The growth of the ICAT family

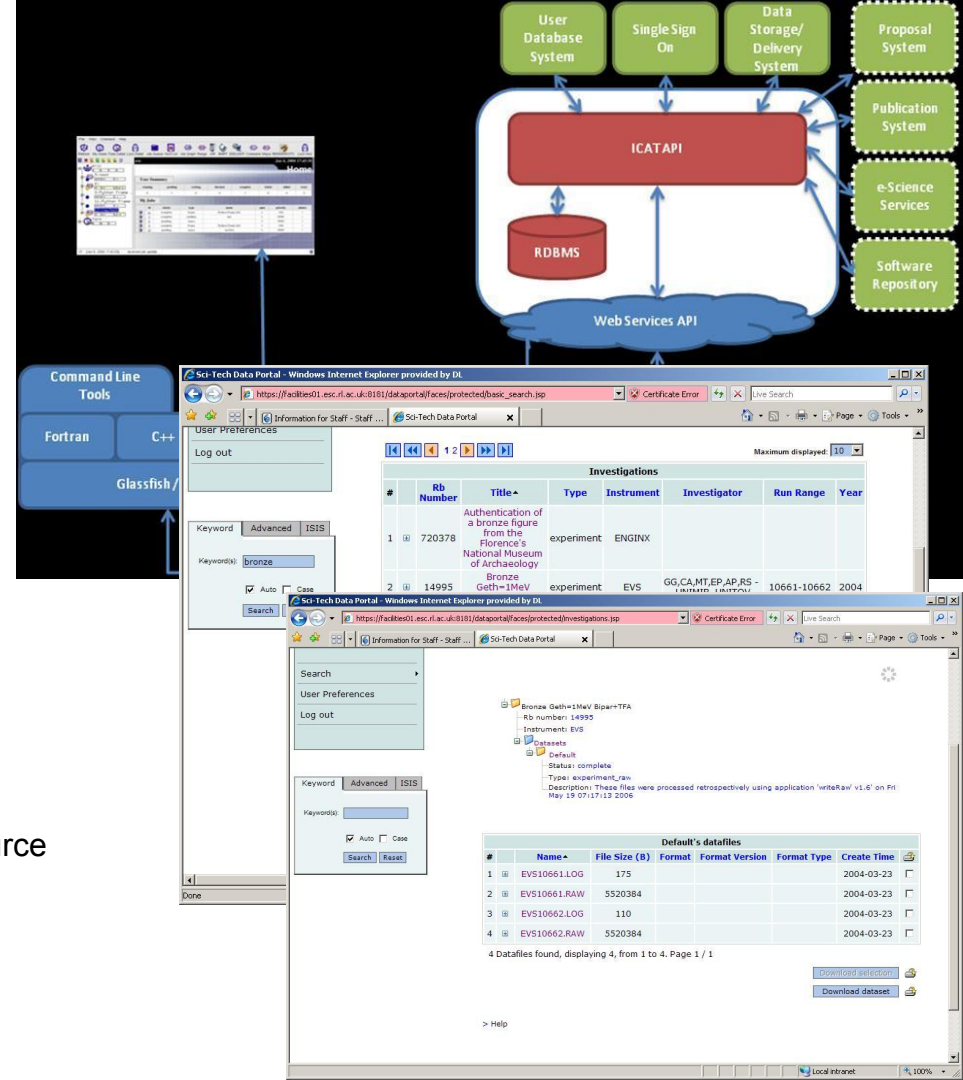
Frazer Barnsley, Wayne Chung, Sylvie Da Graca Ramos, Alex De Maria, Rebecca Fair, Steve Fisher <dr.s.m.fisher@gmail.com>, Andy Gotz, Tom Griffin, Rolf Krahl, Brian Matthews, Peter Parker, Kevin Phipps, Alex Potter-Dixon, Milan Prica, Chris Prosser, Jianguo Rao, Shelly Ren, Brian Ritchie and Jody Salt

“Ancient” History

… provide a metadata catalogue and related components to support large-facility experimental data, ... from proposal through to publication.

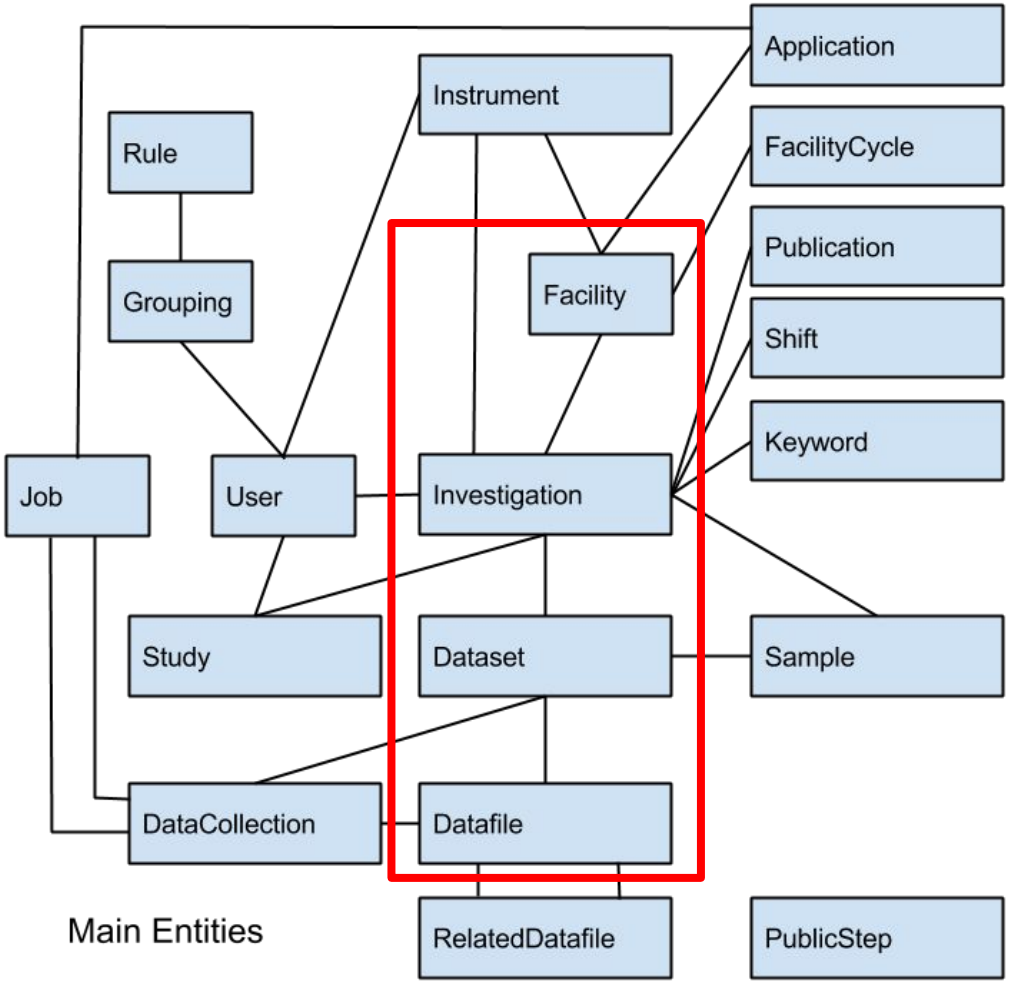
Model - Catalogue - GUI

2001: CLRC Data Portal
2008: ICAT server with SOAP and Glassfish became open source



The model

- Diagram only shows high level view.
- Many to many on diagram relationships are actually implemented by two many to one relationships



The schema

- Entities are identified by an object in the many to one direction and one or more naming fields.
  - For example a Datafile is identified by its Dataset and a name. This also means that a Datafile cannot exist without a Dataset and that it can only be ‘part of’ one Dataset.

