trackmethod</programParameter>
    <values></values>
    <values>Simple</values>
    <values>SLH</values>
    <values>Biggles</values>
    <values>Simulation</values>
  </jobOptions>
  <jobOptions>
    <name>Regular expression for images in directory</name>
    <type>string</type>
    <programParameter>--image-pattern</programParameter>
  </jobOptions>
  <jobOptions>
    <name>Do not clean levels/stats</name>
    <type>integer</type>
    <programParameter>--Levels.no-clean</programParameter>
    <defaultValue>0</defaultValue>
  </jobOptions>
</jobType>
```

Job Options Form in Web Browser

MSMM Viewer Project Options

View type View View beads View whitelights View reg residual frames View reg model frames

Track method

Show variance image instead of image

Do not load traces

Read features/tracks from hdf5 files (slow)

Set min,max for colour scale

Regular expression for images in directory

Do not clean levels/stats (default=0) (min=0) (max=10)

Min number of detected features per frame range of a level/state (default=2)

Threshold for the Chauvenet's outlier test (default=2) (min=1) (max=5)

Set the (real) EM gain by hand

Quantum efficiency (default=0.910000026) (min=-1.0) (max=1.0)

Set the (real) electron/ADU by hand

A unique identifier of the EMCCD (default=Command:Line)

Quit immediately after initialisation completes

Add a string to the view window title

Submit Close

Command Line Interface

- RESTful web service and Python client for job handling
- Alternative to using web browser
- May become preferred interface for some users
- Enables scripted interaction with IJP

```
$> ijp login db username fred password -  
password:  
d3f58cf7-23fb-4e0a-89bc-292dcc142e20
```

```
$> ijp session  
User ingest connected to ICAT 4.2.5 at https:  
//rclsfserv010.rc-harwell.ac.uk:8181 with 1439  
minutes left.
```

```
$> ijp jobtype  
Available job types are:  
view_ingested is interactive  
ingest is batch  
view_project is interactive  
quincy is batch
```

```
$> ijp submit ingest gggg  
2.rclsfserv010.rc-harwell.ac.uk
```

```
$> ijp status  
2.rclsfserv010.rc-harwell.ac.uk, R
```

Status

- System has been implemented deployed and given to LSF for feedback
 - The system has the desired functionality and is responsive
 - Short informal weekly meetings between the developers and LSF have ensured the delivery of the desired product
- Other STFC facilities and groups are interested

Future Developments

- Improvements following user feedback
- Visualisation of Provenance
- Workflow Support
- Administration console
- Alternative remote desktop mechanism
- Alternative batch systems
- Portability