



ICAT Job Portal

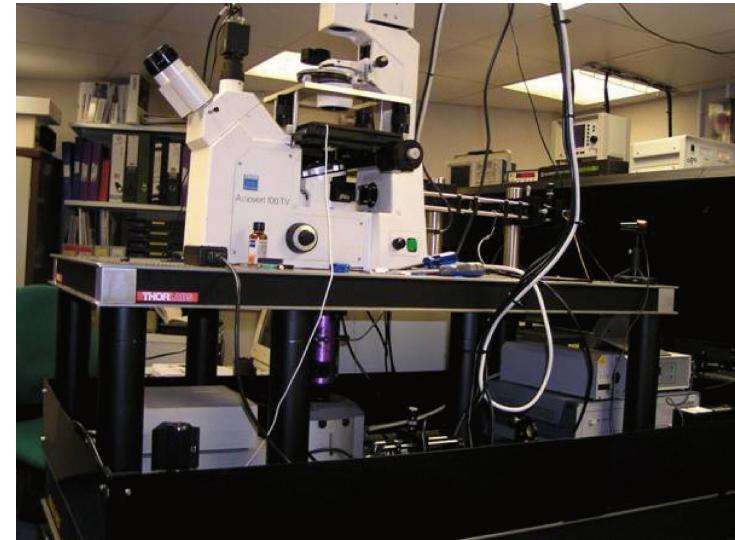
**a generic job submission system
built on a scientific data catalog**

IWSG 2013 ETH, Zurich, Switzerland
3-5 June 2013

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



Problems

- Not scalable
 - not backed by a database - GUI had to scan **all** data each time started
 - user needed a machine allocated to him with a personal account - labour intensive

Some Requirements

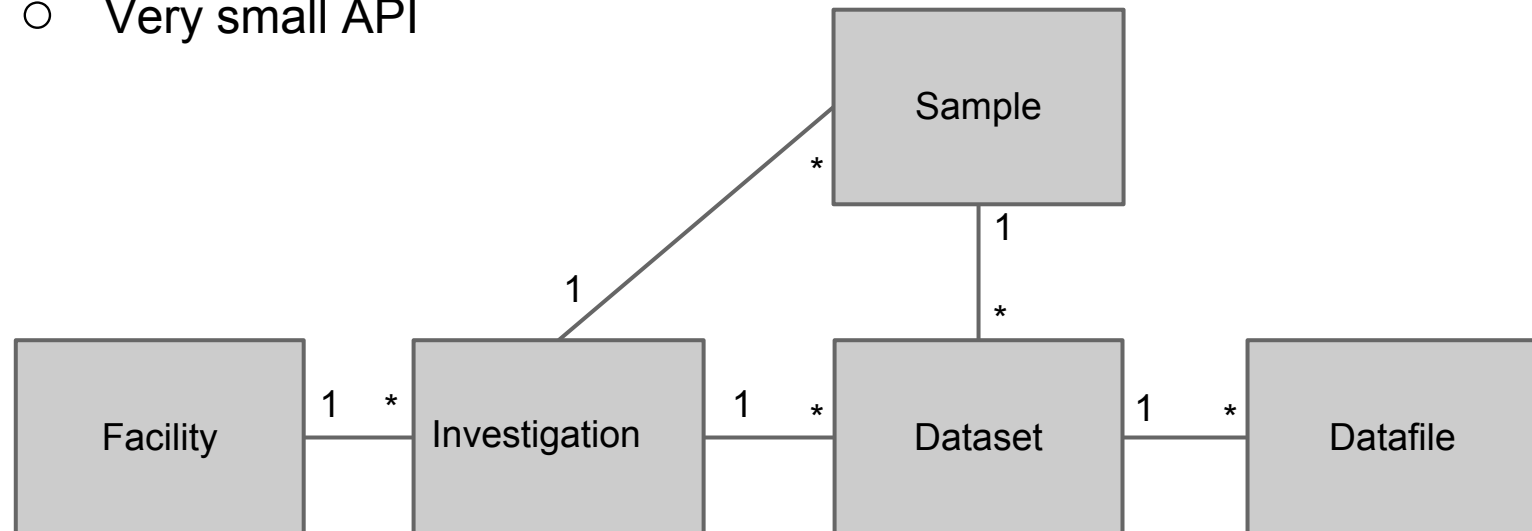
1. GUI and command line from on and off site.
2. Automated installation of OS and software updates.
3. Centralised user/group management.
4. A file server to store raw data, analysed data and other user data.
5. Consult the metadata to locate the data.
6. Upload and download data.
7. Submit batch jobs to Linux nodes; listing, cancelling and retrieving output.
8. Interactive GUI based analysis/visualisation jobs able to access data.
9. Select and submit multiple datasets for processing through applications.
10. No facility dependence: configurable menus, datasets types, jobs and associated job parameters.