Datafile A data file

Uniqueness constraint: dataset, name

From auto generated documentation

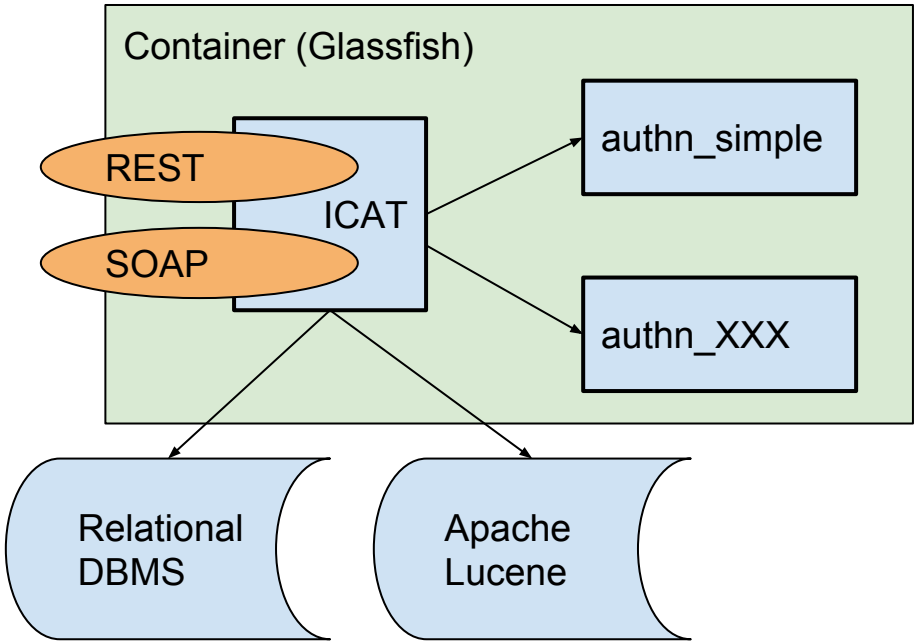
Card	Class	Field	Description
0,*	DataCollectionDatafile	dataCollectionDatafiles	
0,*	RelatedDatafile	sourceDatafiles	
1,1	Dataset	dataset	The dataset which holds this file
0,*	RelatedDatafile	destDatafiles	
0,*	DatafileParameter	parameters	
0,1	DatafileFormat	datafileFormat	

Field	Type	Description
name	String [255] NOT NULL	A name given to the file
fileSize	Long	Expressed in bytes
location	String [255]	The logical location of the file - which may also be the physical location
checksum	String [255]	Checksum of file represented as a string
doi	String [255]	The Digital Object Identifier associated with this data file
description	String [255]	A full description of the file contents

ICAT Server

- Java EE application inside container
- REST and SOAP interfaces
- Pluggable authenticators
- RDBMS and Lucene
- Rule based authorization

- Generic calls to:
  - Write
  - Update
  - Search
  - Delete

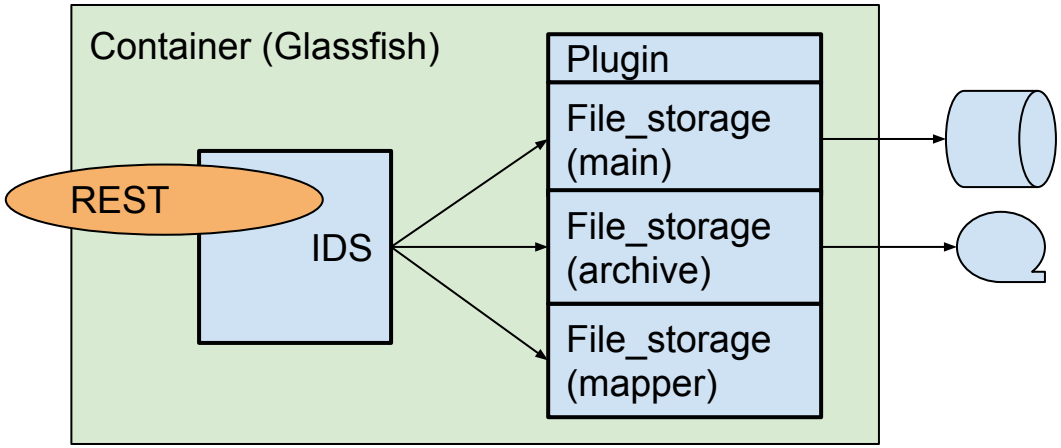


Rule Based Authorization

- Rules to implement a policy
- Such as:
  - All experimental data are public after n days
  - All Investigation records are public
  - Those users related to an investigation can read all Datasets and their related Datafiles and Parameters.
- JPQL SELECT statements define a View.
- Can also define permissions for those in a “group”

IDS Server

- ICAT Data Server for storing and retrieving files
- Can use two level storage if not practical to keep all data on low latency storage
  - IDS manages movement between main and archive storage. Calls to read data not in main storage triggers restore and returns failure.
  - Can use explicit archive and restore calls

