Science & Technology Facilities Council

ICAT Job Portal

ICAT Workshop in Lund, Sweden
12/13 March 2013

Kevin Phipps and Steve Fisher
Scientific Computing Department, RAL

Introduction

- Started as project for Lasers for Science Facility (LSF) at RAL
- Project running for about 1.5 years
- Funded by Harwell Imaging Partnership

Science & Technology Facilities Council

LSF team

- Small team of scientists
- Develop and support analysis software
- Limited IT support
- ~10 analysis nodes with GPUs (Linux)
- No cataloguing of datasets – finding datasets takes longer as more datasets added

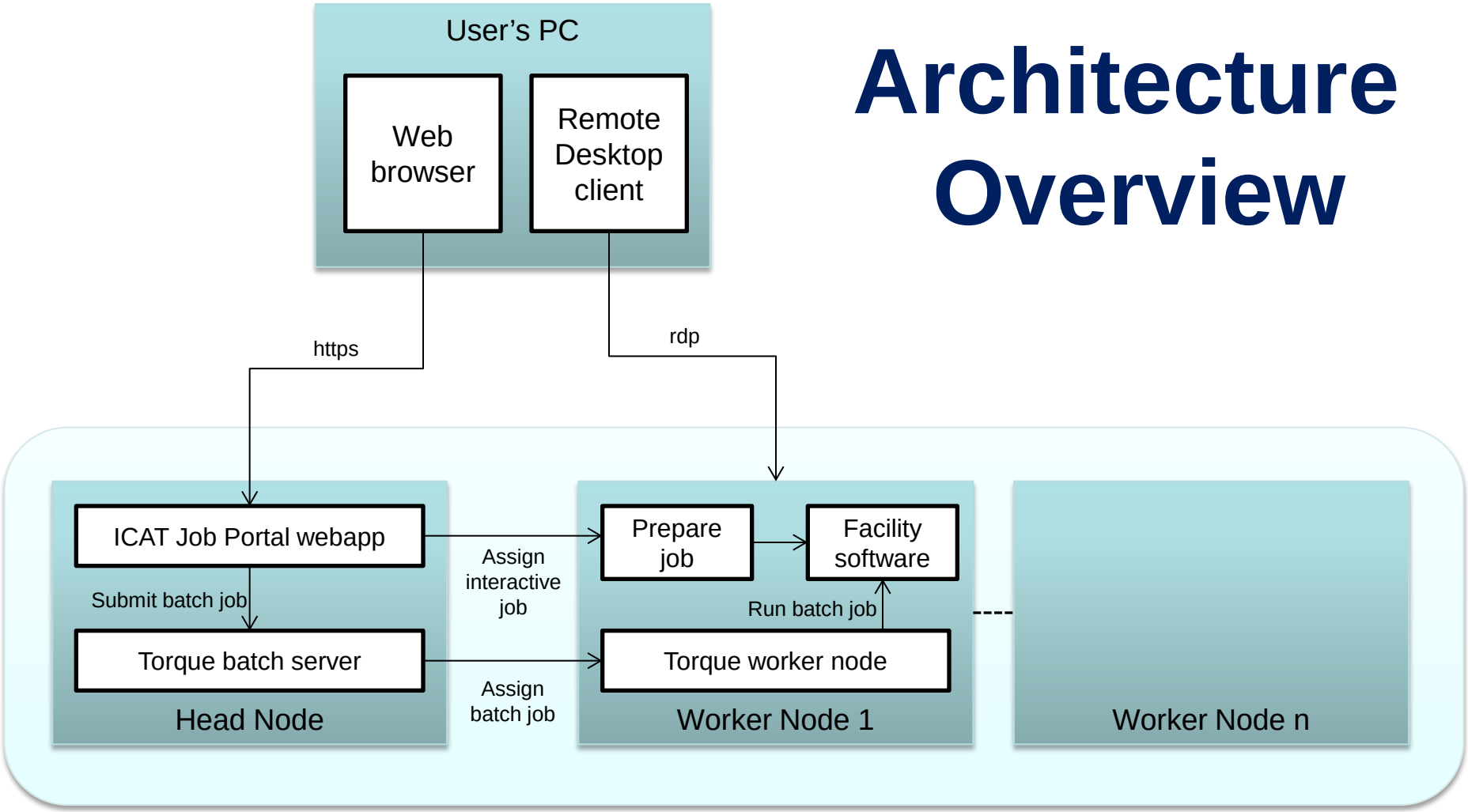
The Aim

- Build a batch and interactive job portal
- Use tried, tested, scalable and preferably open source components
- Hide away the underlying complexity from the end user

