PaNdata ODI 1st Open Workshop

Dublin 24-25/28th of March 2014

Co-located with the RDA 3rd plenary at Croke Park

https://indico.desy.de//event/1stow

Booklet of collected presentations

pandata ODI

ICAT workshop at trinity college in Dublin 24/25th of March

Brian Matthews:	ICAT collaboration and its governance
ICAT site reports:	ALBA, DESY, DLS, ELETTRA, ESRF, HZB, ISIS, SNS
Rolf Krahl:	python-icat - a library for writing ICAT clients in Python
Tom Griffin :	Domain specific ICAT GUI
Tom Griffin:	ISIS authz rules
Rolf Krahl:	Styles of rule and is the authz system adequate?
Jay Rainey:	Mantid and ICAT
Brian Matthews:	ISIS Data Journal and SCAPE
Steve Fisher:	IJP for LSF
Jay Rainey:	Experience of the ICAT API and Documentation
Steve Fisher:	J2EE containers
Steve Fisher:	ICAT, IDS & IJP
Wayne Chung:	TopCAT

ICAT Steering Committee

Brian Matthews

Research Data Group
Scientific Computing Department
STFC Rutherford Appleton Laboratory

brian.matthews@stfc.ac.uk

Science & Technology Facilities Council

Growing number of Partners in ICAT

- ICAT developed with a small number of partners
  - ISIS, DLS
  - SNS
- PaNData has included more
  - ILL, ELLETRA, ESRF etc
- Others also using or are interested
  - CLF, CCFE, …

Science & Technology Facilities Council

Managing an Open Source Collaboration

- To date
  - Developer group
  - Discussion with individual facilities
  - PanData
- But group getting larger
  - Need to make sure all have a say
- Sustainability
  - Current PanData project soon finishing
  - Need to ensure sustainability for ICAT

Science & Technology Facilities Council

ICAT Steering Group

- Made up of representatives of ICAT User facilities
  - Production or near production users

- Purpose:
  - To advise on the likely usage and requirements for ICAT;
  - To provide senior management oversight for the work of the ICAT collaboration partners;
  - To provide strategic guidance on the development and deployment of ICAT.

Science & Technology Facilities Council

Responsibilities of the ICAT Steering Committee

- to provide strategic advice to the ICAT collaboration;
- to protect the interests of the institutions;
- to advise on new initiatives and plans which may affect ICAT;
- to advise on major requirements which may affect ICAT;
- to oversee the ICAT development roadmap,
  - including the incorporation of additional components;
- to ensure that ICAT components continue to remain aligned to the users,
  - a common functionality and code base;
- to advise on ownership, licencing of components and commercialisation opportunities;
- to ensure that the institutions provide appropriate resources to the collaboration.

Science & Technology Facilities Council

Membership

- The Committee has one representative from each the stakeholders who are actively involved in developing, deploying and exploiting ICAT.
  - The Committee will regularly review membership
- The Project Manager of the ICAT Project should attend meetings in a non-voting capacity.
- A Chairperson elected from its members;
- A Secretary provided by STFC Scientific Computing.

Current Members ?

- a representative of ISIS;
- a representative of DLS;
- a representative of CLF;
- a representative of SNS;
- a representative of Scientific Computing at STFC;
- a representative of ILL;
- a representative of ESRF;
- a representative of ELETTRA;
- a representative of the PanData consortium;

Science & Technology Facilities Council

Meetings

- 1 – 2 times a year as needed
- Telco/VC
  - Though we will probably share a room at RAL
- 1 – 1.5 hrs
- Shared area on ICAT web-site

Science & Technology Facilities Council

First Meeting

- ASAP
  - By end April ?
  - Brian Matthews will temporarily chair until a new chair is elected.
- Representatives ?
- Agenda ?
  - Chair of Committee
  - Review ToR and Membership
  - 2014-15 Development Roadmap
  - Facilities Plans and Requirements
  - Future Directions

Science & Technology Facilities Council

ICAT at ALBA

Daniel Salvat – MIS Section

Agenda

- Project overview
- What's next?

ALBA

Project Overview

- Project Team:
- Current versions:
  - ICAT: 4.2.5
  - TopCAT: 1.9
  - Java 6
  - Glassfish 3

September 2013

ALBA

Project Overview

- User Office

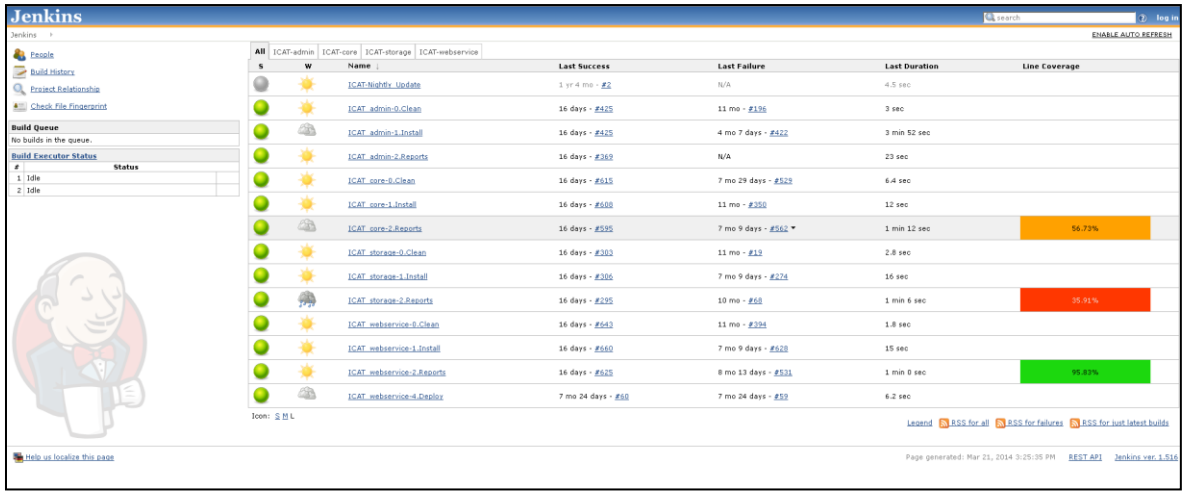
PROPOSALS SYNC
Quartz Jobs

Fire!	SyncBeamlineScientistsWithIcat	ICAT	Sat Mar 22 07:05:00 CET 2014
Fire!	SyncBeamlinesWithIcat	ICAT	Sat Mar 22 07:00:00 CET 2014
Fire!	SyncResearchAreasWithIcat	ICAT	Sat Mar 22 07:15:00 CET 2014
Fire!	SyncPeriodsWithIcat	ICAT	Sat Mar 22 07:10:00 CET 2014
Fire!	SyncProposalsWithIcat	ICAT	Sat Mar 22 07:20:00 CET 2014

ALBA

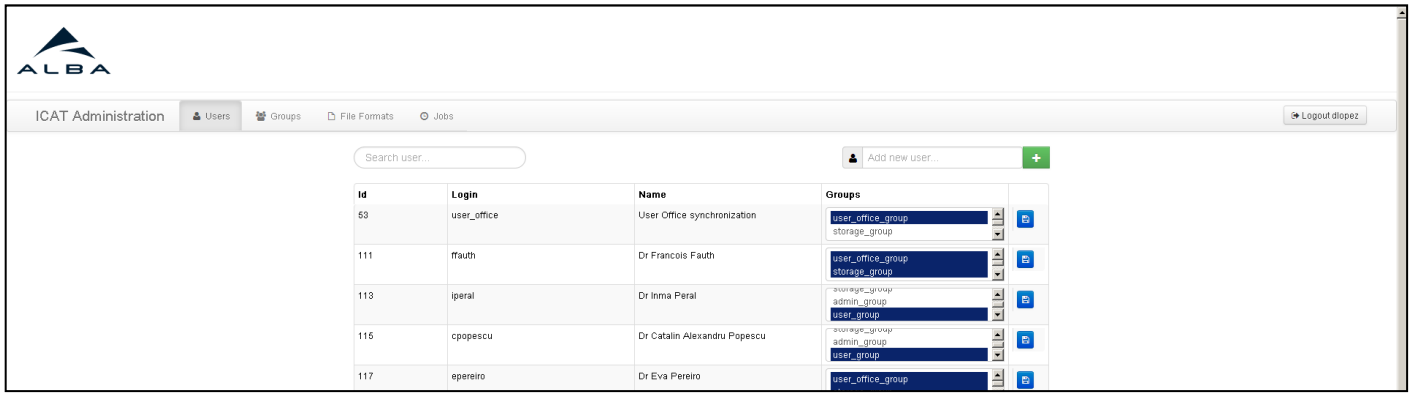
Project Overview

- 4 modules:
  - **Core**: controller of the system interface to ICAT.ear
  - **Admin**: User/Roles/File Format web interface
  - **Storage**: Data ingestion done via file parsing processes.
  - **Webservice**: Interface to the Core.
- CI: Jenkins



Project Overview

- ICAT administration:



ALBA

Project Overview

- TopCAT:

ALBA

What's next?

- Test environment (not production).
  - Need to test with real data.
  - Performance?
- IDS is still to be implemented.
  - Will not go to a production environment without.
- Interested in ingestion via python interface.
- Comments:
  - Upgrades:
    - Java 6 → Java 7
    - Glassfish 3 → Glassfish 4

Effort ?

ALBA

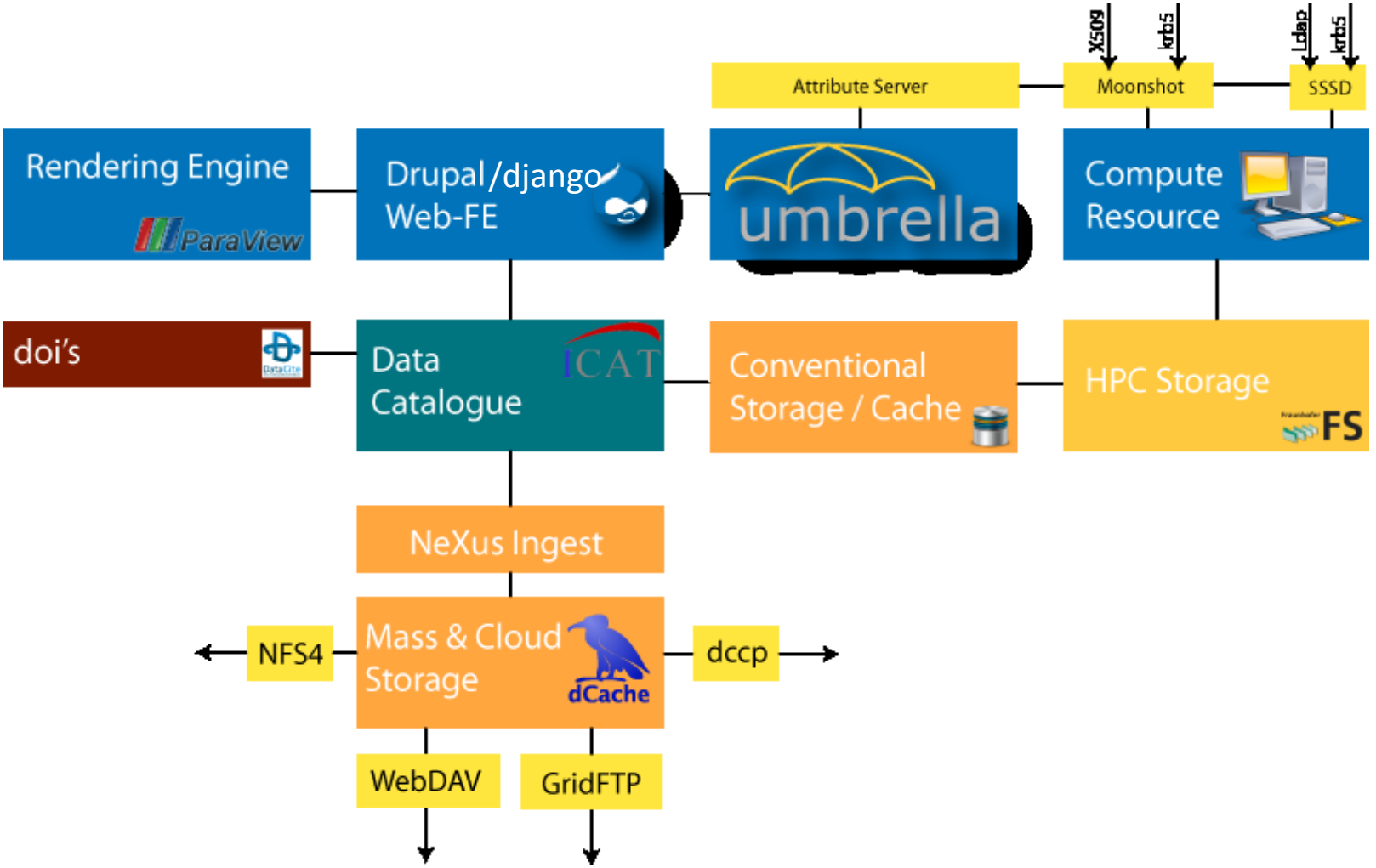
ICAT at ALBA

Thank you

ICAT status @ DESY

- ICAT installed and running (details: Jürgen)
  - Current plan: only for open access data
    - No concrete plan to host all petra3 data in icat
    - The oa ICAT will also serve as a platform to re-evaluate
  - Basic requirement: PI can alter acl's at any time
    - Seems a major obstacle so far
  - Essentially no data ingested yet
    - Problem: open access = published data, but few meta-data recorded
    - Chicken-Egg: don't invest effort for a scarcely filled data catalogue
  - Annotation and ingest of derived data via web-frontend
    - Access control through umbrella & basic (door)
    - WebDAV export