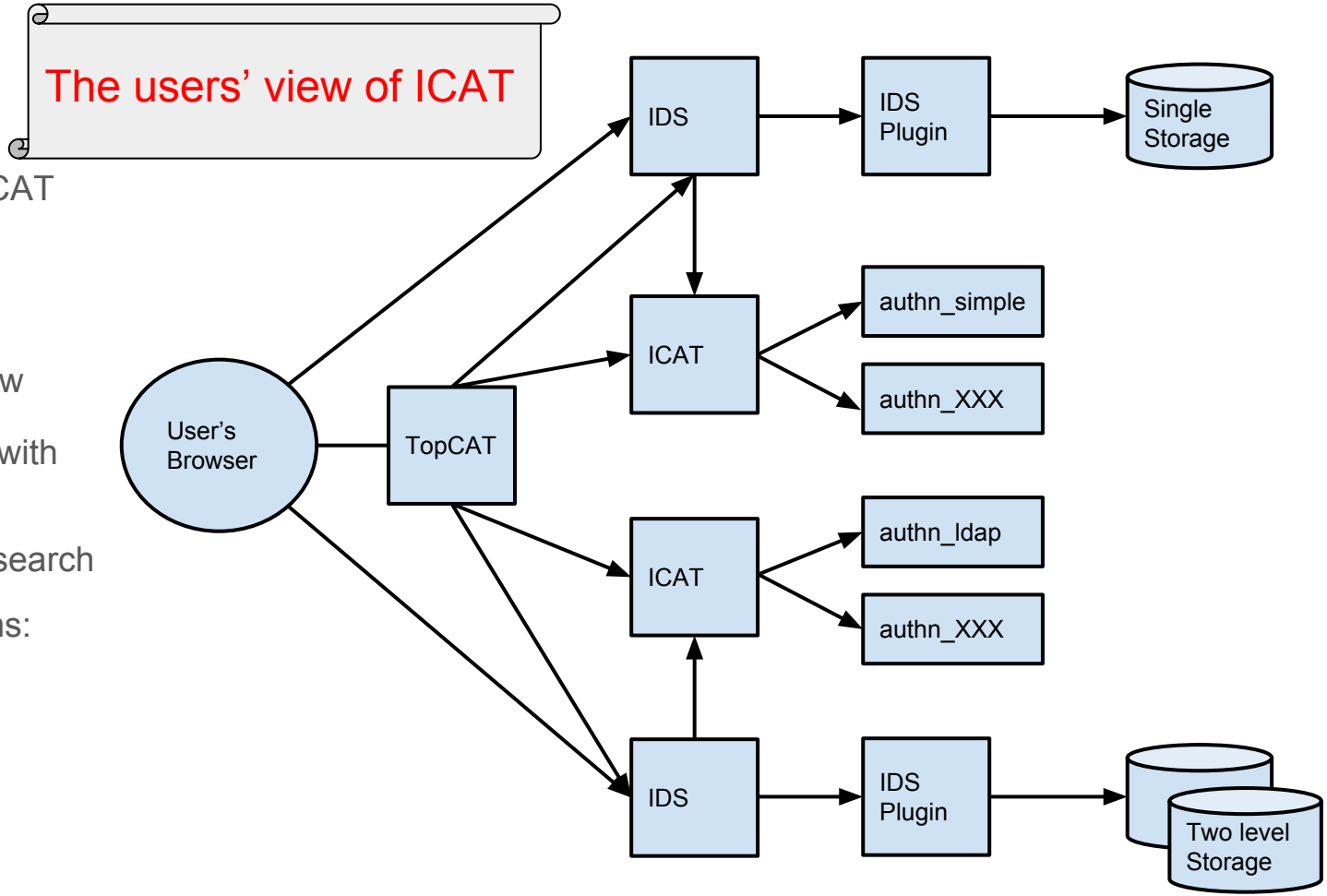


IDS Server - calls

- Storing a file
  - Send in the id of the dataset and name of file
  - Look up dataset in ICAT and check that can write
  - Use the plugin to write the file
  - Catalog the datafile in ICAT
  - Return the id of the new datafile
- Reading one file
  - Check with ICAT that file is readable
  - If file in main storage plugin streams data directly to the user
- Reading multiple files
  - Check with ICAT that all files are readable
  - The same mechanism but a zip file is created on the fly

TopCAT

- Interface to multiple ICAT and IDS servers
- Highly configurable
- Facility dependent view
- New interface written with Angular JS
- Makes use of lucene search
- Download mechanisms:
  - http(s)
  - smartclient
  - various PollCATs
- Pluggable



vanadium

Start Date

End Date

Visit

Dataset

Datafile

Search

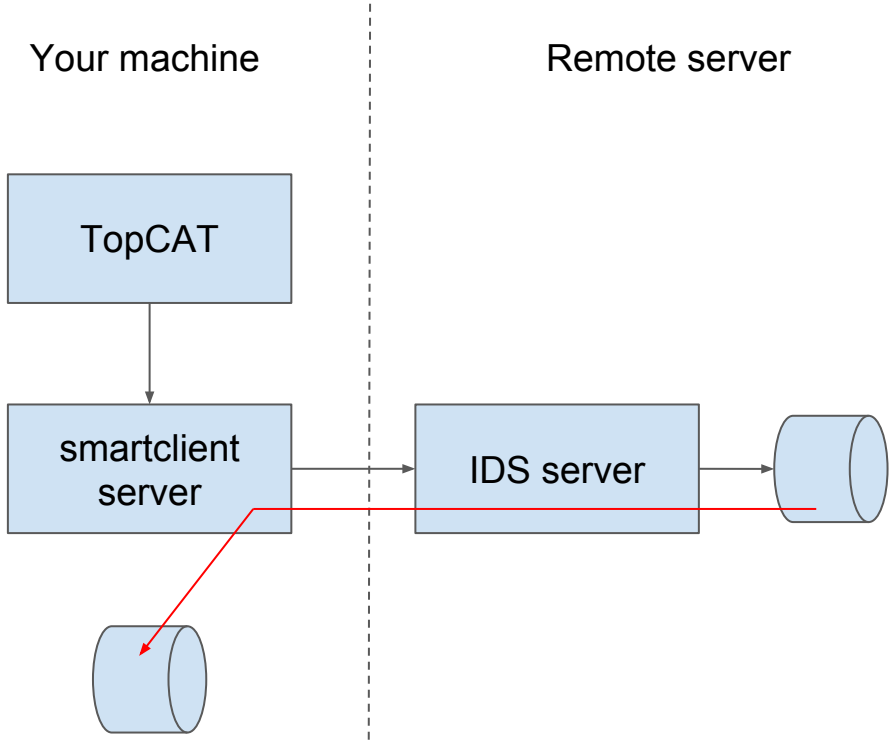
Search Results

Visit Dataset Datafile

Title	Visit Id	Size	Beamline	Start Date	End Date
Containing...	Containing...		Containing...	From... To...	From... To...
[illegible]	EE7999-1	1.26 GB	i11	2012-10-08	2012-10-11
[illegible]	MT10099-1	24.27 GB	i19	2015-02-11	2015-02-13
[illegible]	MT6138-1	55.4 GB	i16	2010-10-06	2010-10-12
[illegible]	SM13188-1	905 B	i22	2015-05-14	2015-05-15
[illegible]	SP11874-1	2.31 GB	b18	2015-05-01	2015-05-05
[illegible]	SI12546-1	1.17 MB	i09	2015-06-29	2015-07-04
[illegible]	SP10434-1	134.45 MB	b18	2014-10-08	2014-10-09
[illegible]	EE7667-1	896.96 MB	i11	2012-08-06	2012-08-09
[illegible]	SI9122-1	55.56 GB	i06-1	2014-03-05	2014-03-11
[illegible]	SI11246-1	64.4 MB	i09	2015-06-24	2015-06-29
[illegible]	MT1036-1	2.27 MB	i16	2009-05-06	2009-05-12
[illegible]	SP6120-1	234.69 GB	b18	2011-02-24	2011-02-27
[illegible]	EE7637-1	2 GB	i11	2012-07-26	2012-07-28
[illegible]	MT556-1	3.99 GB	i16	2008-07-02	2008-07-07
[illegible]	SP12696-1	5.59 GB	i18	2016-02-08	2016-02-12
[illegible]	SP8577-1	822.91 MB	b18	2013-07-24	2013-07-27
[illegible]	SI13517-1	2 kB	i09	2016-01-24	2016-01-29
[illegible]	SI12885-1	905 B	i09-2	2016-02-03	2016-02-07