Project Status

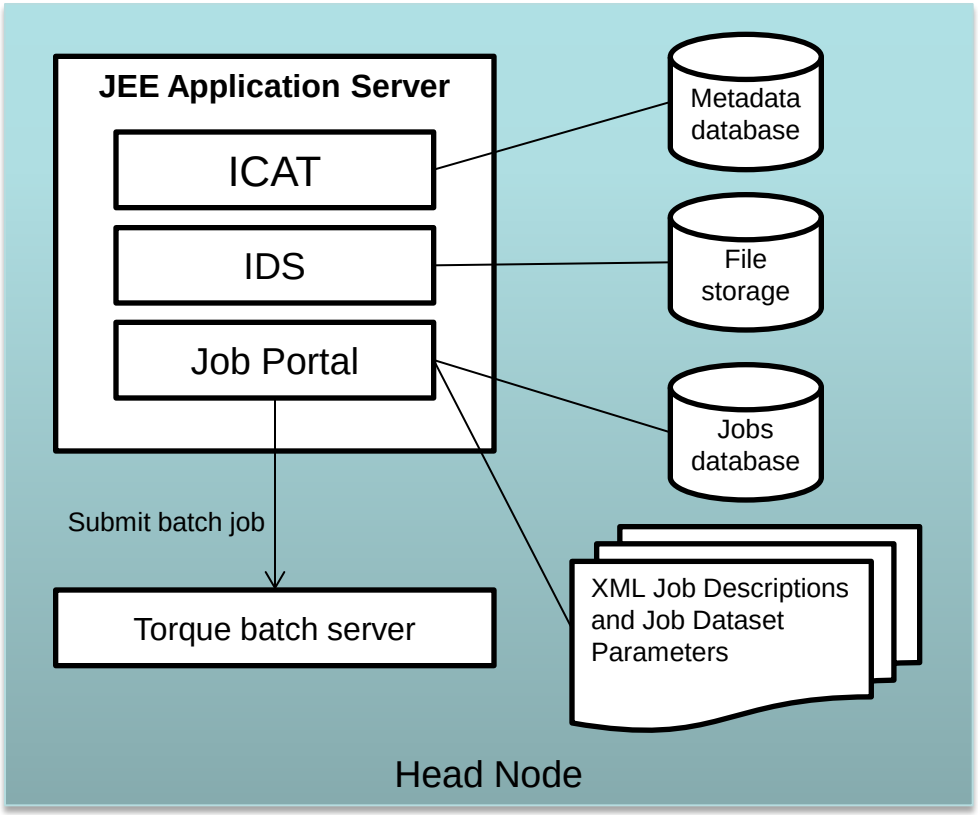
- Prototype demonstrated mid-2012
- Work since then mainly to make it more generic, configurable and installable
- Usable system deployed for LSF within next few months