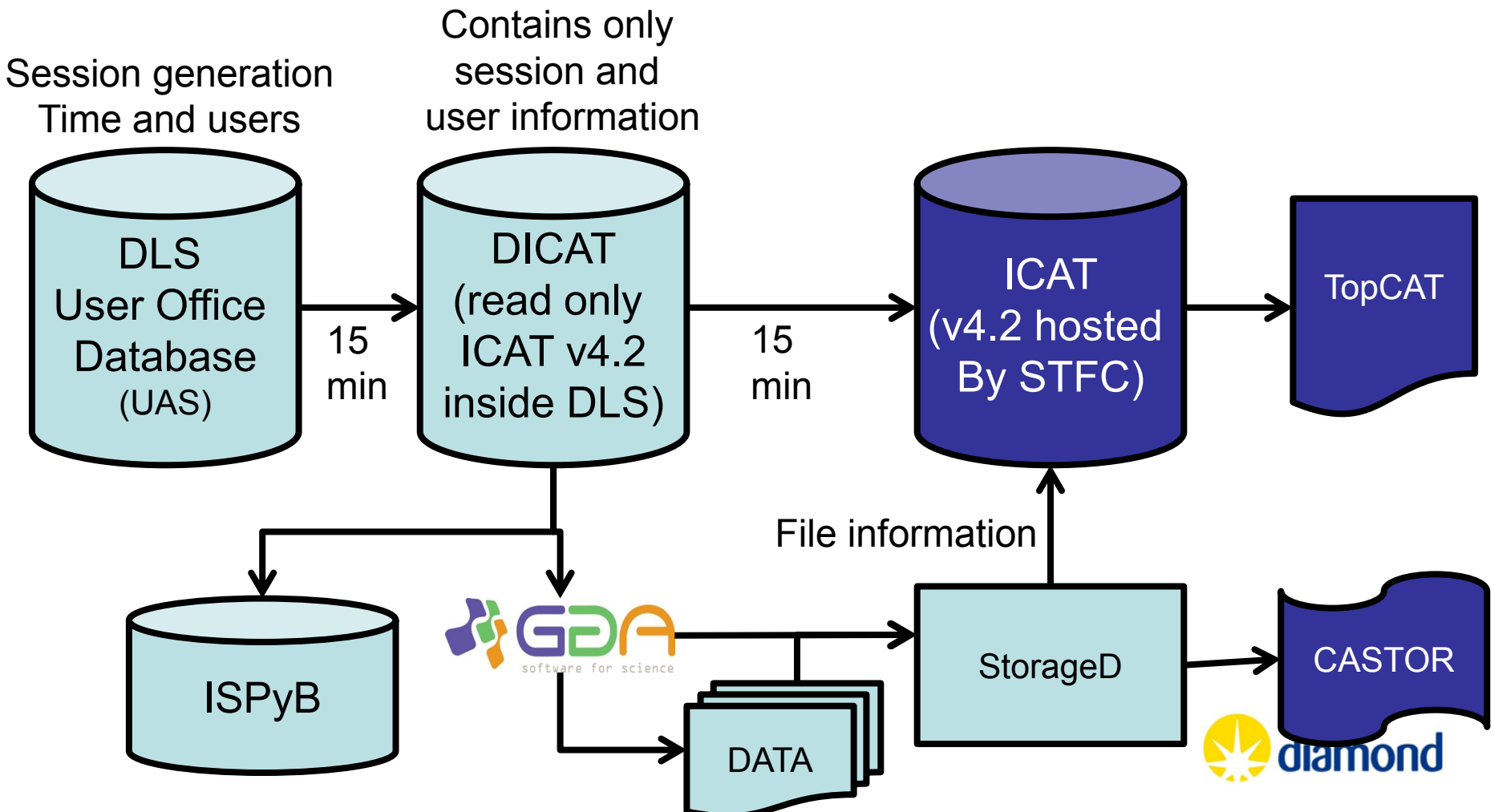
Rough plan


X509
krb5
ldap
krb5
Attribute Server
Moonshot
SSSD
Rendering Engine
ParaView
Drupal/django
Web-FE
umbrella
Compute
Resource
doi's
Data
Catalogue
ICAT
Conventional
Storage / Cache
HPC Storage
FS
NeXus Ingest
NFS4
Mass & Cloud
Storage
dCache
dccp
WebDAV
GridFTP

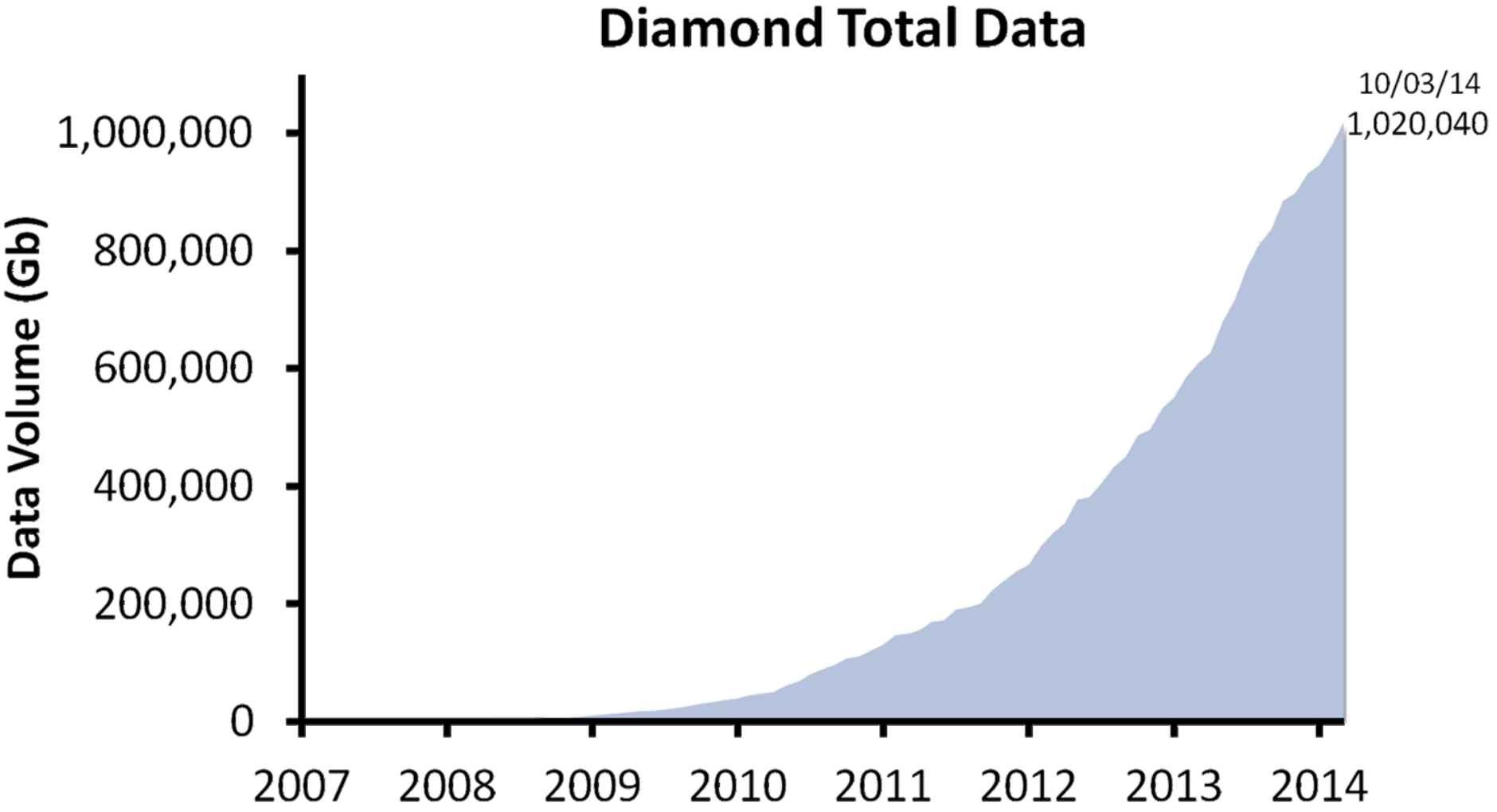
Diamond ICAT Status Report

Diamond migration from 3.3 to 4.2

- Diamond will only be using ICAT 4.2 from March 28.



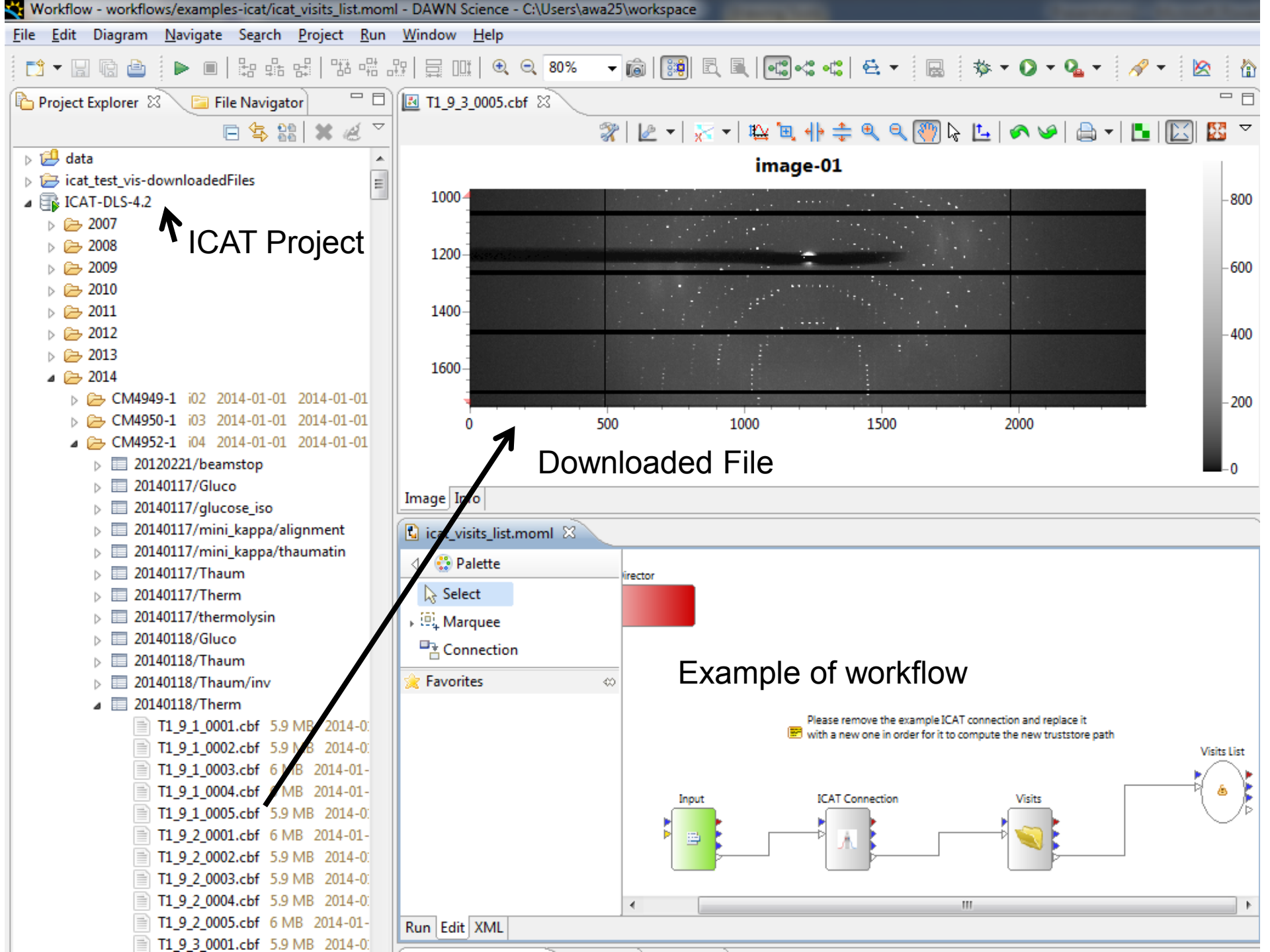
Current status/volumes in Diamond
ICAT = 285,198,074 files