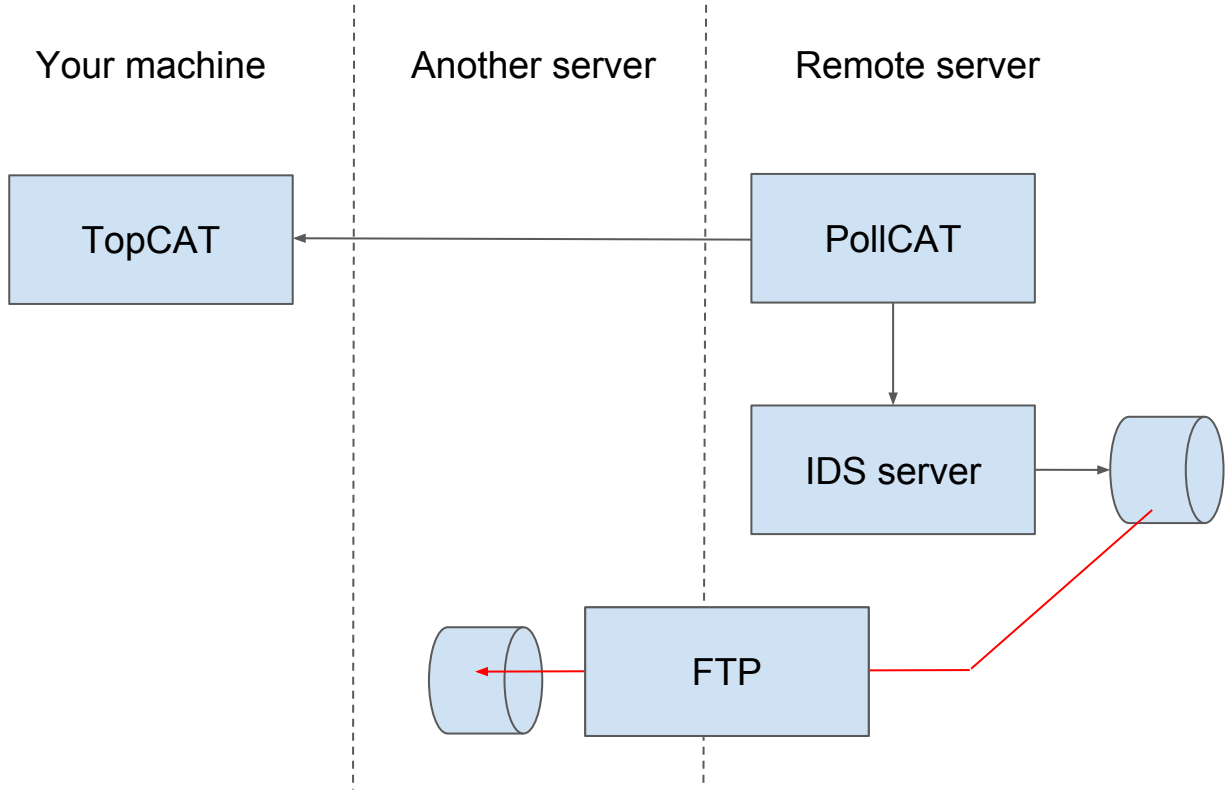
smartclient

- No infrastructure required
- Self contained distributions for Windows, RedHat, Debian, Mac
- Parallel streams of data to your filesystem running in the background