A solution

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

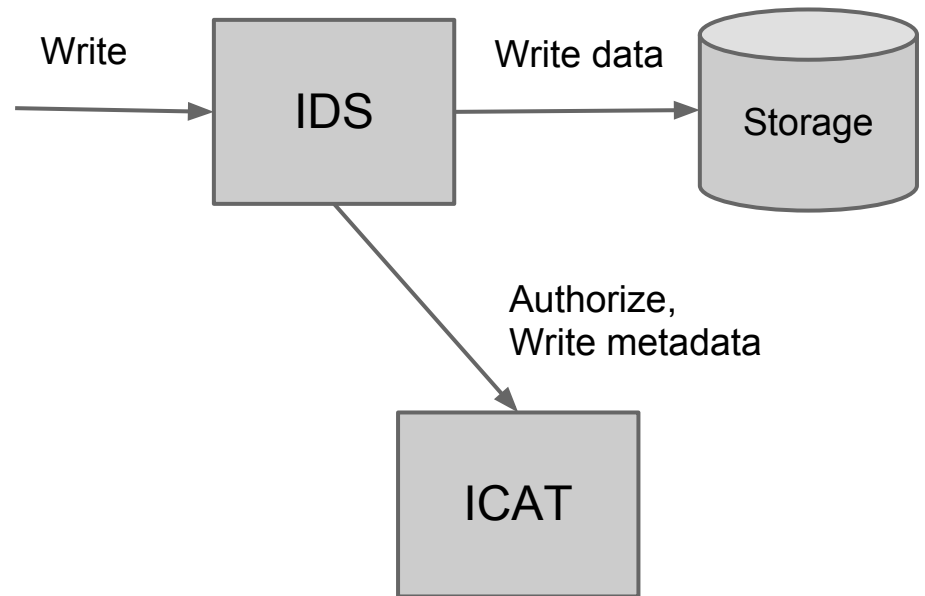
ICAT - metadata catalog

- A database designed to support facility data and exposed as a web service
- Uses any eclipselink supported DBMS
 - Tested with Oracle, MySQL and Derby
- Deployed on an application server
 - Such as Glassfish
- Exposed as a web service
 - Very small API



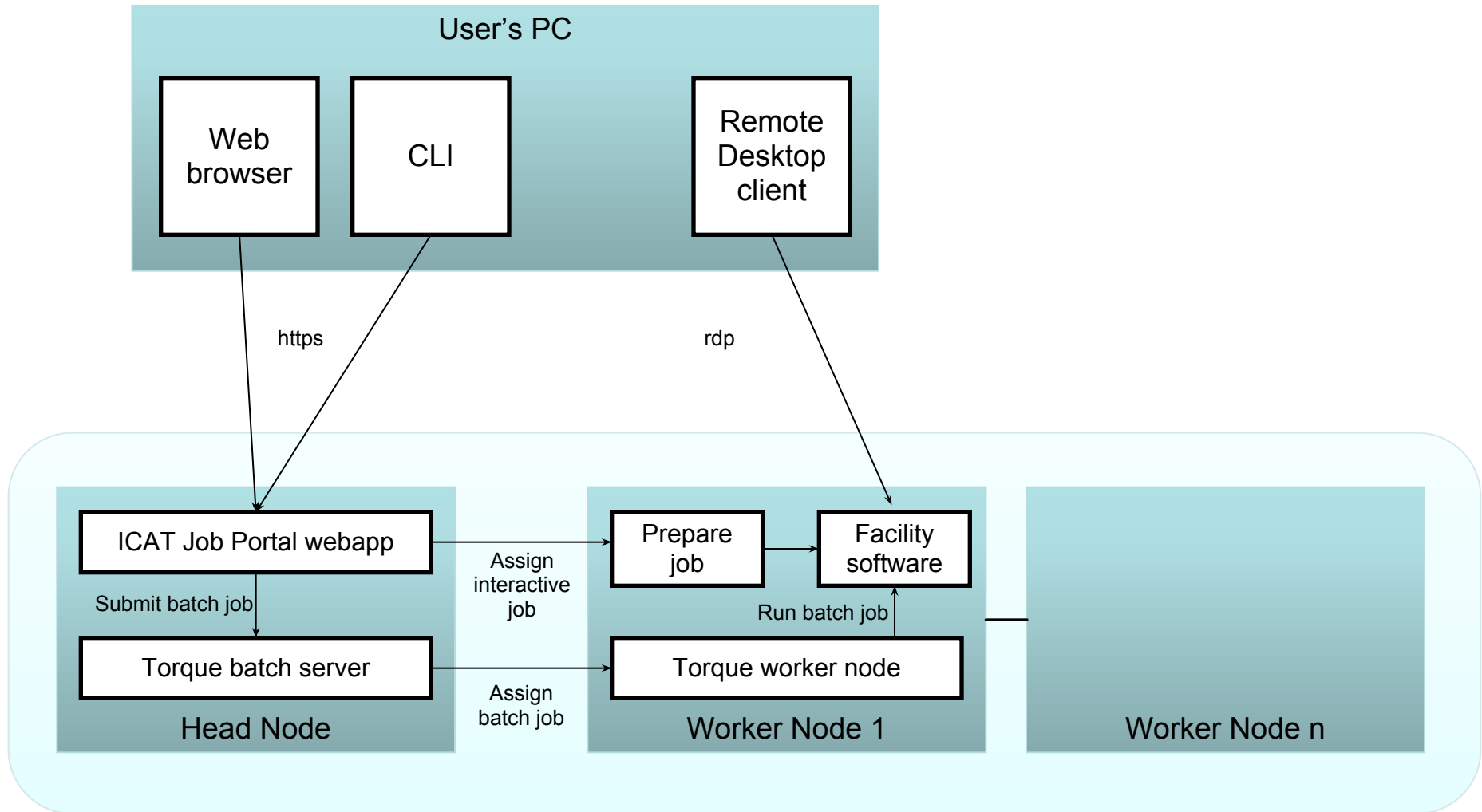
ICAT Data Service (IDS)