ICAT 4.2

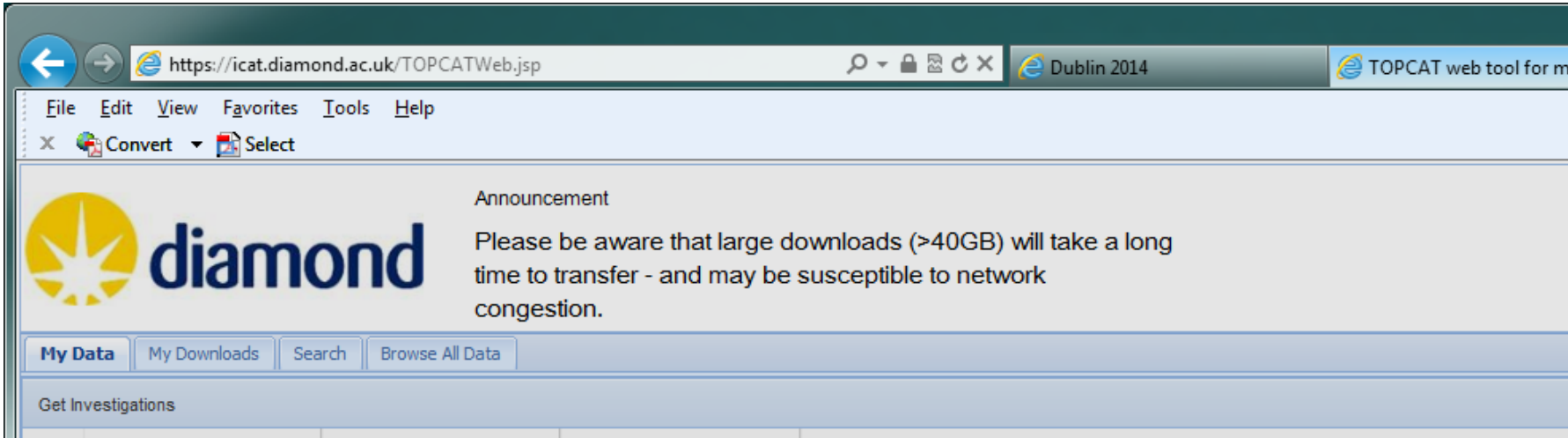
- Successfully integrated and used as source of current visit by
  - Data Acquisition framework in Diamond (GDA)
  - Automated session/visit directory structure generation
  - Projects in DAWN
  - Workflows in Passerelle/DAWN

diamond


Workflow - workflows/examples-icat/icat_visits_list.moml - DAWN Science - C:\Users\awa25\workspace
File Edit Diagram Navigate Search Project Run Window Help
Project Explorer
File Navigator
data
icat_test_vis-downloadedFiles
ICAT-DLS-4.2
ICAT Project
2007
2008
2009
2010
2011
2012
2013
2014
CM4949-1 i02 2014-01-01 2014-01-01
CM4950-1 i03 2014-01-01 2014-01-01
CM4952-1 i04 2014-01-01 2014-01-01
20120221/beamstop
20140117/Gluco
20140117/glucose_iso
20140117/mini_kappa/alignment
20140117/mini_kappa/thaumatin
20140117/Thaum
20140117/Therm
20140117/thermolysin
20140118/Gluco
20140118/Thaum
20140118/Thaum/inv
20140118/Therm
T1_9_1_0001.cbf 5.9 MB
T1_9_1_0002.cbf 5.9 MB
T1_9_1_0003.cbf 6 MB
T1_9_1_0004.cbf MB
T1_9_1_0005.cbf 5.9 MB
T1_9_2_0001.cbf 6 MB
T1_9_2_0002.cbf 5.9 MB
T1_9_2_0003.cbf 5.9 MB
T1_9_2_0004.cbf 5.9 MB
T1_9_2_0005.cbf 6 MB
T1_9_3_0001.cbf 5.9 MB
T1_9_3_0005.cbf
image-01
1000
1200
1400
1600
0
500
1000
1500
2000
800
600
400
200
0
Downloaded File
Image
Info
icat_visits_list.moml
Palette
Select
Marquee
Connection
Favorites
irector
Example of workflow
Please remove the example ICAT connection and replace it
with a new one in order for it to compute the new truststore path
Visits List
Input
ICAT Connection
Visits
Run
Edit
XML

Look out for....

- Cache for data coming off tape ready for download to user via ICAT/IDS
  - DLS now have multi TB cache.
- Increased the filename size in DB.
- Problems with users and the ‘dataset’ concept
  - “I know the filename and directory....”
- Data rates from site ‘improving’



Plans to migrate to 4.3

- None yet.

diamond

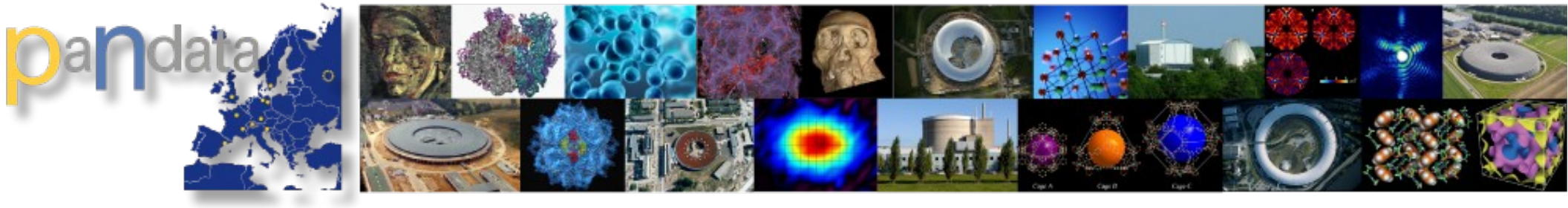
More details in...

- More details in RDA presentations....
- **PaNSIG Meeting Friday 28th 13:30-15:00**
  - Specific topics for future collaboration
    - Data analysis issues and frameworks - Alun Ashton / DLS
    - .....
- **Structural Biology IG Session *Thursday, March 27th, 13.30 -15.00***
  - ***Structural biology towards biomedicine and health***
    - 13.30 -13.50 Data management and processing pipelines at DLS Alun Ashton

diamond

ICAT Deployment

Status @ Elettra



Dublin, 24 March 2014.

ICAT

Elettra Sincrotrone Trieste

Milan Prica

IT – Scientific Computing Group

Elettra – Sincrotrone Trieste S.C.p.A.

Facility overview

- **Two light sources**

  - **ELETTRA** Synchrotron Radiation Facility
    - Operational since 1993 – users since 1994
    - 20 beamlines, 4 more in construction/upgrade

  - **FERMI@ELETTRA** Free Electron Laser
    - Operational since 2011 – users since 2012
    - 3 beamlines, 3 more planned

Virtual User Office (VUO)

- Operational since 1997
- Possible authentication with Umbrella (Nov. 2013); Facility LDAP > 8 years

- Contains plenty of useful metadata
- Uses DOIs
- Enforces Elettra's Data Policy
- Hosts data access service
  - Download / Upload
  - Proposal submitter(s) and selected BL scientist(s) may grant data access to other participants
- **https://vuo.elettra.eu**

Data file formats

- Fermi@Elettra:
  - **Custom HDF5** files on all beamlines
    - Integraged with TANGO
- Elettra:
  - **Custom HDF5** files on some beamlines
    - Part of endstation DAQ TANGO
    - Converters from legacy formats
  - Variety of data formats on older beamlines
    - TIFF, RAW, ASCII...

- **NeXus**: Not adopted

ICAT metadata catalogue

- Three instances deployed (1 public)
  - Updated to the latest ICAT 4.3.2
  - Glassfish 3 and 4 containers
  - Topcat 1.11.0 and ICE
  - Authentication with local and remote db, simple
  - LDAP module not compatible with the Elettra LDAP

Data ingestion software

- **iGEST**
  - Python module
  - Simple to configure (template based mappings)
    - Works with any HDF5 file (NeXus included)
  - Supports parallelism during ingestion
  - Compatible with the FERMI acquisition pipeline

iGEST mapping example

```
[icat]
icatwsdl=https://icat-elettra.grid.elettra.trieste.it:8181/ICATService/ICAT?wsdl
user=user@gmail.com
password=12345678901112
plugin=db

[template]
h5template=temp_vlabsample.h5

[icatmeta]
facility=entry/instrument/source/name
instrument=entry/instrument/name
investigationtype=entry/experiment_type
investigation=entry/experiment_identifier
datasettype=entry/dataset/type
dataset=entry/dataset/name
sample=entry/sample/name
sampletypename=entry/sample/type/name
sampleformula=entry/sample/chemical_formula
facilitycycle=entry/instrument/facilitycycle
```

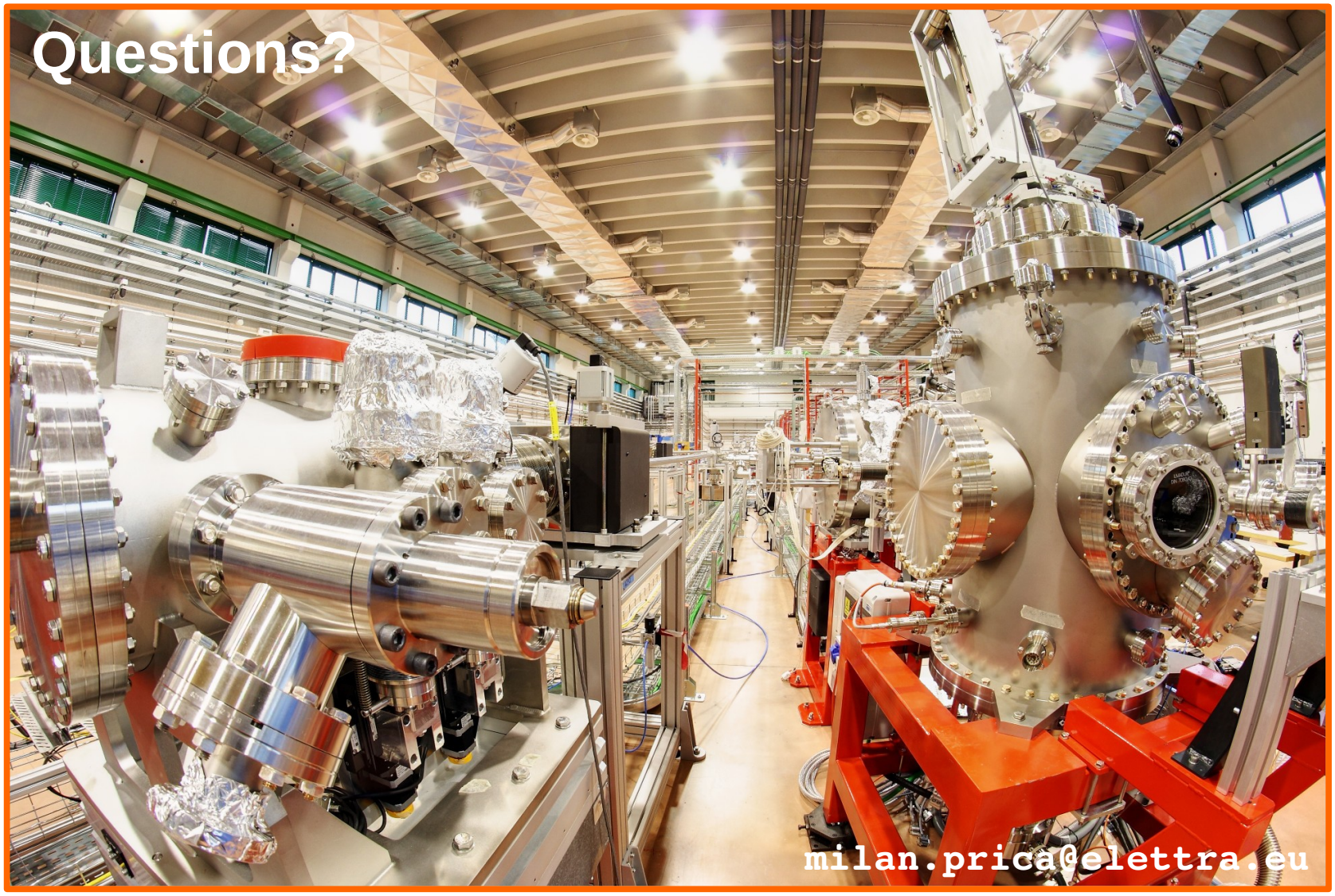
Issues

- LDAP Authentication:
  - Custom module needs to be developed for ELETTRA
  - Umbrella – ICAT authentication could be used?
- Rules and Users:
  - Rules to be defined for ELETTRA
  - Users should be inserted automatically from VUO

Issues

- Front-end (TopCAT):
  - A custom front-end integrated with VUO would be a better solution for ELETTRA
  - Users often perceive the front-end as “the catalogue”
  - Configuration tools should be part of the front-end
- Schema:
  - Many cross dependencies
  - Quite complex to match to the facility needs

Thank you!