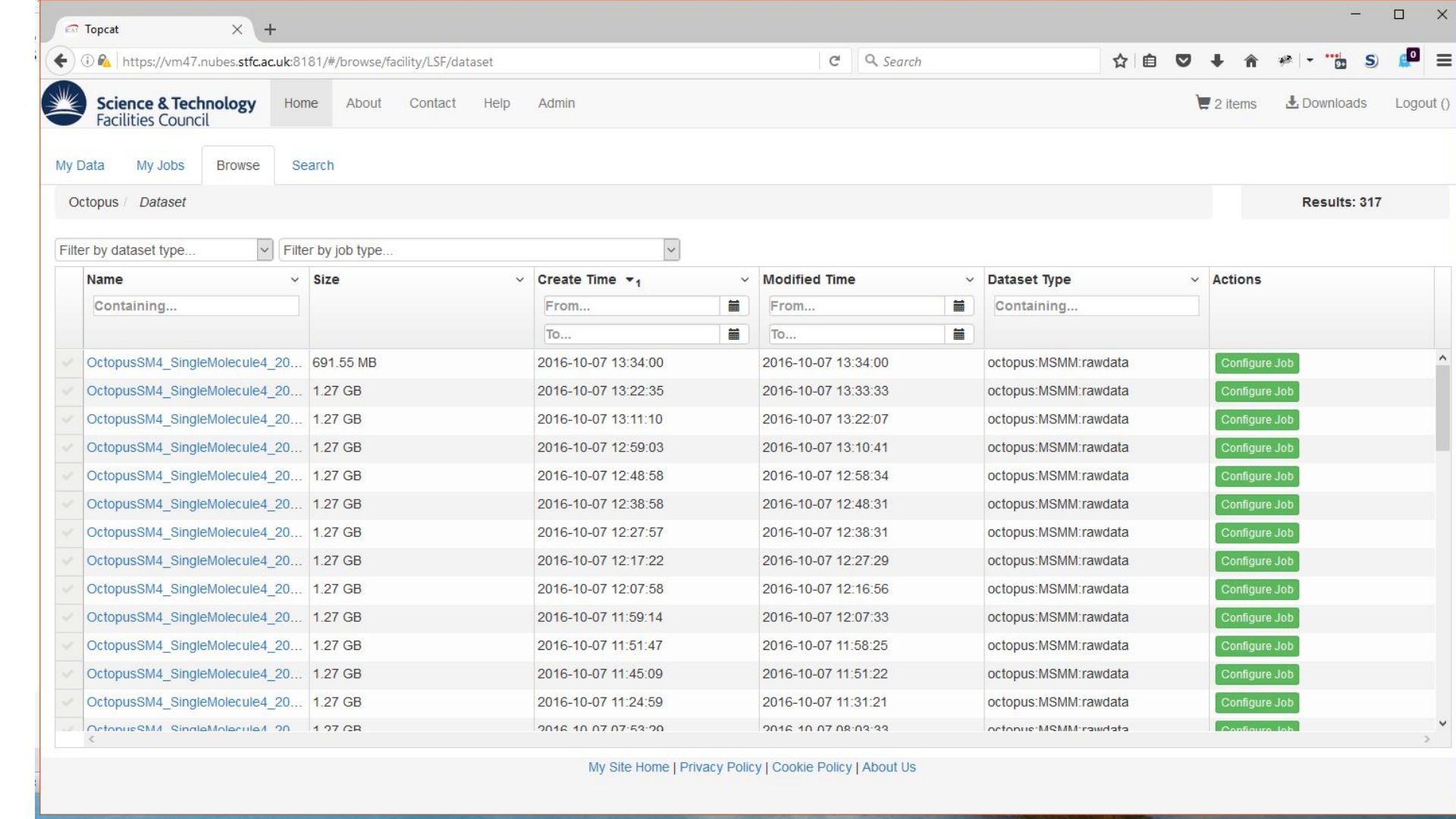
PollCAT

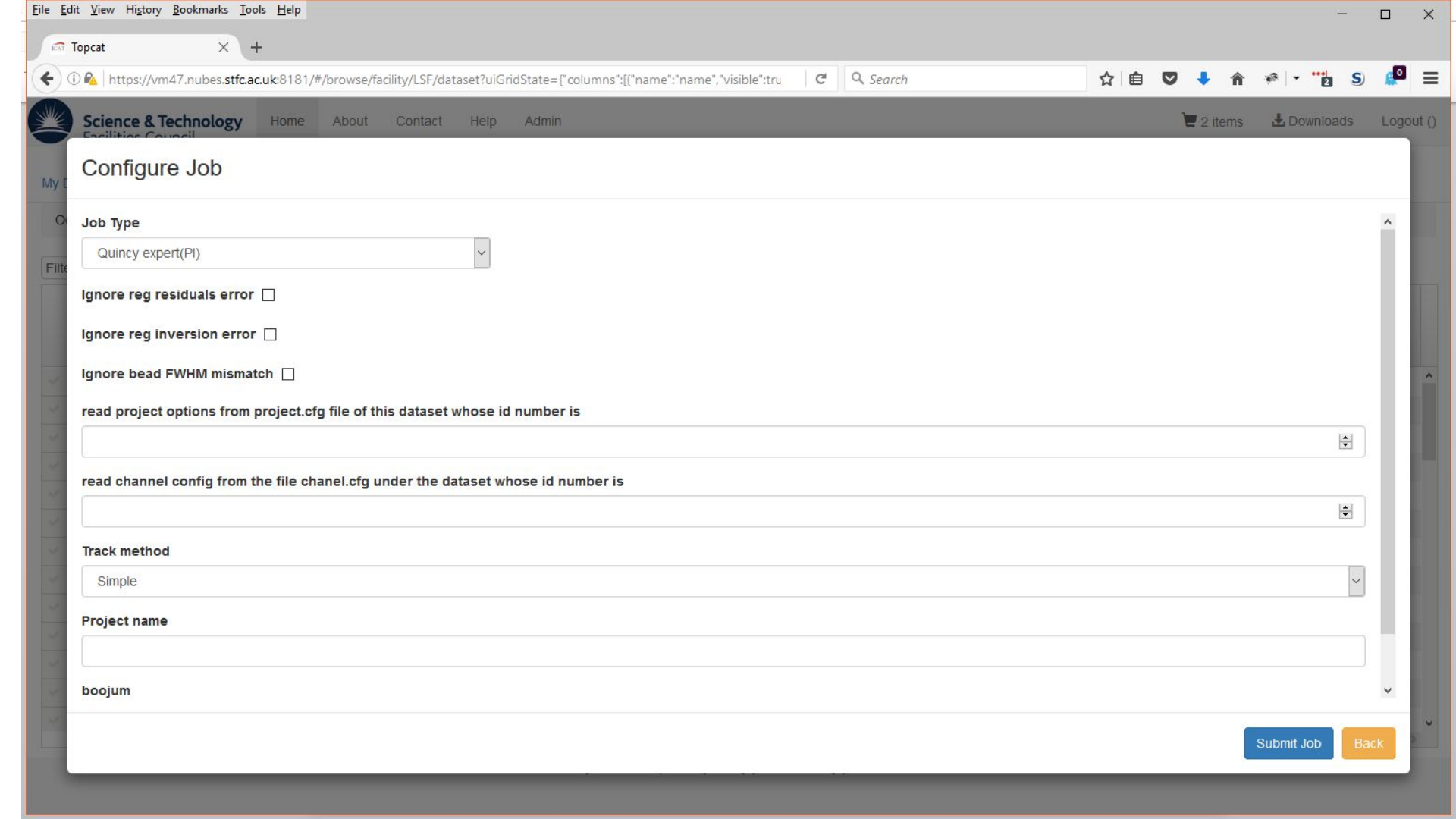
- PollCAT polls the IDS until data ready
- Then moves it via a plugin