This is defined by an interface (RESTful web service) which is able to store files and register their metadata in ICAT. Various implementations depending on needs of facility.

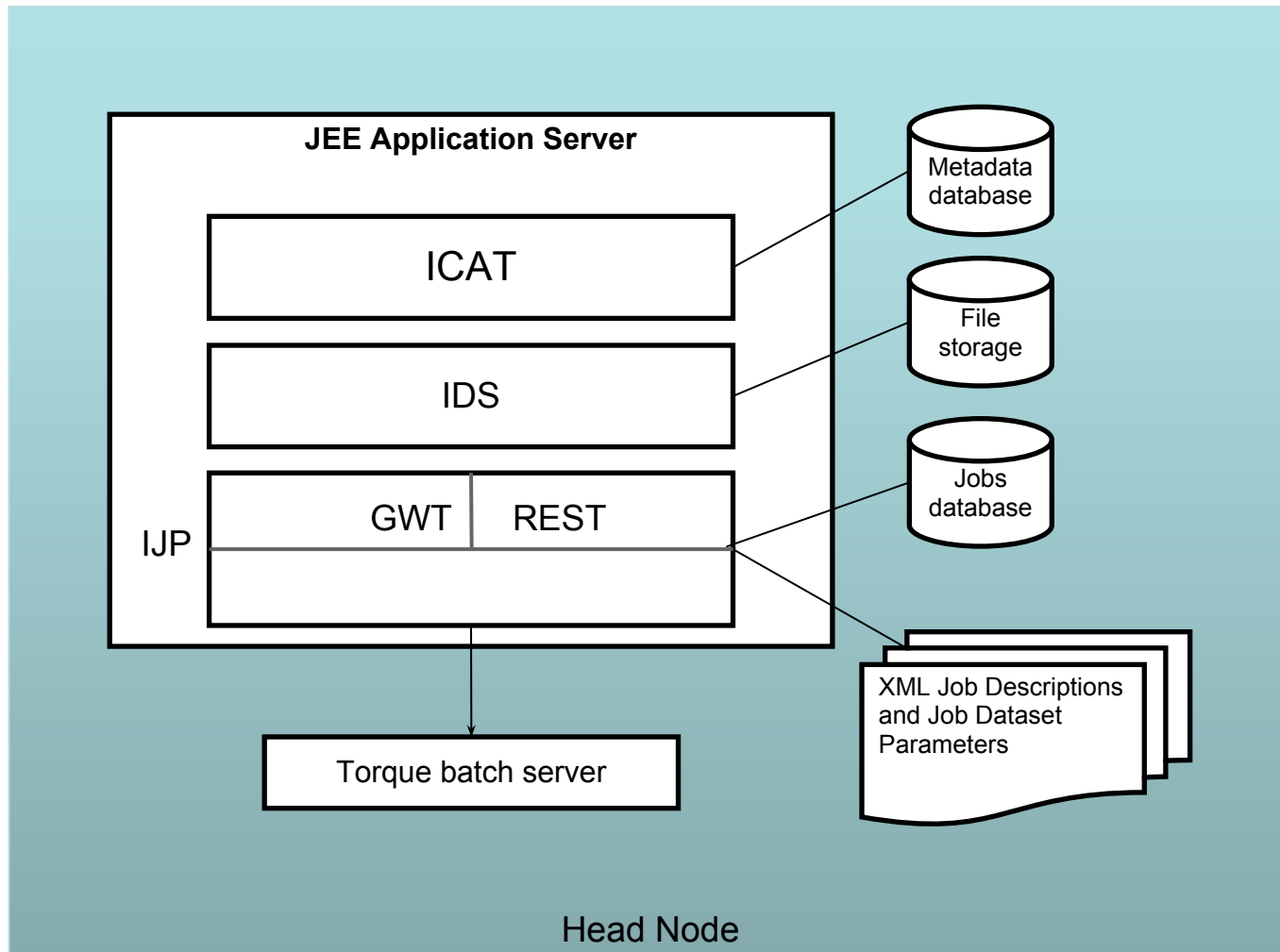


ICAT metadata authorization => IDS data authorization

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

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

nframes >= 500

7 datasets found.

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

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	

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

← 🔒 Google 🔍 🏠 ★ ▾ 🌐 ▾

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	xezhuycms.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	iqkkvabkk.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 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

Recycle Bin

Firefox

ICAT Job Portal

81/jjp_portal-2.0.0-SNAPSHOT/

project

Any user

Any instrument

Any instrument

Unknown instrument

OctopusSM1

OctopusSM2

Unknown experiment type

Colocalisation

Undefined

5 datasets found.

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

comments

endDate

2013-05-16T13:22:00+0

experiment_type

Undefined

id

18714

instrument

OctopusSM4

location

Dummy Investigation 1/

name

20130301_0005_0010_e

nchannels

1

nframes

60

sampledescription

Images for ccd spec est

startDate

2013-05-16T13:22:00+0