ESRF
The European Synchrotron

Status of metadata at ESRF

ICAT deployment
Metadata ingestion
Current status
Next steps

ICAT deployment

Current software stack

- ✓ Glassfish 4.0
- ✓ ICAT 4.3.1 + authn_db 1.1.1
- ✓ TopCAT 1.11 + IDS 1.0.0 (+ custom plugin)

Databases

- ✓ Oracle (test and production)
- ✓ Derby (development)

Machines

- ✓ Oracle VM
- ✓ Sun JDK 1.7.0

Security

- ✓ Full SSL with signed certificates (TERENA)
- ✓ Apache reverse proxy (mod proxy http, address translation)

Some Issues during deployment

IDS not fully suitable yet for ESRF setup

- We need more operations delegated to the plugin, including logic for available/not available, direct/archived, path reconstruction, …

Problem with TopCAT behind a reverse proxy (GWT issue)

Address translation ignored:

- All .gwt.rpc files need to be copied

or

- Archive needs to be modified to change the context root

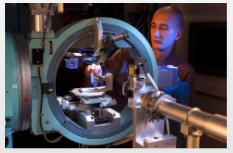
Other issues with proxy or firewall

- Problems connecting to external ICAT on port 8181 (solved)
- Problems connecting to ILL using ill.eu URL (ill.fr works)