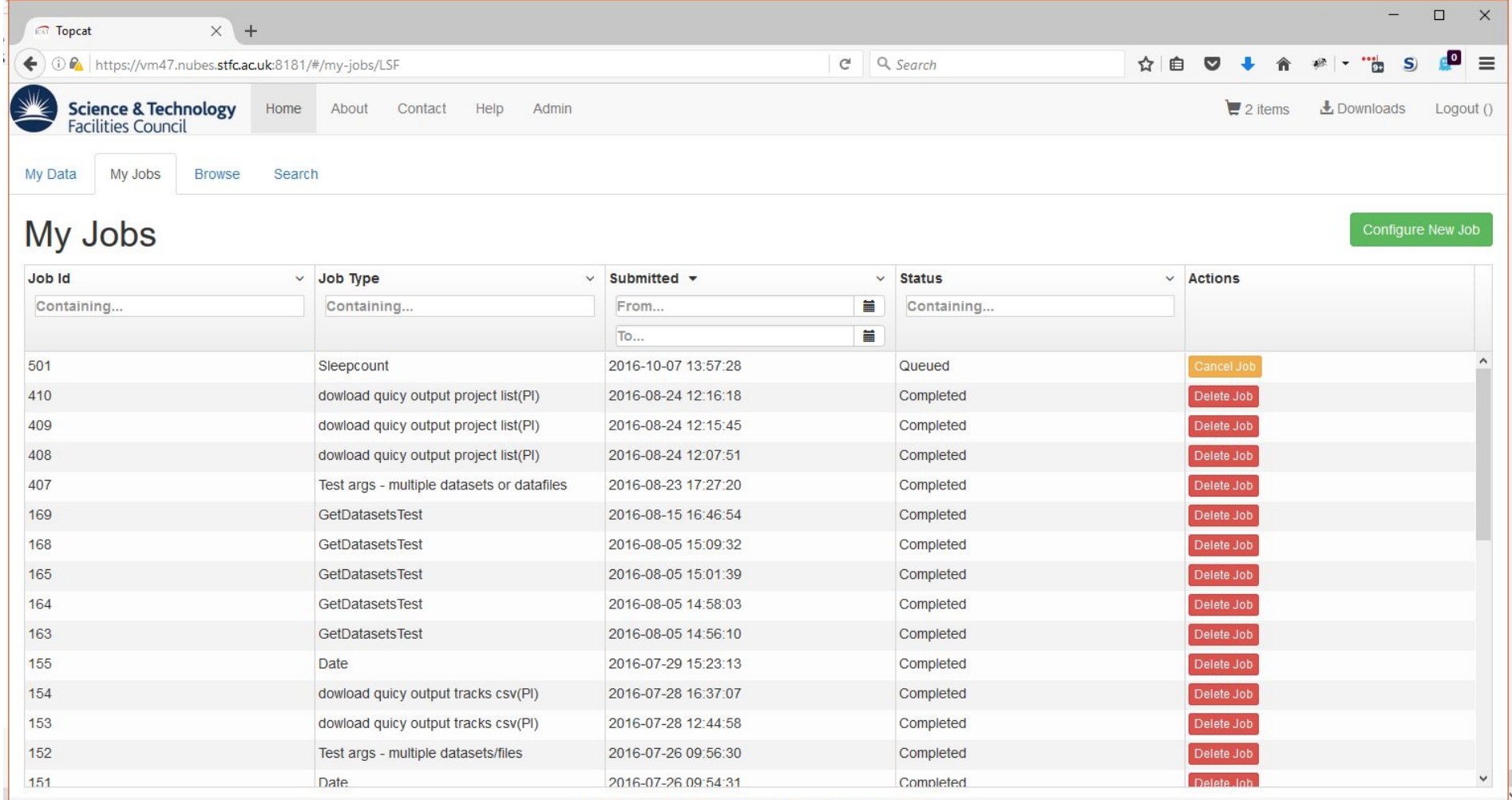


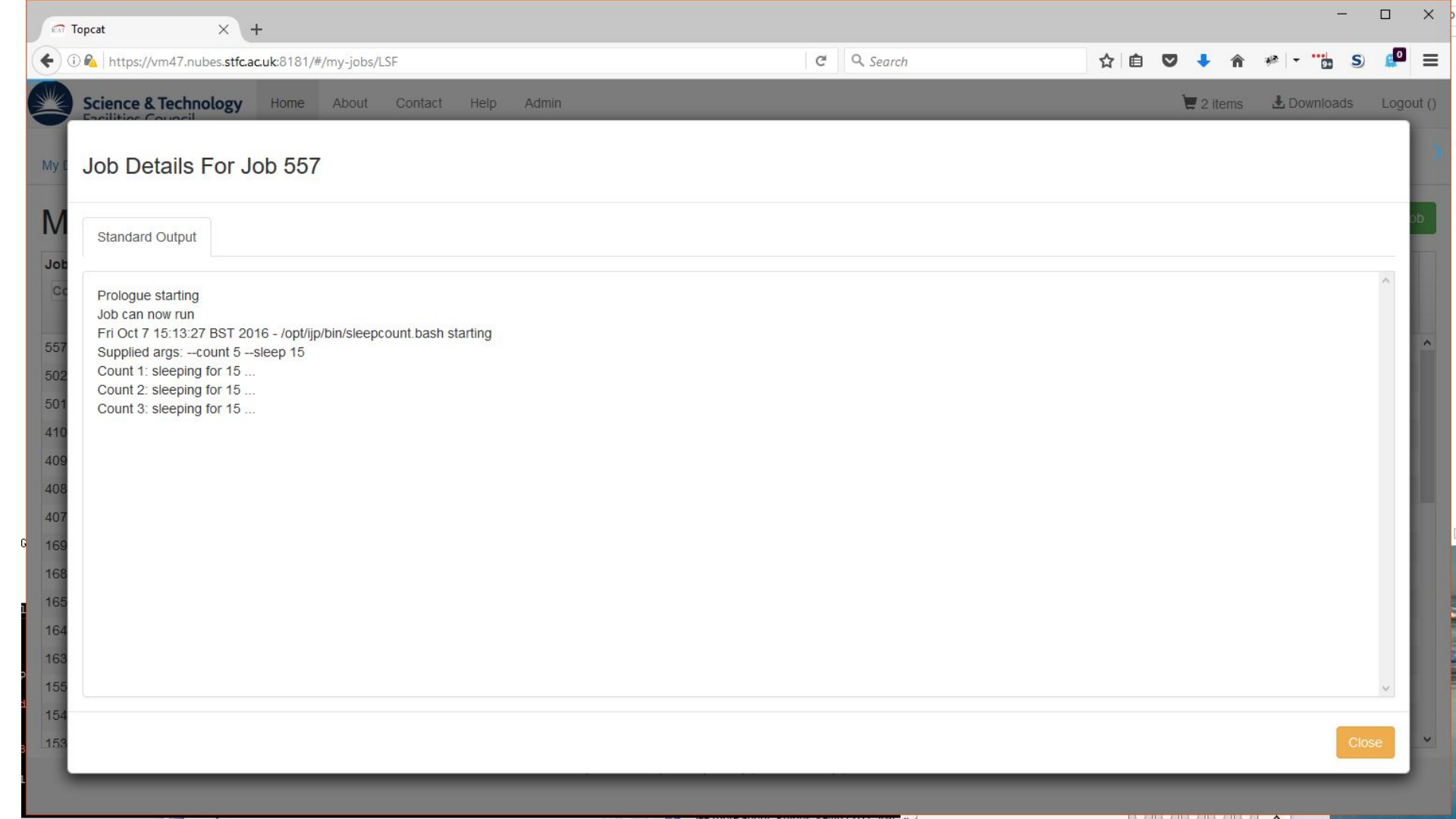
IJP

- ICAT Job Portal
- Was a standalone component
- Now implemented as a TopCAT plugin so feels like TopCAT


Topcat
https://vm47.nubes.stfc.ac.uk:8181/#/browse/facility/LSF/dataset
Search
Science & Technology Facilities Council
Home
About
Contact
Help
Admin
2 items
Downloads
Logout ()
My Data
My Jobs
Browse
Search
Octopus / Dataset
Results: 317
Filter by dataset type...
Filter by job type...
Name
Containing...
Size
Create Time
From...
To...
Modified Time
From...
To...
Dataset Type
Containing...
Actions
OctopusSM4_SingleMolecule4_20... 691.55 MB 2016-10-07 13:34:00 2016-10-07 13:34:00 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 13:22:35 2016-10-07 13:33:33 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 13:11:10 2016-10-07 13:22:07 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 12:59:03 2016-10-07 13:10:41 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 12:48:58 2016-10-07 12:58:34 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 12:38:58 2016-10-07 12:48:31 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 12:27:57 2016-10-07 12:38:31 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 12:17:22 2016-10-07 12:27:29 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 12:07:58 2016-10-07 12:16:56 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 11:59:14 2016-10-07 12:07:33 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 11:51:47 2016-10-07 11:58:25 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 11:45:09 2016-10-07 11:51:22 octopus:MSMM:rawdata Configure Job
OctopusSM4_SingleMolecule4_20... 1.27 GB 2016-10-07 11:24:59 2016-10-07 11:31:21 octopus:MSMM:rawdata Configure Job
My Site Home | Privacy Policy | Cookie Policy | About Us