Architecture Overview



Science & Technology Facilities Council

Head Node Architecture



Science & Technology Facilities Council

Installation and Configuration

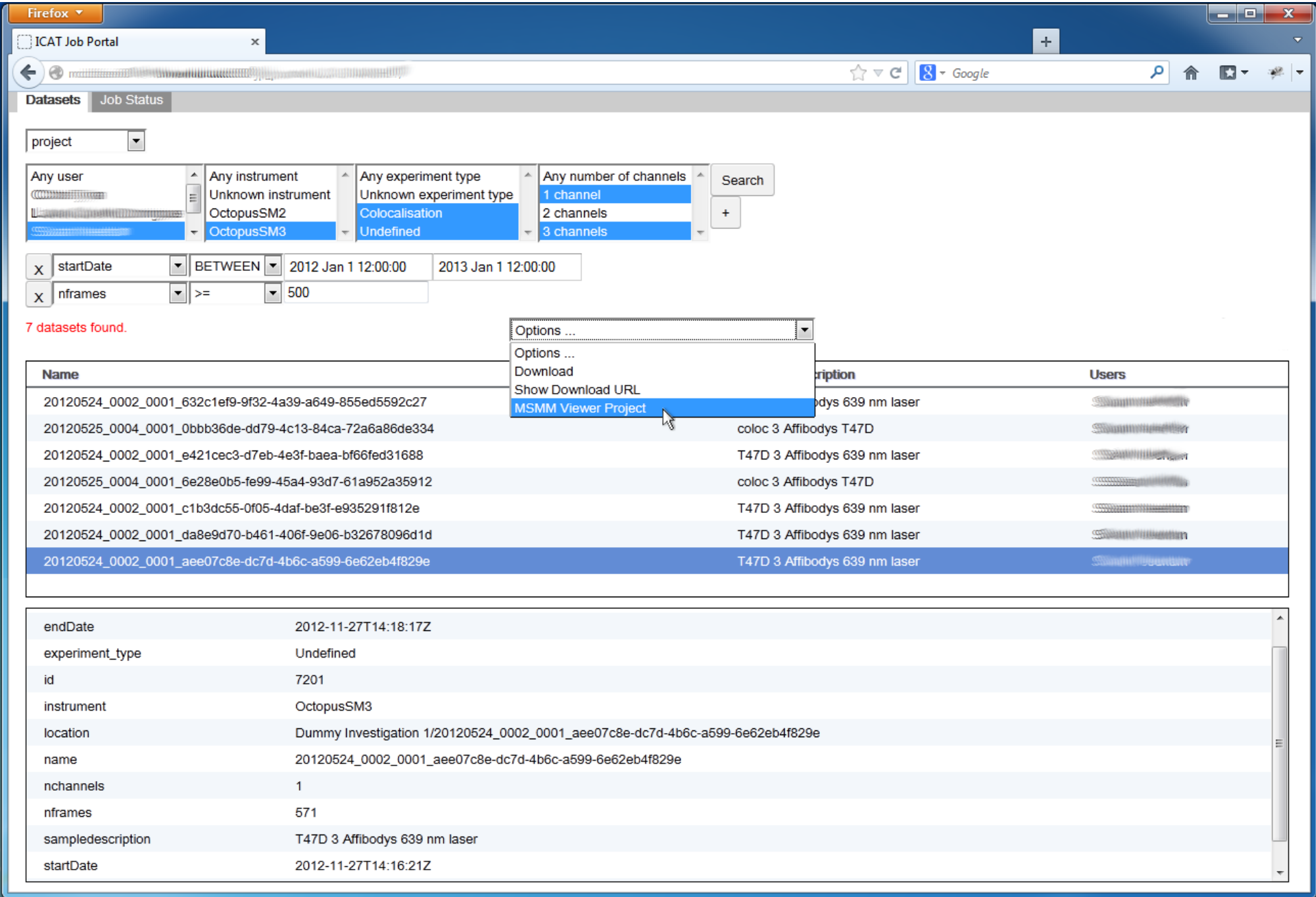
- Puppet Automation Software used to install head node and worker nodes
- Start with a clean Ubuntu 12.04 LTS installation and a network connection on all machines
- Installation takes an hour or two and includes:
  Java Development Kit, Glassfish Application Server including ICAT, IDS and Job Portal Software, MySQL database server and required databases, Torque batch system, Ganglia monitoring and the facility specific software
- Facility software updates rolled out automatically to all worker nodes by Puppet

‘Admin’ User Tasks

- Create XML Job Descriptions
- Create XML files for each dataset type picking out dataset features relevant to Job Options
- Write wrapper scripts for each application – saving and loading datasets from IDS and recording provenance
- A Python library of utilities is being created to help with the creation of wrapper scripts

Science & Technology Facilities Council

Job Portal Main Panel (Datasets)