Metadata ingestion

Click to edit Master text styles

The beamline



The ICAT ingester

Second level

Third level

Fourth level

- *Fifth level*

The ICAT installation

ICAT API

ICAT

RDBMS

NEW beamline control system

New Sequencer

Data Manager

Spec session

TANGO

Experiment metadata Management

WEB

Scientist controlling the Experiment

SMIS API

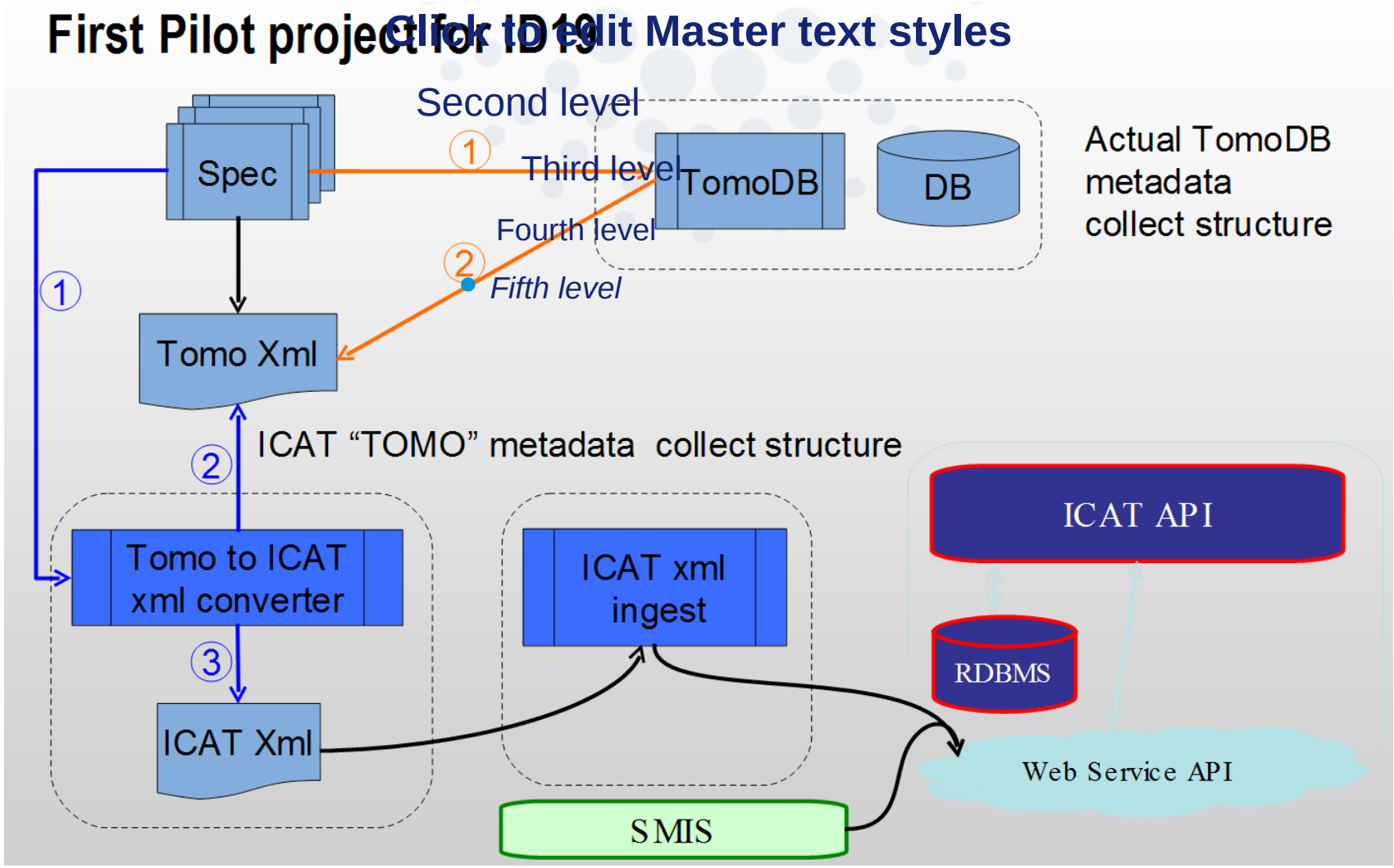
RDBMS

Web Service API

ICAT API

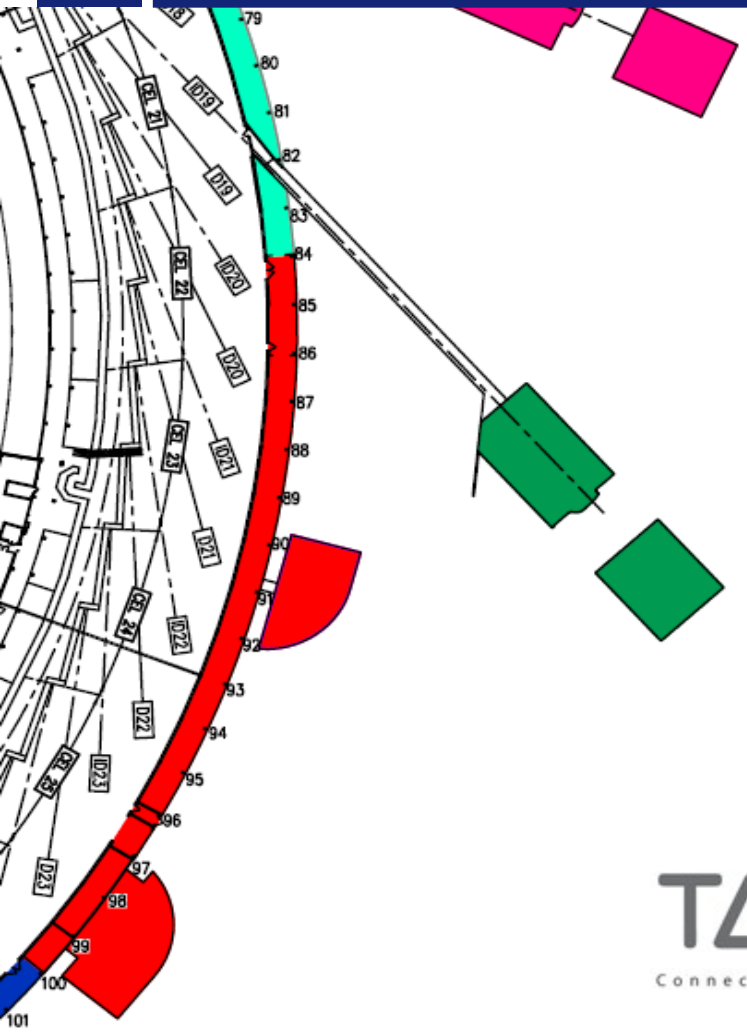
RDBMS

Web Service API



Data used for service verification imported using the Pilot

General case



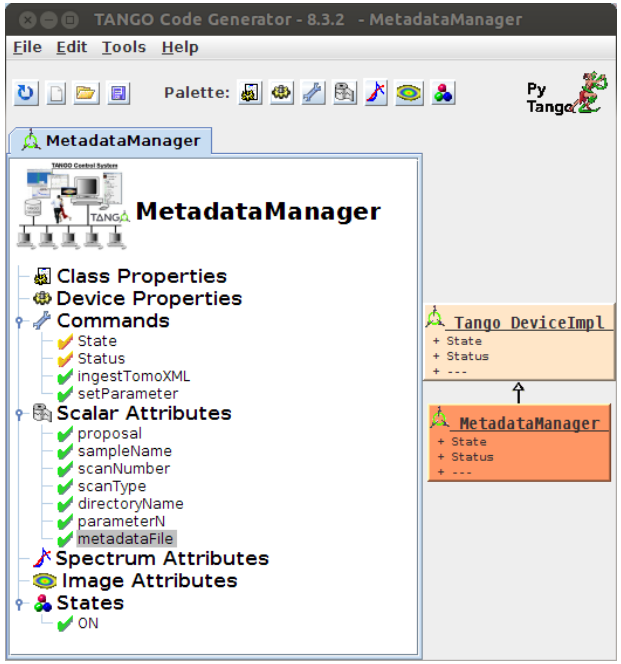
TANGO
Connecting things together

ISPyB
Information System for Protein crystallography Beamlines

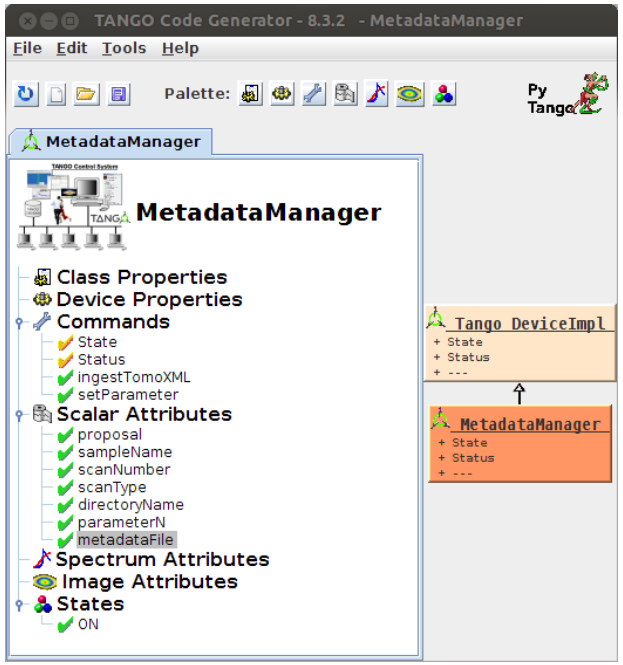
TANGO
Connecting things together

TomoDB II
Tomographic information system

TANGO
Connecting things together

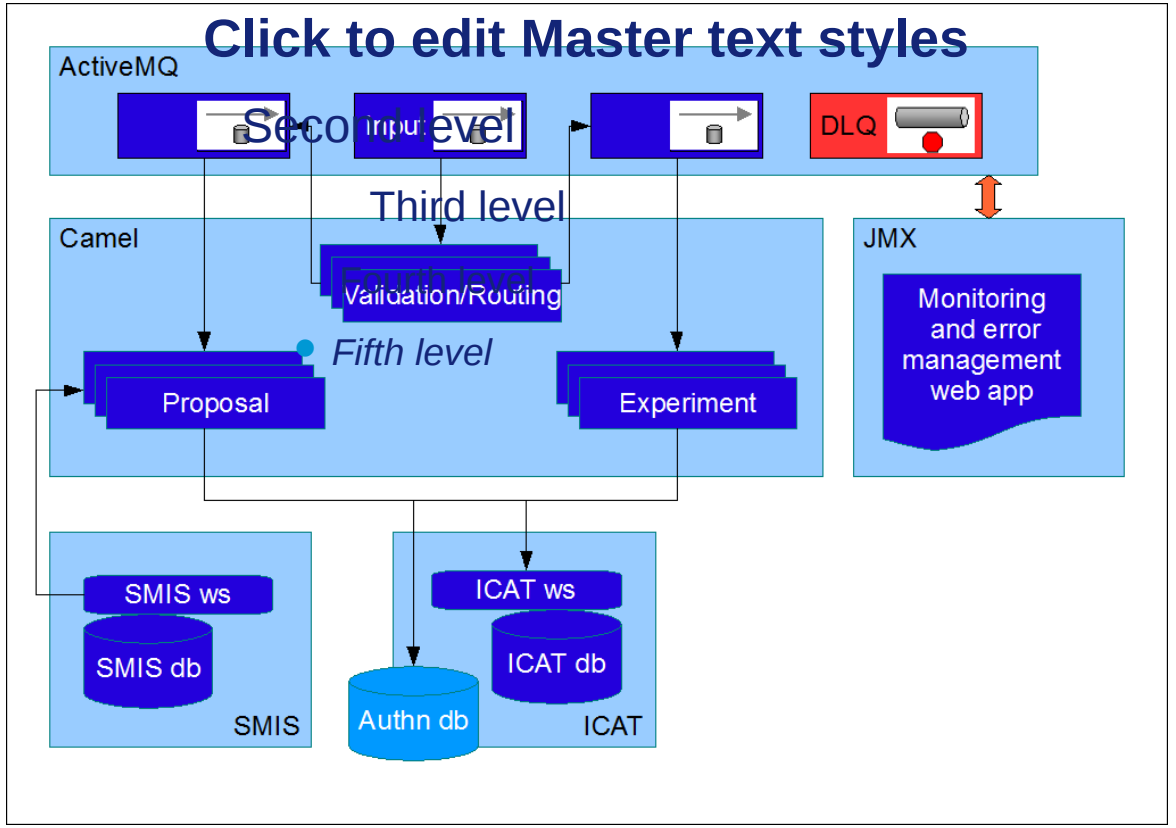


Proposed implementation

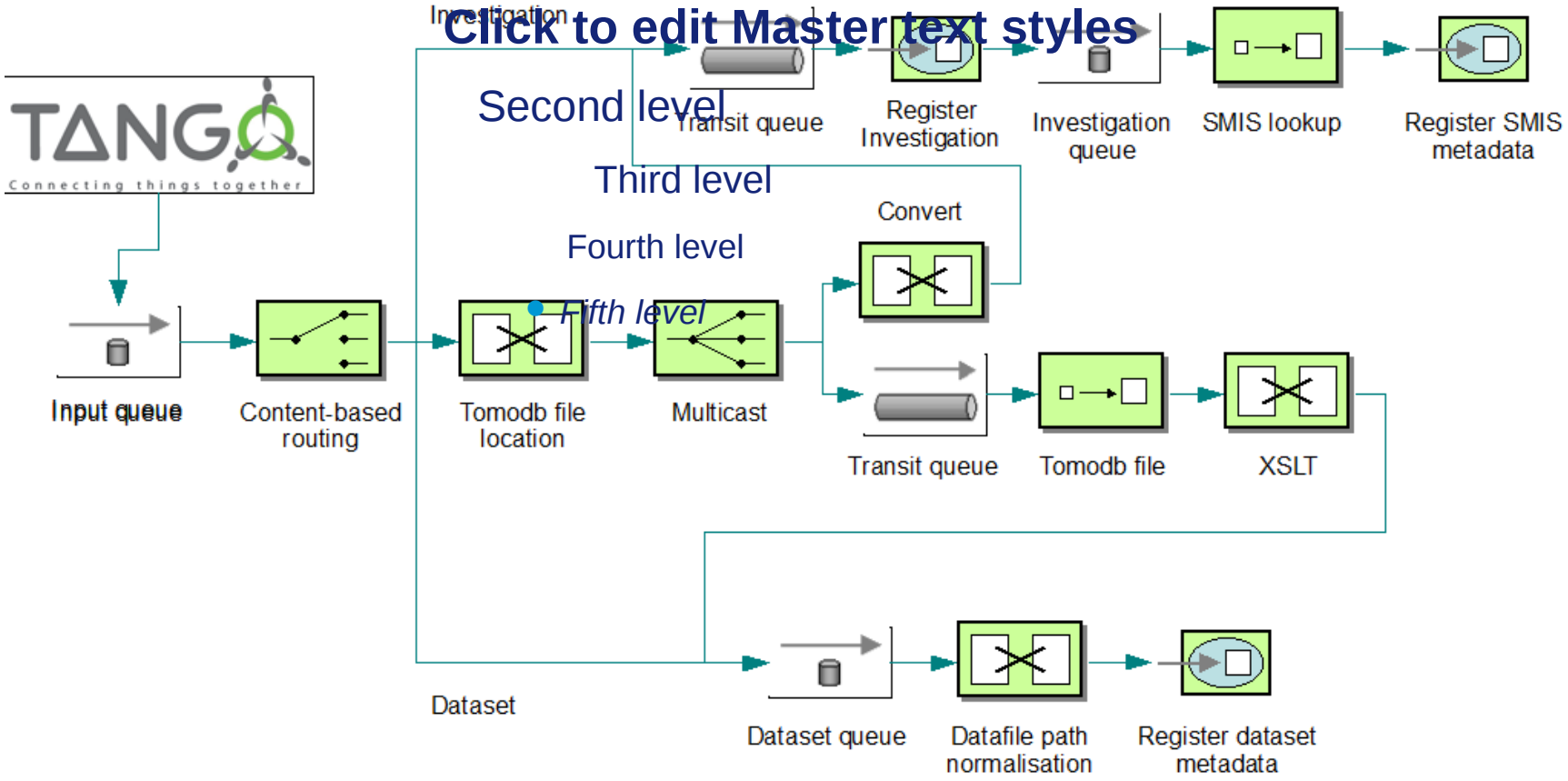


1 stateful TANGO device on each beamline

1 stateless queuing and processing system to bind them all



Processing details



Risks and benefits

Asynchronous processing

- Non-blocking TANGO devices (fire and forget)
- Queues serve as buffer in case of peak activity or slow/dead process

Data integrity

- Message persistence
- Transacted processing
- Error management, redelivery policy

Scalability

- Concurrent processing with configurable thread pools
- Several instances of queues and/or routes can be deployed

Input queue is a single point of failure

- Failover configuration / cluster deployment
- Local hdf5 file for offline processing

Advantages of apache camel

Most features already implemented

- ✓ Concurrent processing
- ✓ Transactions
- ✓ Automatic message translation (marshall/unmarshall, conversion)
- ✓ Logging / message tracing
- ✓ Error management

Technology agnostic

- ✓ Implements well-defined Enterprise Integration Patterns
- ✓ Independent of queuing system, transport technology, …

Very well integrated

- ✓ DSL for Java, Spring, Scala, Groovy, Annotations, Blueprint, …
- ✓ Options for deployment (standalone, EJB container, OSGi container)
- ✓ Pure Maven (archetypes, plugins), JUnit support, Spring integration

Current status

Prototype development

- Embedded Apache ActiveMQ
- Standalone Java application (using Spring)

Project metrics

- c.a. 1 month in the making
- c.a. 600 LoC (250 tests, 200 ICAT client, 150 data objects, beans)
- c.a. 800 lines of XML (incl. blank lines) (600 test, 100 configuration, 100 XSLT)
- Stored on ESRF gitlab

Not done yet

- ❑ SMIS integration
- ❑ Monitoring and error management interface

Authentication with Umbrella

Reusing SMIS mechanism already in place

Next steps

Prototype almost ready for testing with TANGO device

Validation of the design

Testing and performance analysis

Stress testing, failure recovery

Dimensioning of the different components

Data volume for queues, number of concurrent processors for each steps

1 year = 1200 experimental session, 4800 visits, 1800 publications

on 30 public beamlines (+10 CRG)

Live testing on one volunteer beamline

We still need to define

- ESRF data policy
- Infrastructure needed for deployment
- Maintenance and support policy

The European Synchrotron

Thank you for your attention

Issues with ICAT

Solved issues

- No migration script for Derby db (no python driver), no migration of rules
  - Python code (partially) ported to Java, rules recreated
- Some modification of the package required for TopCAT
  - To change the JPA log level
- Problem with TopCAT behind a reverse proxy (GWT issue)
  - All .gwt.rpc files need to be copied (address translation ignored)

Current issues

- ICE 1.0.0 not compatible with ICAT 4.3.x
  - Need to be updated at the same time than ICAT if considered as an admin module
- IDS not fully suitable for ESRF setup
  - We need more operations delegated to the plugin, including logic for available/not available, direct/archived, path reconstruction, …
  - The plugin currently needs a special packaging

Other issues

Issues with clients

- DAWN ICAT plugin
  - Only supports LDAP as authentication method
  - Cannot configure new ICAT repositories
- Most Python tools need at least Python 2.6
  - Version 2.4 stated in documentation
- XML ingestion tool does not work with ICAT 4.3.x
  - Would need a complete rebuild as it uses a modified 4.2 client

Issues with other ICATs

- Problems connecting to external ICAT on port 8181 (solved)
- Problems connecting to ILL using ill.eu URL (ill.fr works)
- Failed to connect to ISIS during SV 7 (investigation ongoing)
  - **Likely linked to our proxy and firewall configuration too**

HZB Helmholtz Zentrum Berlin

HZB

Report on ICAT Deployment

Rolf Krahl

ICAT Meeting, Dublin, Mar 2014

ICAT Deployment

ICAT deployment is in experimental stage at HZB.

- Virtual machine `icat.helmholtz-berlin.de`
- openSUSE 13.1 (x86_64)
- OpenJDK 1.7.0
- GlassFish Server Open Source Edition 4.0
- ICAT components:
  - authn_simple 1.0.0 (users: root, useroffice, idsreader)
  - authn_ldap 1.1.0
  - ICAT 4.3.2
  - ids.storage_file 1.0.0
  - ids.server 1.0.1
  - TopCAT 1.11.0
- Use central Oracle database as backend
- Access is proxied through Apache using mod_proxy_http

GATE and LDAP

Users:

- ICAT users are authenticated by the HZB LDAP server.
- Users are imported from LDAP into ICAT by a Python script.

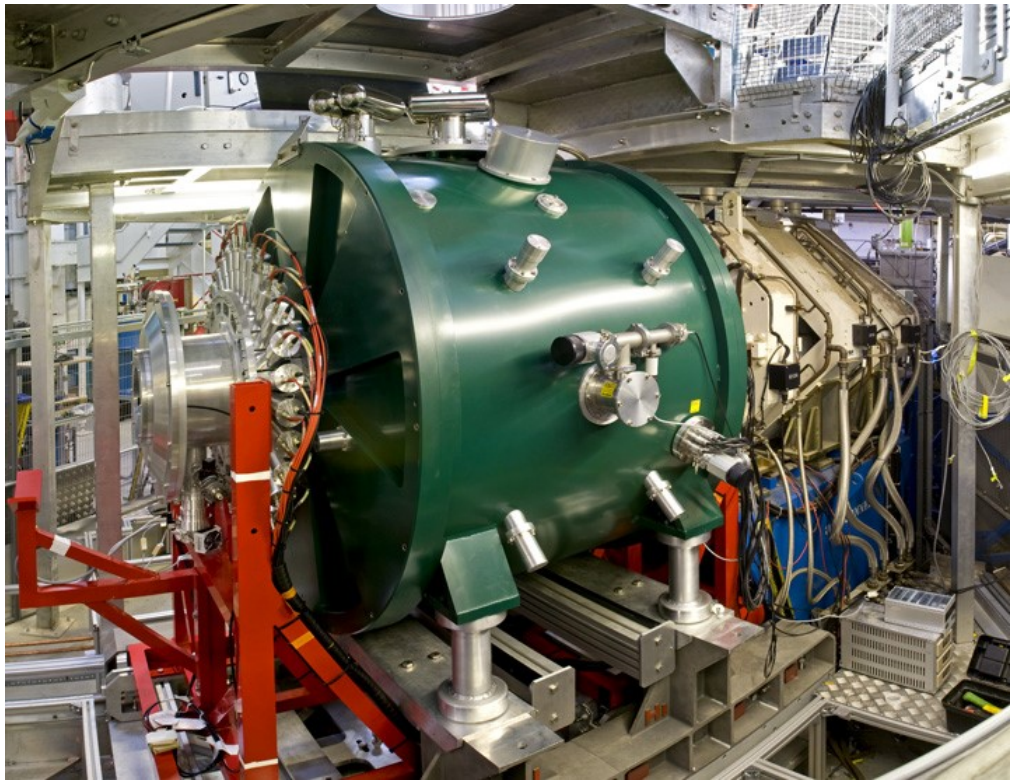
HZB user portal GATE:

- GATE is the HZB user portal and proposal submission system.
- Import instruments and proposals (as investigations) from GATE into ICAT using Python scripts.
- Problem: GATE has its own user base. Some GATE users do have an HZB account (LDAP), but not all.
- Plan: GATE users should get an HZB account routinely if their proposal is accepted, But workflow for this is not settled yet.

Storage

- Systematic storage of experimental data is not yet operational for all instruments at HZB.
- Purchase of storage hardware is planned.
- E.g. the physical storage that IDS should be deployed on, does not exist yet.
- A customized IDS storage plugin might be needed before going into production.
- Question: How to keep data on storage and datafile objects in ICAT in sync?