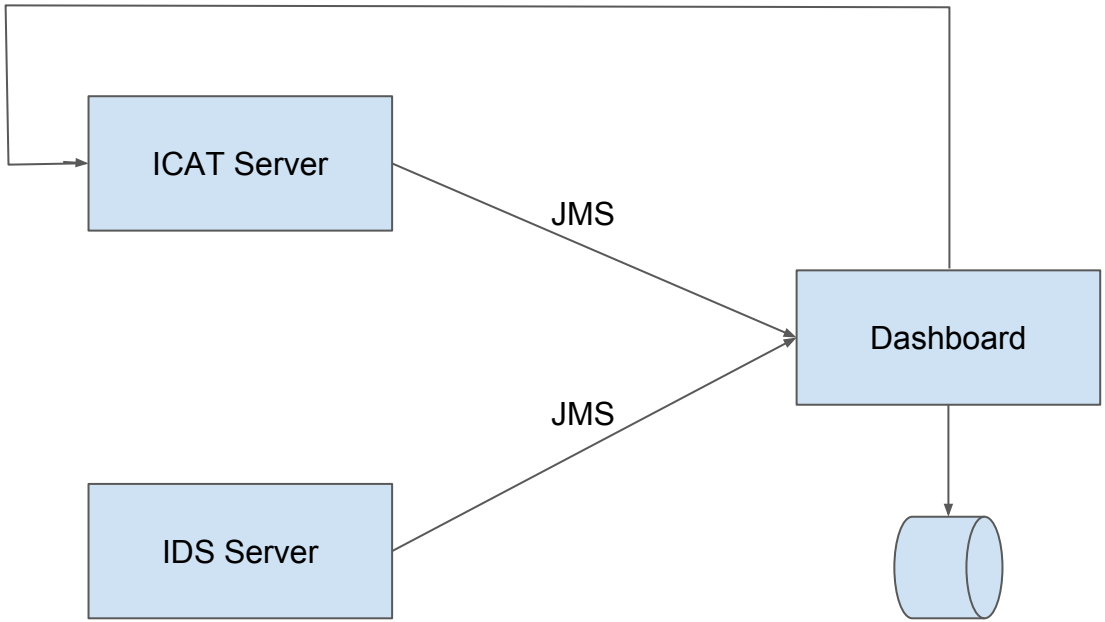

File Edit View History Bookmarks Tools Help
Topcat
https://vm47.nubes.stfc.ac.uk:8181/#/browse/facility/LSF/dataset?uiGridState={"columns":[{"name":"name","visible":tru
Search
Science & Technology Facilities Council
Home
About
Contact
Help
Admin
2 items
Downloads
Logout ()
Configure Job
Job Type
Quincy expert(PI)
Ignore reg residuals error
Ignore reg inversion error
Ignore bead FWHM mismatch
read project options from project.cfg file of this dataset whose id number is
read channel config from the file chanel.cfg under the dataset whose id number is
Track method
Simple
Project name
boojum
Submit Job
Back


Topcat
https://vm47.nubes.stfc.ac.uk:8181/#/my-jobs/LSF
Search
Science & Technology Facilities Council
Home
About
Contact
Help
Admin
2 items
Downloads
Logout ()
My Data
My Jobs
Browse
Search
My Jobs
Configure New Job
Job Id
Job Type
Submitted
Status
Actions
Containing...
Containing...
From...
To...
Containing...
501 Sleepcount 2016-10-07 13:57:28 Queued Cancel Job
410 dowload quicy output project list(PI) 2016-08-24 12:16:18 Completed Delete Job
409 dowload quicy output project list(PI) 2016-08-24 12:15:45 Completed Delete Job
408 dowload quicy output project list(PI) 2016-08-24 12:07:51 Completed Delete Job
407 Test args - multiple datasets or datafiles 2016-08-23 17:27:20 Completed Delete Job
169 GetDatasetsTest 2016-08-15 16:46:54 Completed Delete Job
168 GetDatasetsTest 2016-08-05 15:09:32 Completed Delete Job
165 GetDatasetsTest 2016-08-05 15:01:39 Completed Delete Job
164 GetDatasetsTest 2016-08-05 14:58:03 Completed Delete Job
163 GetDatasetsTest 2016-08-05 14:56:10 Completed Delete Job
155 Date 2016-07-29 15:23:13 Completed Delete Job
154 dowload quicy output tracks csv(PI) 2016-07-28 16:37:07 Completed Delete Job
153 dowload quicy output tracks csv(PI) 2016-07-28 12:44:58 Completed Delete Job
152 Test args - multiple datasets/files 2016-07-26 09:56:30 Completed Delete Job
151 Date 2016-07-26 09:54:31 Completed Delete Job
My Site Home | Privacy Policy | Cookie Policy | About Us


Topcat
https://vm47.nubes.stfc.ac.uk:8181/#/my-jobs/LSF
Search
Science & Technology Facilities Council
Home
About
Contact
Help
Admin
2 items
Downloads
Logout ()
Job Details For Job 557
Standard Output
Prologue starting
Job can now run
Fri Oct 7 15:13:27 BST 2016 - /opt/ijp/bin/sleepcount.bash starting
Supplied args: --count 5 --sleep 15
Count 1: sleeping for 15 ...
Count 2: sleeping for 15 ...
Count 3: sleeping for 15 ...
Close

