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

User's PC
- Web browser
- CLI
- Remote Desktop client

Head Node
- ICAT Job Portal webapp
- Torque batch server
  - Submit batch job

Worker Node 1
- Prepare job
- Facility software
- Torque worker node
  - Assign interactive job
  - Assign batch job
  - Run batch job

Worker Node n
Head Node Architecture

- JEE Application Server
  - ICAT
  - IDS
  - GWT
  - REST

- Torque batch server

- Metadata database
- File storage
- Jobs database
- XML Job Descriptions and Job Dataset Parameters
### Job Portal Main Panel (Datasets)

#### Datasets

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>20120524_0002_0001_502c1ef9-9f32-4a39-a649-855e65592c27</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
<tr>
<td>20120526_0004_0001_0b306de-9179-4c13-84ca-72a6a85de334</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
<tr>
<td>20120526_0002_0001_e421cc3-75ec-4aed-9fba-865pf631688</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
<tr>
<td>20120526_0004_0001_6e28f065-f699-46a4-93d7-61a62a36912</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
<tr>
<td>20120524_0002_0001_c1be35c5-0f06-4caf-83f6-a935291f1e2</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
<tr>
<td>20120524_0002_0001_da8e9d70-0bde-40bf-9e6e-b328670896d1</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
<tr>
<td>20120526_0002_0001_aee07c8e-dc7d-4b6c-a699-6e62ebf929e</td>
<td>T47D 3 Affibods 639 nm laser</td>
<td></td>
</tr>
</tbody>
</table>

#### End Date
- 2012-11-27T14:18:17Z

#### Experiment Type
- Undefined

#### ID
- 7201

#### Instrument
- OctopusSM3

#### Location
- Dummy investigation 1/20120524_0002_0001_aee07c8e-dc7d-4b6c-a699-6e62ebf929e

#### Name
- 20120524_0002_0001_aee07c8e-dc7d-4b6c-a699-6e62ebf929e

#### Channels
- 1

#### Frames
- 571

#### Sample Description
- T47D 3 Affibods 639 nm laser

#### Start Date
- 2012-11-27T14:16:21Z
Job Options

MSMM Viewer Project Options

View type
- View
- View beads
- View white lights
- View reg residual frames
- View reg model frames

Track method

Show variance image instead of image

Do not load traces

Read features/track 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=1.0)

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=0.010000026) (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 window title

Submit

Close
### Job Status Panel

<table>
<thead>
<tr>
<th>Job ID</th>
<th>Worker Node</th>
<th>Batch Filename</th>
<th>Submitted</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>81.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>qmybdzrphr.sh</td>
<td>01-03-2013 14:41:54</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>78.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>icfhkvhjhf.sh</td>
<td>12-02-2013 13:51:57</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>77.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>agefjfhjzfwf.sh</td>
<td>12-02-2013 13:51:51</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>76.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>xezhuyccms.sh</td>
<td>12-02-2013 13:40:39</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>75.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>fcebrhyxvp.sh</td>
<td>12-02-2013 13:40:29</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>74.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>ilqkkvabkk.sh</td>
<td>12-02-2013 10:50:48</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>73.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>dnfsmuakvy.sh</td>
<td>12-02-2013 10:48:21</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>64.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>ahtpltkhzc.sh</td>
<td>11-02-2013 16:27:14</td>
<td>COMPLETED</td>
</tr>
<tr>
<td>65.sig-10.esc.rl.ac.uk</td>
<td>sig-12.esc.rl.ac.uk</td>
<td>phqrzrbcki.sh</td>
<td>11-02-2013 16:27:14</td>
<td>COMPLETED</td>
</tr>
</tbody>
</table>
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
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
Job Options from XML

XML Job Description on Head Node

```xml
<jobType>
  <name>NSM Viewer Project</name>
  <executable>/usr/local/summ/bin/run_summ_viewer</executable>
  <multiple>false</multiple>
  <type>interactive</type>
  <datasetTypes>project</datasetTypes>
  <jobOptions>
    <name>View</name>
    <groupName>View</groupName>
    <type>boolean</type>
    <programParameter>--programParameter
    <condition></condition>
  </jobOptions>
</jobType>

<jobOptions>
  <name>View reg beads</name>
  <groupName>View</groupName>
  <type>boolean</type>
  <programParameter>--programParameter
  <condition>numBeadFiles>0 && numChannels>1</condition>
</jobOptions>

<jobOptions>
  <name>Track method</name>
  <type>enumeration</type>
  <programParameter>--trackmethod
  <values></values>
  <values>Simple</values>
  <values>SLH</values>
  <values>Biglee</values>
  <values>Simulation</values>
</jobOptions>

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

<jobOptions>
  <name>Do not clean levels/stats</name>
  <type>integer</type>
  <programParameter>--levels_no-clean
  <defaultValue>0</defaultValue>
</jobOptions>
```

Job Options Form in Web Browser

MSMM Viewer Project Options

- View type: View, View beads, View white lights, View reg residual frames, View reg model frames
- Track method: dropdown
- 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

Min number of detected features per frame range of a level/state: (default=2) (min=0) (max=10)

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