ICAT at ISIS



Tom Griffin, STFC ISIS Facility
ICAT Workshop
Dublin

March 2014

tom.griffin@stfc.ac.uk

Science & Technology Facilities Council
ISIS

Status

- ICAT 3.3 : production for 5+ years
- ICAT 4.3 : production - parallel
- ICAT Data server: production
- Authentication : Full user office plugin
- Large Rule set - (mostly) implements our data policy

Review – Past Issues

- Rules system now supports most of what we need
- IDS reference implementation required only minor changes
- Investigation <-> Instrument change painful, but useful
- SampleType constraint changed
- TopCAT single sign on

Current Issues

- Network reliability at RAL
  - Due to network topology ISIS staff see this as an ICAT problem
- Periodic server hangs
  - Not always sure why (memory, load, ?)
  - Multi node setup for balancing and reliability

1 year plan – Lund

- Go live with 4.2 / 4.3 / 4.4 ✓
- Mantid integration with 4.x ✓
- Switch off 3.3 ✗
- Upload of processed data
  - from Mantid ✓
  - From TopCAT ✗

<1 year Requirements - Lund

- Permissions control in TopCAT ✗
  - ICAT auth rules to support this ?
- GUI tweaks to TopCAT ✓
- IDS reference implementation ✓
- Data upload from Mantid ✓

2014/15 plan

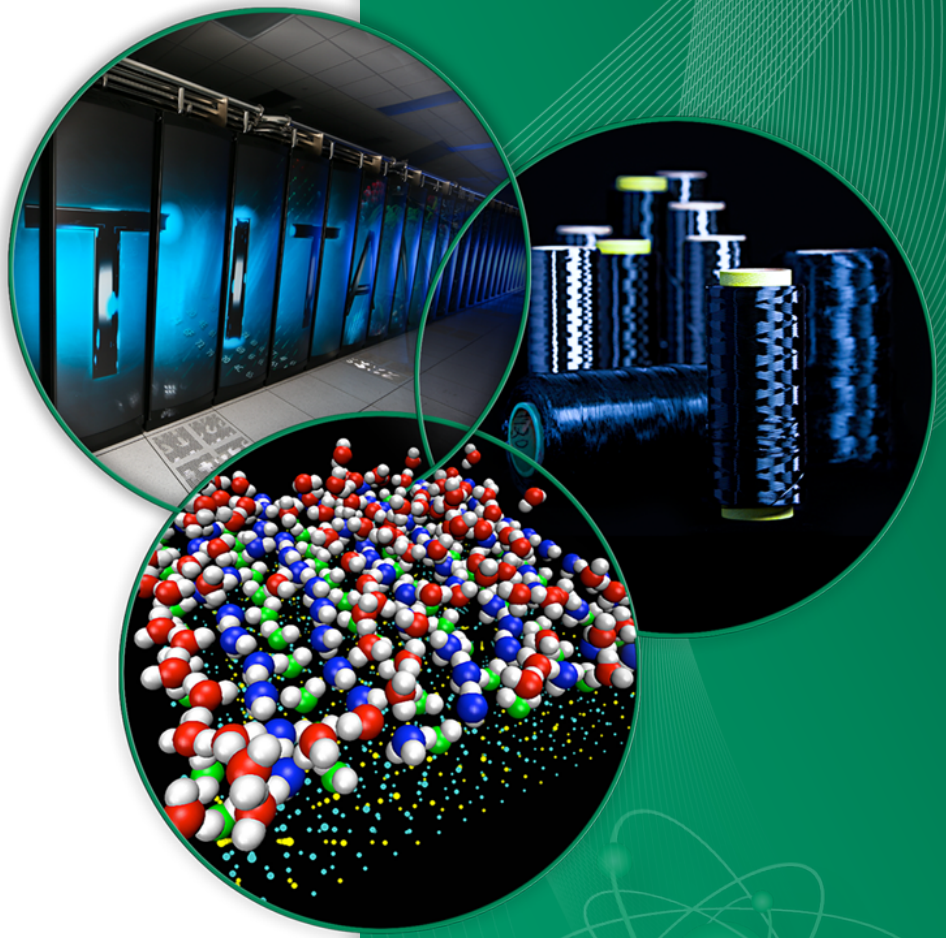
- Continue verification testing of 4.3
- Continue Mantid integration with 4.3
- Switch off 3.3
- ----------
- New interfaces (see later presentation)
- ISIS 'Scientific Computing Strategy'

Requirements

- See the roadmap (?)
- Further enhancements to rules
- GUI changes to TopCAT….maybe more
- Stability – **no** *breaking* changes
- Performance generally very good
  - Function of ingest server, database and network
  - Some rules can cause ‘slow’ queries

Questions...

ICAT4 Experience at ORNL/SNS



Shelly Ren

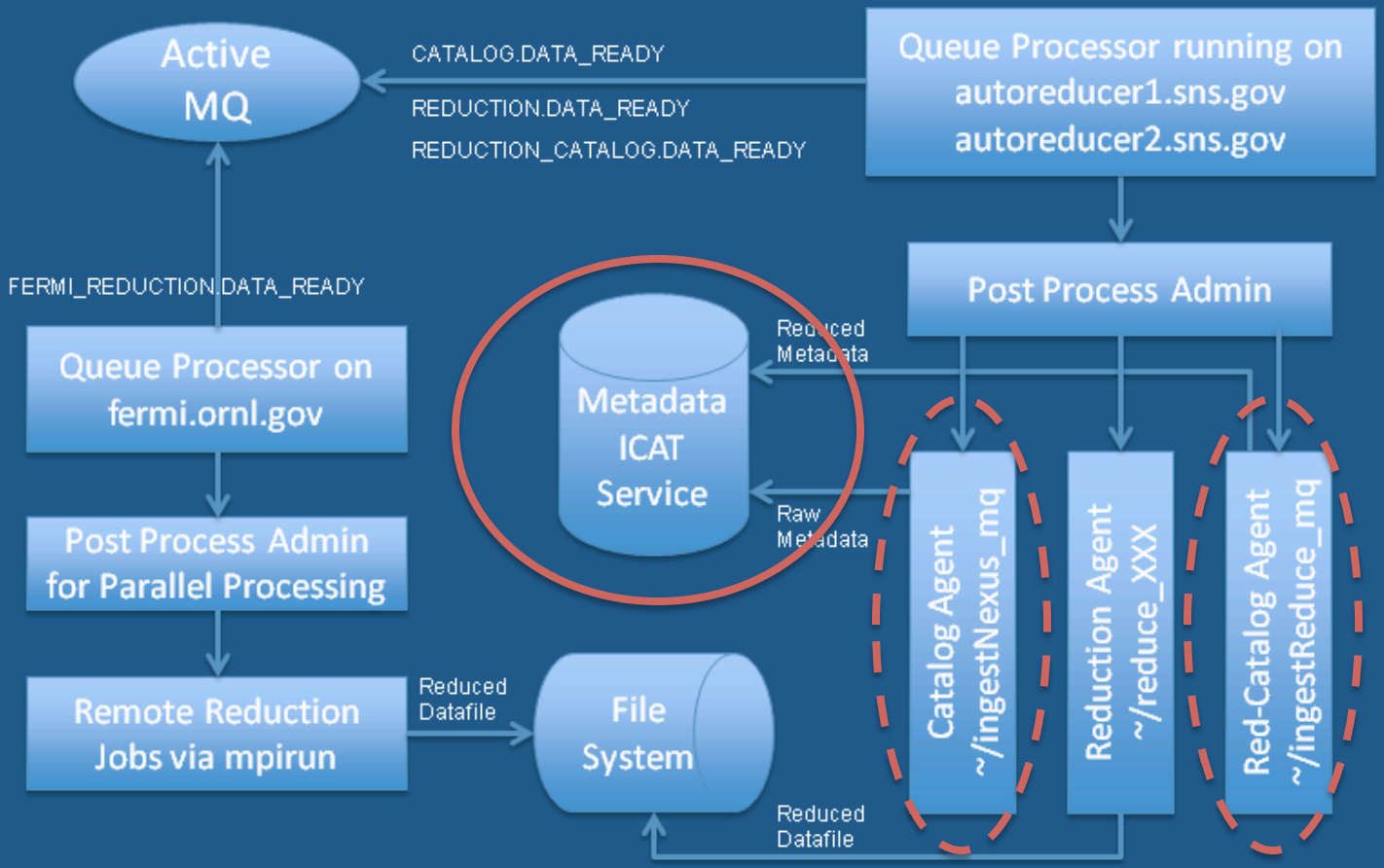
Scientific Information Systems

March 24, 2014

How is ICAT used at SNS today?

- Fully integrated with SNS post processing
  - Auto catalog of raw experiment data
  - Auto catalog of auto reduced data
- Web service to serve live experiment monitor
  - Up to date information about a proposal
  - Up to date information about runs of a proposal
- Web service to serve for search engines (findnexus, finddata, Filefinder):
  - Summary of a proposal
  - Metadata of an experiment/run
  - Location of experiment data

Automated Data Catalog and Data Reduction



Experiment Monitor

SEQ IPTS-10633

3qr | admin | logout

home › seq › ipts-10633 live monitoring: status | runs | PVs

List of SEQ runs for IPTS-10633:

Run	Created on	Status
48832	March 23, 2014, 6:07 p.m.	complete
48831	March 23, 2014, 5:09 p.m.	complete
48830	March 23, 2014, 4:09 p.m.	complete
48829	March 23, 2014, 3:11 p.m.	complete
48828	March 23, 2014, 2:07 p.m.	complete
48827	March 23, 2014, 1:51 p.m.	complete
48826	March 23, 2014, 12:38 p.m.	complete
48825	March 23, 2014, 11:35 a.m.	complete
48824	March 23, 2014, 10:07 a.m.	complete
48823	March 23, 2014, 9:07 a.m.	complete
48822	March 23, 2014, 7:46 a.m.	complete
48821	March 23, 2014, 6:45 a.m.	complete
48820	March 23, 2014, 5:46 a.m.	complete
48819	March 23, 2014, 4:44 a.m.	complete
48818	March 23, 2014, 3:46 a.m.	complete
48817	March 23, 2014, 2:43 a.m.	complete
48816	March 23, 2014, 1:44 a.m.	complete
48815	March 23, 2014, 12:43 a.m.	complete
48814	March 22, 2014, 11:41 p.m.	complete
48813	March 22, 2014, 10:40 p.m.	complete
48812	March 22, 2014, 9:41 p.m.	complete

21 of 223 runs shown – show all runs

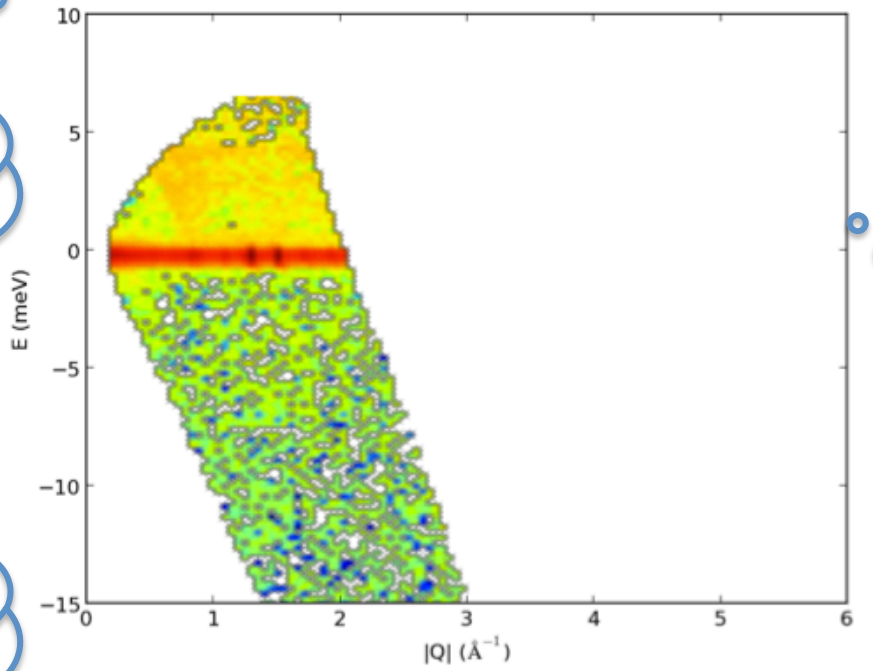
Experiment status

List of experiments

Experiment Data Display

Run title: Run=42172,Ba3NiSb2O9, powder,T=1.51 K,7.5 meV,Fermi=180 Hz><
Run start: March 23, 2014, 11:18 p.m.
Run end: March 24, 2014, 6:23 a.m.