LSF_remote_session-23 - sig-12.escri.ac.uk - Remote Desktop Connection

20130301_0005_0010 - Images for ccd spec estimation

Show channels ☒ Layout: Vertical Options: Load Save Save default Reset Annotation: Verbose Font: Sans 7 Figure Movie Help

FrameViewer

Zoom in Clear Zoom out Image options Trail: 0 Compact Show: All Reg error

(R) 1: Cy3 (Undefined)

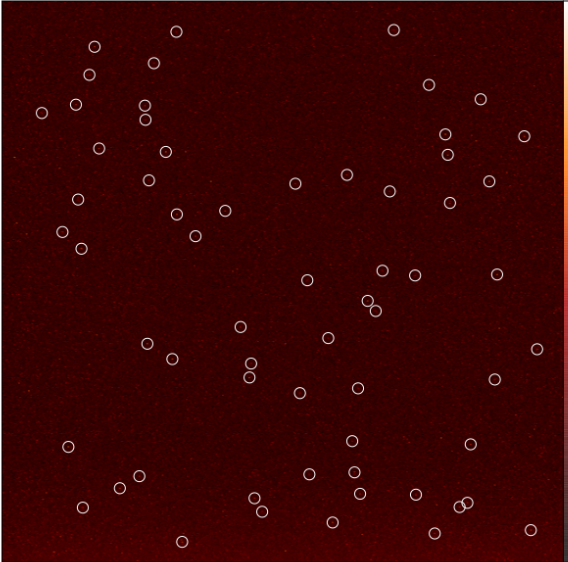


Image coords (x,y) = (263.15, 231.43) pix. Chid 1 value = 97

Messages

Frameid 2 (number 1 of 60) - 0.000sec - FrameRole:Sample - procframe_0002.h5 - Unregistered - NOT Bias-subtracted

job status panel

Show dataset info

Chris Tynan

Chris Tynan

Chris Tynan

Chris Tynan

nan

09:54

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

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 as described has been implemented deployed and given to LSF for feedback
 - A number of known deficiencies
 - Documentation very incomplete
 - Apache License Version 2
 - If you want to try it, suggest that you contact us first (dr.s.m.fisher@gmail.com)

Future Developments

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

and finally ...

- The system has the desired functionality and is responsive
- Other STFC facilities and groups are interested in using it
- Going from prototype to a generic, maintainable system took a long time
- Short informal weekly meetings between the developers and LSF have ensured the delivery of the desired product