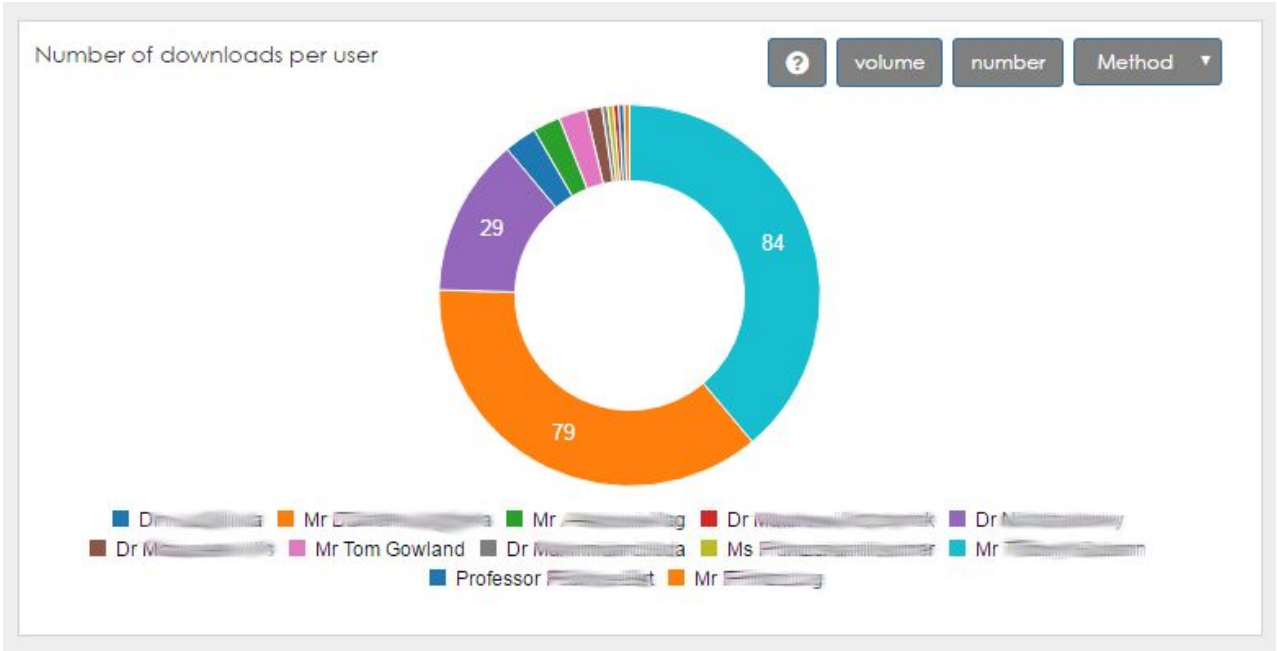
Dashboard

- Web based GUI (Angular JS) to give overview of ICAT Usage



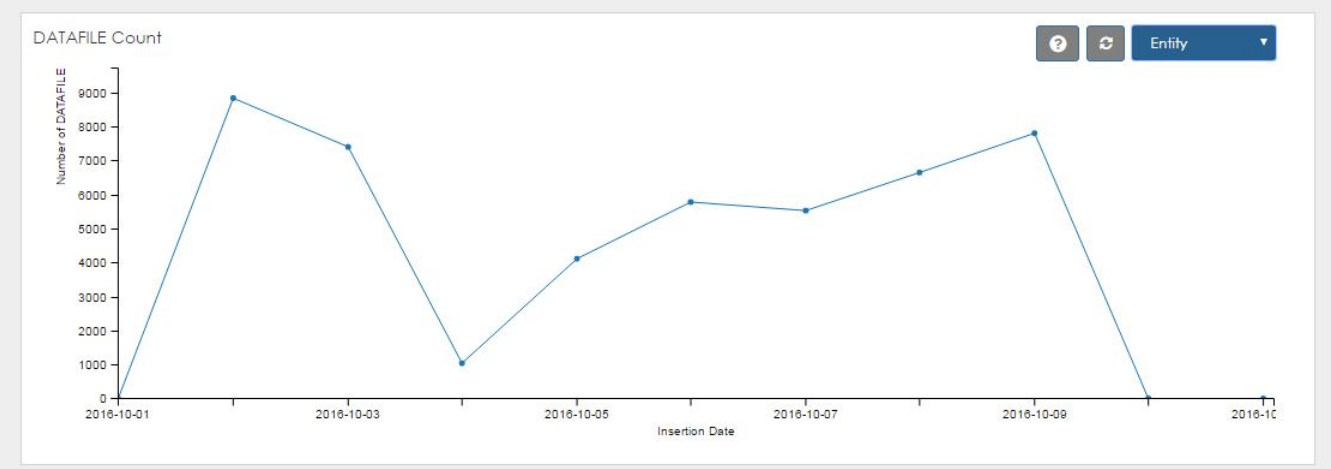
Dashboard

- Web based GUI (Angular JS) to give overview of ICAT Usage



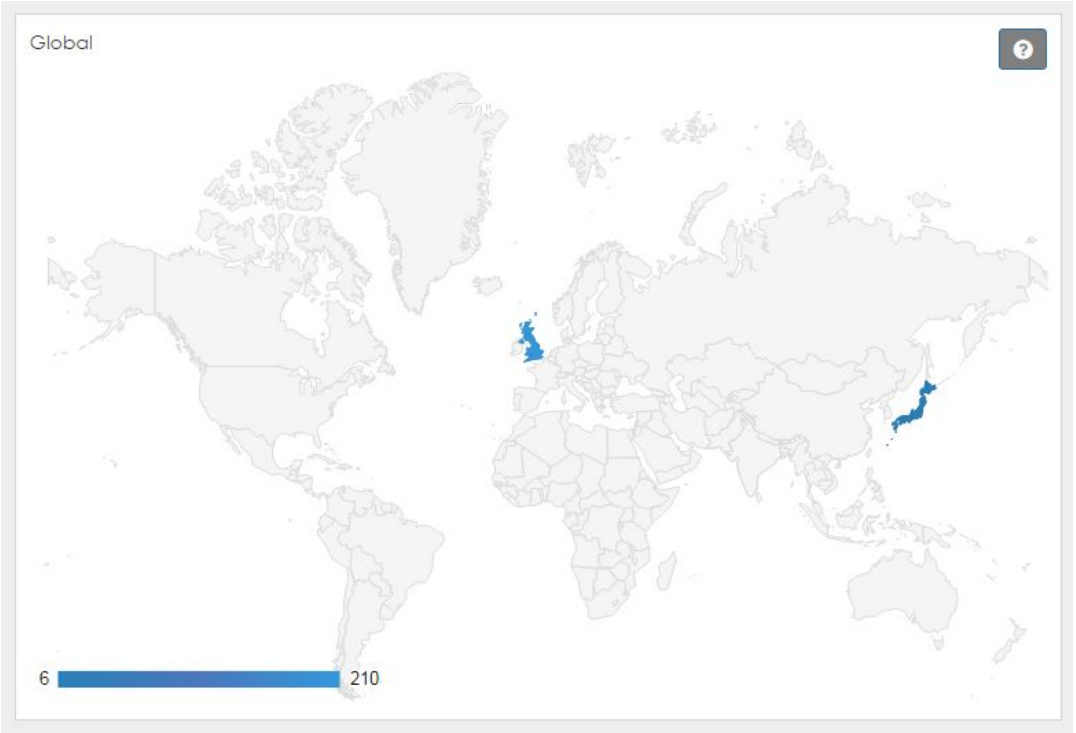
Dashboard

- Web based GUI (Angular JS) to give overview of ICAT Usage



Dashboard

- Web based GUI (Angular JS) to give overview of ICAT Usage



Conclusion

- The model is practical
- Loose coupling and plugins provide flexibility to support many facilities
- The ICAT project is very much alive

https://icatproject.org