Experiment parameters



Plot of auto reduced data

Experiment data location

Data files:

- /SNS/HYS/IPTS-10498/adara/HYS_42172.adara
- /SNS/HYS/IPTS-10498/nexus/HYS_42172.nxs.h5

Reduced files:

- /SNS/HYS/IPTS-10498/shared/autoreduce/4pixel/HYS_42172_4pixel.nxspe
- /SNS/HYS/IPTS-10498/shared/autoreduce/msk_tube/HYS_42172_msk_tube_spe.nxs
- /SNS/HYS/IPTS-10498/shared/autoreduce/reduction_log/HYS_42172.nxs.h5.log
- /SNS/HYS/IPTS-10498/shared/autoreduce/4pixel/HYS_42172_4pixel_spe.nxs
- /SNS/HYS/IPTS-10498/shared/autoreduce/msk_tube/HYS_42172_msk_tube.nxspe
- /SNS/HYS/IPTS-10498/shared/autoreduce/msk_tube
  /HYS_42172_msk_tube_spe.nxs.png

More about ICAT tomorrow

- Web service to encapsulate database/objects from application users
  - Integrated user or guest portal
  - Experiment hall monitors
- ICAT driven fully automate data reduction
  - Auto reduce via the latest calibration, characterization runs
  - Auto reduce via strategies defined in ICAT
- Statistical distribution of experiment parameters, aka process variables
- ICAT assisted experiment design

ICAT Installation, Performance, Concerns

- Plan to roll ICAT 4.3 in the summer break
- Still need to work on and test out Rules
- ICAT 4.2 seems to run smoothly in production
- Potential performance concerns: load and search of large volume of dataset and datafile parameters

HZB Helmholtz Zentrum Berlin

python-icat

A Library for Writing ICAT Clients in Python

Rolf Krahl

ICAT Meeting, Dublin, Mar 2014

Introduction

- SOAP is used as the access protocol for ICAT.
- Clients exist for different programming languages, including Java and Python.
- The most popular SOAP library for Python is Suds.
- python-icat aims to make writing ICAT clients with Python simpler.

Design Goals

python-icat is build on top of Suds.

Goals

- Keep the general structure and flexibility of Suds.
- Simplify things where possible.
- Try to remove annoying details.
- Make use object oriented design.

A typical python-icat program might be mistaken for a generic Suds program at first glance. It's just somewhat simpler.

Example: Add a Datafile

Using plain Suds

```
dataset = client.service.search(sessionId,
                "Dataset[name='e201215']")[0]
format = client.service.search(sessionId,
                "DatafileFormat[name='NeXus']")[0]
datafile = client.factory.create("datafile")
datafile.dataset = dataset
datafile.datafileFormat = format
datafile.name = "e201215-7.nxs"
datafile.id = client.service.create(sessionId, datafile)
```

Example: Add a Datafile

Using plain Suds

```
dataset = client.service.search(sessionId,
                "Dataset[name='e201215']")[0]
format = client.service.search(sessionId,
                "DatafileFormat[name='NeXus']")[0]
datafile = client.factory.create("datafile")
datafile.dataset = dataset
datafile.datafileFormat = format
datafile.name = "e201215-7.nxs"
datafile.id = client.service.create(sessionId, datafile)
```

Using python-icat

```
dataset = client.search("Dataset[name='e201215']")[0]
format = client.search("DatafileFormat[name='NeXus']")[0]
datafile = client.new("datafile")
datafile.dataset = dataset
datafile.datafileFormat = format
datafile.name = "e201215-7.nxs"
datafile.create()
```

Example: Add a Datafile

Or even:

Using python-icat

```
dataset = client.search("Dataset[name='e201215']")[0]
format = client.search("DatafileFormat[name='NeXus']")[0]
client.new("datafile",
           dataset=dataset,
           datafileFormat=format,
           name="e201215-7.nxs").create()
```

Notes

- `client.new(...)` creates a new ICAT entity objects. Use this in place of `client.factory.create(...)`.
- `client.new(...)` optionally accepts keyword/value arguments to set attributes.
- ICAT API methods are defined as methods in the python-icat client. Replace `client.service.<method>(...)` by `client.<method>(...)`.
- Don't care about the session id, the python-icat client remembers it and adds it to the ICAT method calls as needed.
- ICAT entity objects have their own methods: e.g. `datafile.create()`.

Example: Add Keywords to an Investigation

Using plain Suds

```
investigation = client.service.search(sessionId,
        "Investigation[name='2010-E2-0489-1']")[0]
keywords = []
for k in ["Foo", "Bar", "Baz"]:
    keyword = client.factory.create("keyword")
    keyword.name = k
    keyword.investigation = investigation
    keywords.append(keyword)
client.service.createMany(sessionId, keywords)
```

Example: Add Keywords to an Investigation

Using plain Suds

```
investigation = client.service.search(sessionId,
        "Investigation[name='2010-E2-0489-1']")[0]
keywords = []
for k in ["Foo", "Bar", "Baz"]:
    keyword = client.factory.create("keyword")
    keyword.name = k
    keyword.investigation = investigation
    keywords.append(keyword)
client.service.createMany(sessionId, keywords)
```

Using python-icat

```
investigation = client.search(
        "Investigation[name='2010-E2-0489-1']")[0]
investigation.addKeywords(["Foo", "Bar", "Baz"])
```

Example: Create a Group

Using plain Suds

```
users = [ jbotu , jdoe , nbour ]
group = client.factory.create("group")
group.name = "investigation_42_reader"
group.id = client.service.create(sessionId , group)
ugs = []
for u in users:
    ug = client.factory.create("userGroup")
    ug.user = u
    ug.group = group
    ugs.append(ug)
client.service.createMany(sessionId , ugs)
```

Example: Create a Group

Using plain Suds

```
users = [ jbotu , jdoe , nbour ]
group = client.factory.create("group")
group.name = "investigation_42_reader"
group.id = client.service.create(sessionId , group)
ugs = []
for u in users:
    ug = client.factory.create("userGroup")
    ug.user = u
    ug.group = group
    ugs.append(ug)
client.service.createMany(sessionId , ugs)
```

Using python-icat

```
users = [ jbotu , jdoe , nbour ]
group = client.createGroup("investigation_42_reader", users)
```

Example: Login

Using plain Suds

```
client = suds.client.Client(url)
credentials = client.factory.create("credentials")
credentials.entry.append(
        [ { 'key':'username', 'value':username },
          { 'key':'password', 'value':password } ])
sessionId = client.service.login(auth, credentials)

# ...

client.service.logout(sessionId)
```

Example: Login

Using plain Suds

```
client = suds.client.Client(url)
credentials = client.factory.create("credentials")
credentials.entry.append(
        [ { 'key':'username', 'value':username },
          { 'key':'password', 'value':password } ])
sessionId = client.service.login(auth, credentials)

# ...

client.service.logout(sessionId)
```

Using python-icat

```
client = icat.client.Client(url)
credentials = { 'username':username, 'password':password }
client.login(auth, credentials)
```

ICAT Version

- Drawback of python-icat: it depends on the ICAT version.
- When the ICAT API changes, the library needs to get adapted to the new version.
- Currently supported: 4.2.* and 4.3.*, the API version is checked automatically.
- A module `icat.icatcheck` tests compatibility helps to adapt the library to new versions.
- Advantage: some incompatibities between ICAT versions are handled by python-icat and hidden from the application.

Example: Add Instrument to an Investigation

Using plain Suds

```
investigation = client.service.search(sessionId,
        "Investigation INCLUDE 1 [name='2010-E2-0489-1']")[0]
instrument = client.service.search(sessionId,
        "Instrument[name='HIKE']")[0]
if client.service.getApiVersion() < '4.3.0':
    investigation.instrument = instrument
    client.service.update(sessionId, investigation)
else:
    ii = client.factory.create('investigationInstrument')
    ii.investigation = investigation
    ii.instrument = instrument
    client.service.create(sessionId, ii)
```

Example: Add Instrument to an Investigation

Using plain Suds

```
investigation = client.service.search(sessionId,
        "Investigation INCLUDE 1 [name='2010-E2-0489-1']")[0]
instrument = client.service.search(sessionId,
        "Instrument[name='HIKE']")[0]
if client.service.getApiVersion() < '4.3.0':
    investigation.instrument = instrument
    client.service.update(sessionId, investigation)
else:
    ii = client.factory.create('investigationInstrument')
    ii.investigation = investigation
    ii.instrument = instrument
    client.service.create(sessionId, ii)
```

Using python-icat

```
investigation = client.search(
        "Investigation[name='2010-E2-0489-1']")[0]
instrument = client.search("Instrument[name='HIKE']")[0]
investigation.addInstrument(instrument)
```

Config

- A typical ICAT client always needs the same set of command line arguments: URL of the ICAT service, authentication plugin name, username, and password.
- A module `icat.config` takes care of this: it defines the command line arguments.
- Configuration options may be set via command line arguments, environment variables, configuration files, and default values (in this order, first match wins). The password may also be read from interactive keyboard input.
- Of course, a program may define additional custom arguments.

Example: Config

Using plain Suds

```
url = "https://" + sys.argv[1] + ":" + sys.argv[2] \
    + "/ICATService/ICAT?wsdl"
auth = sys.argv[3]
username = sys.argv[5]
password = sys.argv[7]
client = suds.client.Client(url)
credentials = client.factory.create("credentials")
credentials.entry.append(
        [ { 'key':'username', 'value':username },
          { 'key':'password', 'value':password } ])
sessionId = client.service.login(auth, credentials)
```

Example: Config

Using plain Suds

```
url = "https://" + sys.argv[1] + ":" + sys.argv[2] \
    + "/ICATService/ICAT?wsdl"
auth = sys.argv[3]
username = sys.argv[5]
password = sys.argv[7]
client = suds.client.Client(url)
credentials = client.factory.create("credentials")
credentials.entry.append(
        [ { 'key':'username', 'value':username },
          { 'key':'password', 'value':password } ])
sessionId = client.service.login(auth, credentials)
```

Using python-icat

```
config = icat.config.Config()
conf = config.getconfig()
client = icat.Client(conf.url, **conf.client_kwargs)
client.login(conf.auth, conf.credentials)
```

Default Command Line Arguments

```
usage: login-icat-config.py [options]

optional arguments:
  -h, --help            show this help message and exit
  -c CONFIGFILE, --configfile CONFIGFILE
                        config file
  -s SECTION, --configsection SECTION
                        section in the config file
  -w URL, --url URL     URL to the web service description
  --http-proxy HTTP_PROXY
                        proxy to use for http requests
  --https-proxy HTTPS_PROXY
                        proxy to use for https requests
  -a AUTH, --auth AUTH  authentication plugin
  -u USERNAME, --user USERNAME
                        username
  -p PASSWORD, --pass PASSWORD
                        password
  -P, --prompt-pass     prompt for the password
```

Example: Exception Handling

Using plain Suds

```
try:
    sessionId = client.service.login(auth, credentials)
except suds.WebFault as e:
    if e.fault.detail.IcatException.type == 'SESSION':
        print "Login failed: %s" % e
    else:
        raise
```

Example: Exception Handling

Using plain Suds

```
try:
    sessionId = client.service.login(auth, credentials)
except suds.WebFault as e:
    if e.fault.detail.IcatException.type == 'SESSION':
        print "Login failed: %s" % e
    else:
        raise
```

Using python-icat

```
try:
    client.login(conf.auth, conf.credentials)
except ICATSessionError as e:
    print "Login failed: %s" % e
```

Example: Searching

Using plain Suds

```
searchres = client.service.search(sessionId, "Facility")
if len(searchres) != 1:
    raise RuntimeError("Expected to find one facility")
else:
    facility = searchres[0]
```

Example: Searching

Using plain Suds

```
searchres = client.service.search(sessionId, "Facility")
if len(searchres) != 1:
    raise RuntimeError("Expected to find one facility")
else:
    facility = searchres[0]
```

Using python-icat

```
facility = client.assertedSearch("Facility")[0]
```

Example: Searching

Using plain Suds

```
searchres = client.service.search(sessionId, "Facility")
if len(searchres) != 1:
    raise RuntimeError("Expected to find one facility")
else:
    facility = searchres[0]
```

Using python-icat

```
facility = client.assertedSearch("Facility")[0]
```

Using python-icat (more)

```
# Assert there is at least one Investigation
investigation = client.assertedSearch("Investigation",
                                      assertmax=None)[0]
# Assert there is at most one Instrument
res = client.assertedSearch("Instrument", assertmin=0)
```

Misc

- A module `icat.cgi` helps writing CGI scripts. It does session management: the ICAT session Id is set as a cookie in the user's browser.
- Methods `Entity.getUniqueKey()` and `Client.searchUniqueKey()` to create a unique object identifier and to search for the object corresponding to an identifier respectively.
- Example scripts `icatdump.py` and `icatrestore.py` that dump the whole content of an ICAT to a file (YAML) and restore it from the dump file respectively.

Internals

- `icat.client.Client` is a `suds.client.Client`. Everything you can do with a Suds client, you can do with a python-icat client.
- The ICAT entity objects created by `client.new(...)` or returned by a search live in a hierarchy of classes based on `icat.entity.Entity`.
- The ICAT entity objects mimic very closely the behavior of corresponding Suds objects. They are converted transparently from and to Suds objects as appropriate.

System Requirements and Download

System Requirements

- Python 2.6 or newer (Python 2.6 requires a patch).
- Suds, either 0.4 or jurko fork, the latter is recommended.
- argparse (in system library in Python 2.7 or newer).
- The example scripts use PyYAML, but this is not needed to use the library itself.

Download

- python-icat 0.4.0 available at `http://code.google.com/p/icatproject/wiki/PythonIcat`
- BSD license.