Job Options

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>
    <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>
    <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>
    <jobOptions>
        <name>Regular expression for images in directory</name>
        <type>string</type>
        <programParameter>--image-pattern</programParameter>
    </jobOptions>
    <jobOptions>
        <name>Regular expression for images in directory</name>
        <type>string</type>
        <programParameter>--image-pattern</programParameter>
    </jobOptions>
```

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 [ ]

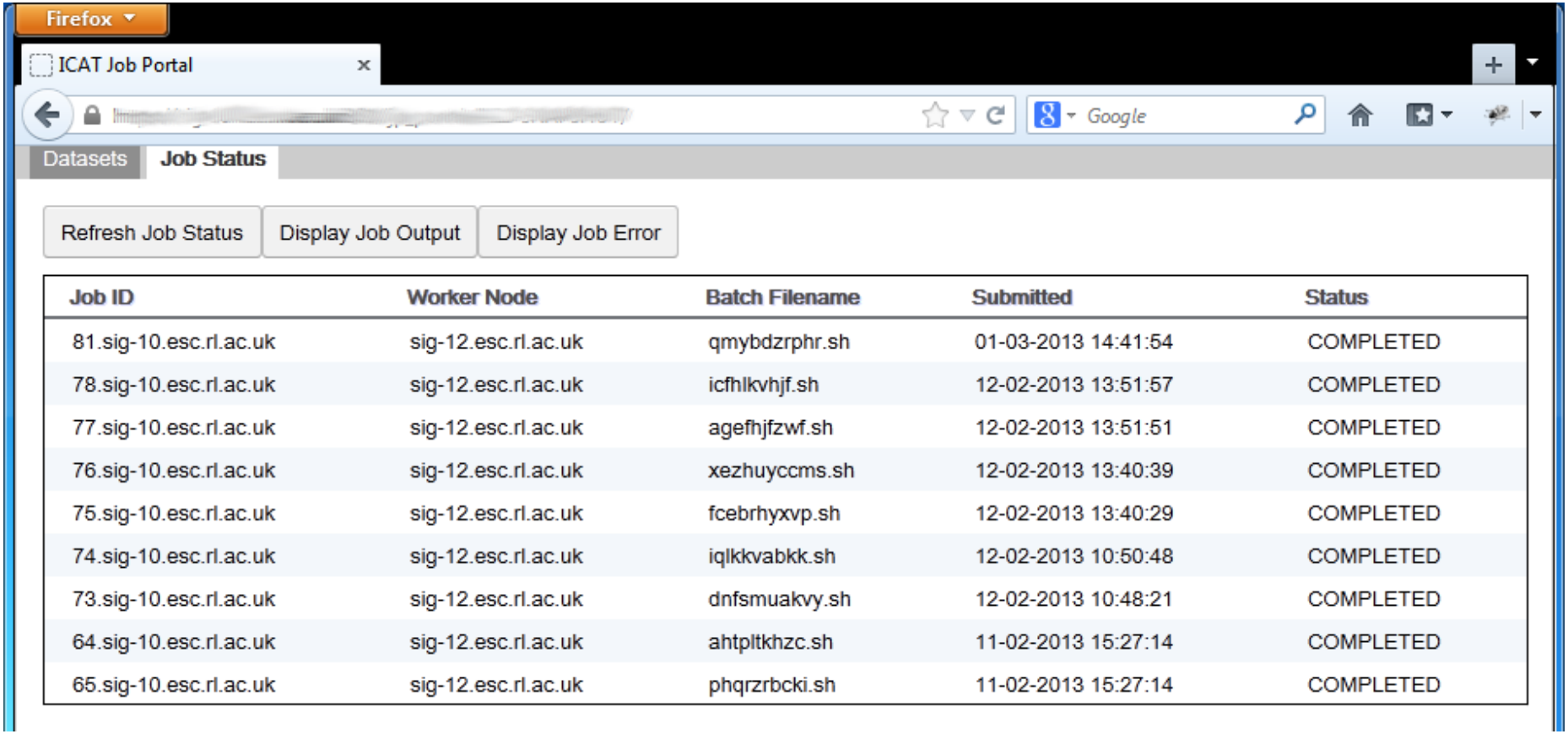
A unique identifier of the EMCCD [ ] *(default=Command:Line)*

Quit immediately after initialisation completes [ ]

Add a string to the view window title [ ]

Submit Close

Job Status Panel



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	agefhjfzwf.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	phqrzrbcki.sh	11-02-2013 15:27:14	COMPLETED

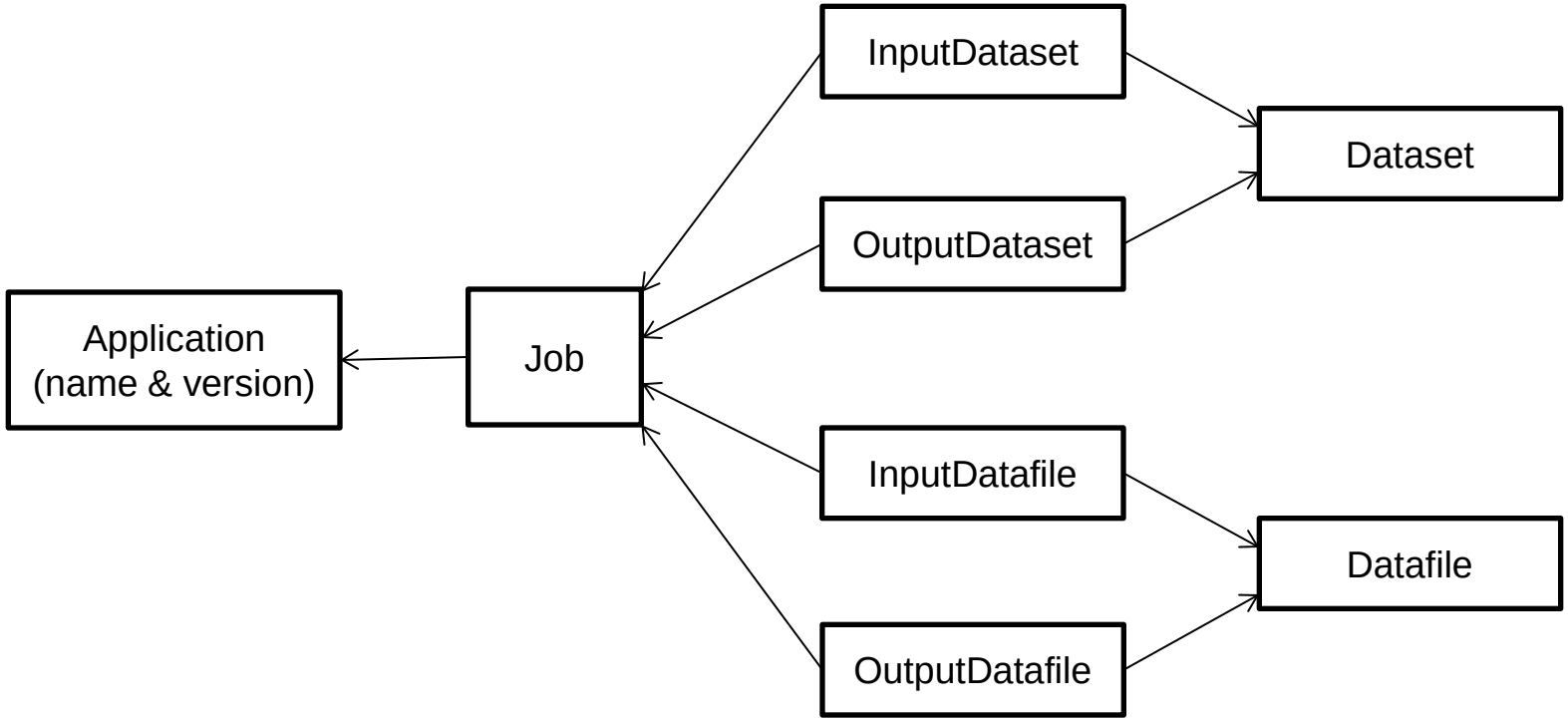
Multiple Dataset Handling

- Jobs can accept a single or multiple datasets (specified in XML Job Description)
- Multiple datasets can be selected in the Portal
- 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

Command Line Interface

- RESTful web service and Python client added for job handling
- Alternative to using web browser
- May become preferred interface for more proficient users
- Enables scripted interaction with Job Portal

Provenance Support in ICAT



Science & Technology Facilities Council

Provenance in Job Portal

- Job Portal will include a tool to visualise dataset provenance
- When a new dataset is added to ICAT the wrapper script must register the provenance information
- Relevant to PanData WP 6 – Data Provenance

Science & Technology Facilities Council

Future Developments

- Improvements following user feedback
- LSF secured funding for 30 new dedicated nodes on the EMERALD GPU cluster at RAL
- Port installation to different OS/any OS?
- Make it batch system independent
- Find new groups of users for it