System Requirements and Download

System Requirements

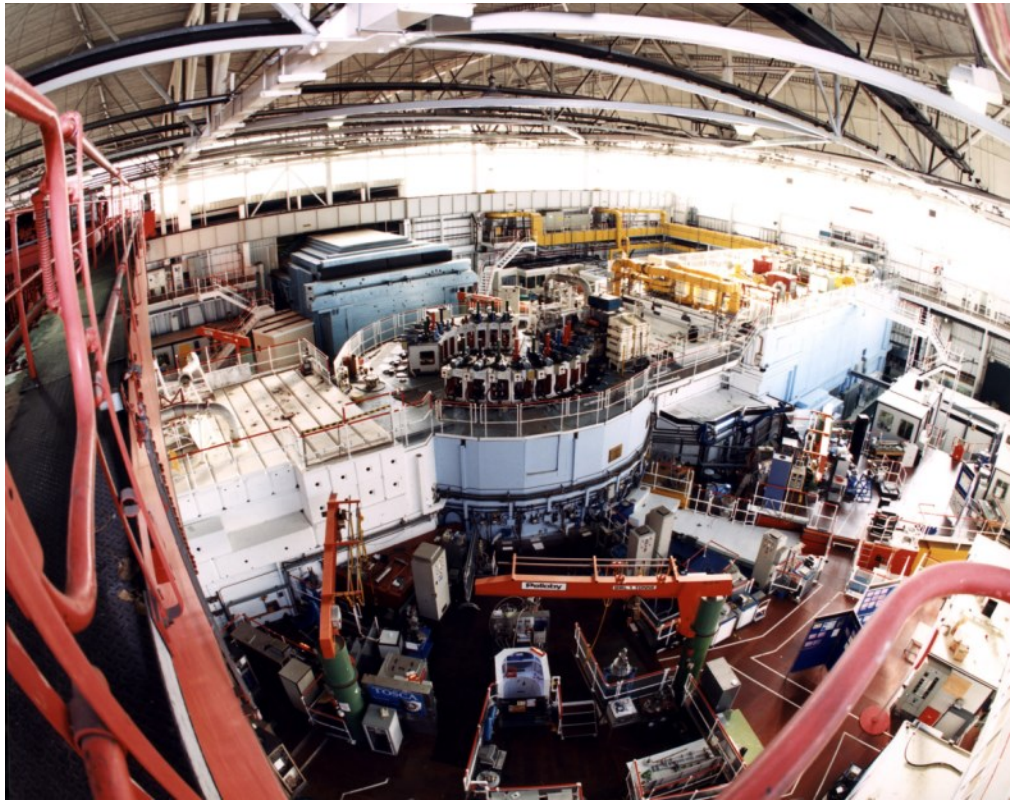
- Python 2.6 or newer (Python 2.6 requires a patch).
- Suds, either 0.4 or jurko fork, the latter is recommended.
- argparse (in system library in Python 2.7 or newer).
- The example scripts use PyYAML, but this is not needed to use the library itself.

Download

- python-icat 0.4.0 available at `http://code.google.com/p/icatproject/wiki/PythonIcat`
- BSD license.

Thank you for your attention! Questions?

A customisable domain specific ICAT interface



Frazer Barnsley, Tom Griffin
ISIS Facility

Elettra, Trieste

November 2013

tom.griffin@stfc.ac.uk

Science & Technology Facilities Council
ISIS

Contents

- What?
- Why?
- Who?
- How?
- Where?
- When?

Why?

- TopCAT is very generic, so does not provide information to our users in a domain specific way
- This is fine for some things
- Some domains have specific workflows and standard ways of working
- Can we have ‘custom’ portals for them, without too much work?

Who?

- Customers
  - ISIS Disordered Materials Group
  - Inelastic spectra database
- Creators
  - ISIS Computing group
  - Me
  - Frazer Barnsley (graduate secondment)

- Will be added to contrib

How?

- Uses ICAT as the database and IDS to store and serve data
- Configuration file defines
  - Names
  - Numbers of samples
  - Parameters (Investigation, sample, dataset, datafile)
  - Datafiles (uploads)
  - Tooltips
  - Friendly names
- Creates ICAT entries (including ParameterTypes)
- Grails – Groovy on Rails
- Iterative development with science groups
- Not much code. Lots of nice things for free
- Tomcat, Glassfish, Jetty, etc

Where?

- Currently on a development server (RAL only)

Live Demo……

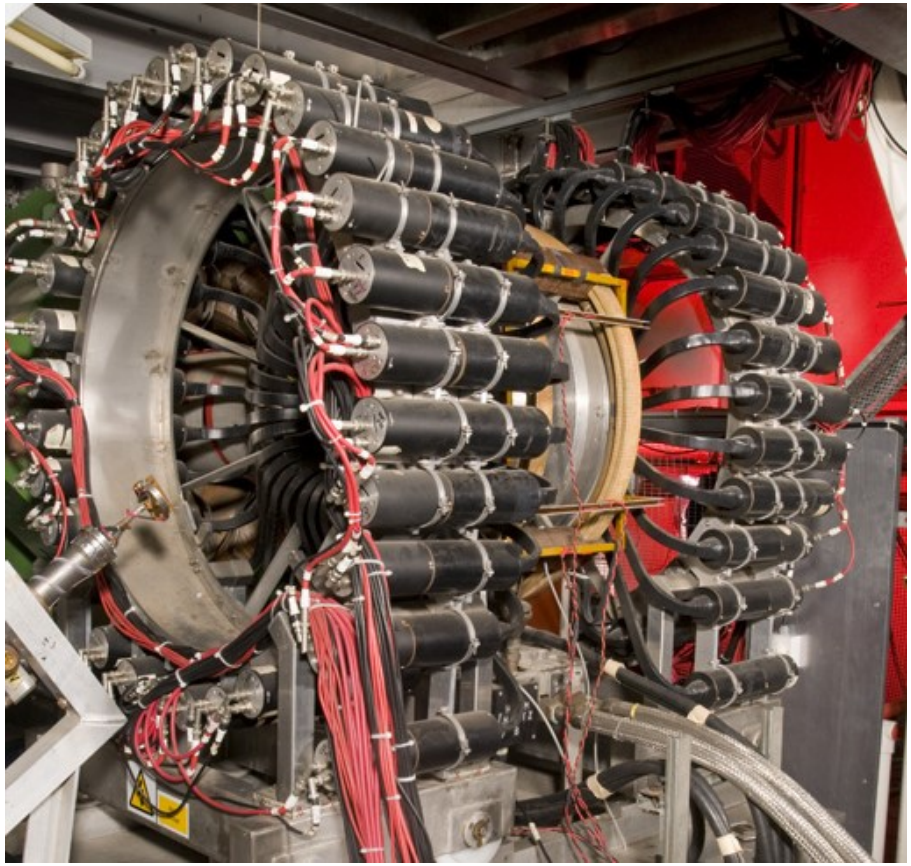
http://130.246.49.212:8080/SimpleICATBrowser/

When?

- Next release to customers – end of this week
- Must be finished by end of November

Questions...

Authorisation Rules at ISIS



Tom Griffin, STFC ISIS Facility
ICAT Workshop
Dublin

March 2014

tom.griffin@stfc.ac.uk

Science & Technology Facilities Council
ISIS

Introduction

- Implements a defined data policy
  - PaNdata policy based.
- Public Tables
- Public Steps
- Rules
- UserGroups

Background

- Rules – by default everything is closed.
- ISIS data policy requires >100 rules…..still not finished
- Are (now) sanity checked on creation
- Required at every level for **direct** access (see PublicSteps)
- Can be checked:

```
isAccessAllowed(String sessionId,
EntityBaseBean bean, AccessType
accessType)
```

Public Tables

- For read only access to open tables
- Cached
- Instrument, Application, DatasetType, InvestigationType etc
- Can be used for link tables : DataCollectionDatafile, InvestigationUser

Public Tables

```
List<String> publicTables = new ArrayList<>();
publicTables.add("Application");
publicTables.add("DatafileFormat");
publicTables.add("DatasetType");
publicTables.add("Facility");
...<cut>...

 publicTables.add("DataCollectionDatafile");
 publicTables.add("DataCollectionDataset");
       ...<cut>...

 List<EntityBaseBean> publicRules = new ArrayList<>();
 for (String publicTableName : publicTables) {
         Rule publicRule = new Rule();
         publicRule.what = publicTableName;
         publicRule.crudFlags = "R";
         publicRules.add(publicRule);
}
port.createMany(sessionId, publicRules);
```

Public Steps

- Allow access to a related object (attribute)
- ONLY used in INCLUDE processing.
- Offer a significant speedup on INCLUDE
- Thing -> ThingParameter
- Investigation -> samples, publications, users, investigationInstrument etc

Public Steps

```
String[] parameterBasedPublicSteps = new
String[]{"Investigation", "Dataset", "Datafile",
                "Sample"};
for (String table : parameterBasedPublicSteps)
{
        PublicStep paramPublicStep = new PublicStep();
        paramPublicStep.origin = table;
        paramPublicStep.field = "parameters";
        publicSteps.add(paramPublicStep);
}
```

Public Steps

```
String[] publicStepsFromInvestigation = new
String[]{"samples", "publications", "shifts",
"investigationUsers", "keywords",
"investigationInstruments"};

for (String step : publicStepsFromInvestigation)
{
        PublicStep invToSomething = new PublicStep();
        invToSomething.origin = "Investigation";
        invToSomething.field = step;
        publicSteps.add(invToSomething);
}
```

Rule Structure

- Administrators
- Safe Admin (read all)
- [Raw] Data Ingestor
- Instrument Scientists
- Investigators
- Disordered Materials Database
- DOI service
- Unembargoed data

Administrators

- Easy

```
List<String> allTables = port.getEntityNames();
for (String table : allTables)
{
      Rule rule = new Rule();
      rule.grouping = facilityAdmins;
      rule.crudFlags = "CRUD";
      rule.what = table;
      port.create(sessionId, rule);
}
```

Other super groups

- Safe admins – same rules, no ‘CUD’
- Data Ingestors: ‘CRU’, fewer tables

Instrument Scientists

- Defines access relative to instruments

```
SELECT i FROM Investigation i
JOIN i.investigationInstruments ii
JOIN ii.instrument inst
JOIN inst.instrumentScientists instSci
JOIN instSci.user u
WHERE u.name = :user
```

Investigation, Dataset, Datafile, Sample, SampleType + 4x Parameters

Instrument Scientists

```
SELECT df FROM Datafile df
JOIN df.dataset d
JOIN d.investigation i
JOIN i.investigationInstruments ii
JOIN ii.instrument inst
JOIN inst.instrumentScientists instSci
JOIN instSci.user u
WHERE d.name='Default'
AND u.name = :user";
```

Investigators

- Defines access relative to investigation role

```
DatafileParameter <-> Datafile <-> Dataset <->
Investigation <-> InvestigationUser  <-> User <->
User [name = :user]
```

```
Investigation, Dataset, Datafile, Sample,
SampleType + 4x Parameters
```

Disordered Materials Database

- Write (authenticated) and open read to domain specific ‘database’

```
SELECT df FROM Datafile df
JOIN df.dataset ds
JOIN ds.investigation i
JOIN i.type it
WHERE it.name Disordered Materials

Grouping = Disordered Materials Publishers
Access = CRUD
```

DOI Account

- Requires read access to generate landing pages

```
SELECT i FROM Investigation i WHERE i.doi IS NOT NULL

SELECT ds FROM Dataset ds WHERE ds.doi IS NOT NULL
```

- DOI creation runs as Data Ingestor

Unembargoed Data

- Read for all authenticated users

```
SELECT i FROM Investigation i WHERE i.releaseDate <
CURRENT_TIMESTAMP

SELECT dfp FROM DatafileParameter dfp
JOIN dfp.datafile df
JOIN df.dataset d
JOIN d.investigation i
WHERE d.name = 'Default'
AND i.releaseDate < CURRENT_TIMESTAMP

Investigation, Dataset, Datafile, Sample, SampleType +
4x Parameters
```

Complications

- Lock updates when DOI <> null
- Granting permissions
  - Create a group per investigation
  - What about finer grain?
  - Adding new users
  - What is ICAT, what is User office?
- Allowing partial updates
  - Investigation.release_date only
- Performance

Questions...

HZB Helmholtz Zentrum Berlin

Authorization Rules

Setup of Authorization Rules based on Groups of Users

Rolf Krahl

ICAT Meeting, Dublin, Mar 2014

Why?

Why using groups to setup the authorization rules?

- Authorization based on groups is very flexible.
- Easy to grant or to revoke permissions: simply add the user to or remove him from the corresponding group.
- Access policies may be individually defined for each investigation.
- Users may manage permissions themselves. They only need CRUD permission on UserGroup related to the corresponding group.
- InvestigationUser is also used for other purposes (e.g. TopCAT). One might wish to setup permissions independently from this.

Access Groups

For each investigation, create three access groups:

`investigation_<name>_writer`: Shall get CRUD permission on objects related to the investigation, such as Datafiles, Datasets, Samples, Keywords, Parameters and so on.

`investigation_<name>_reader`: Shall get R permission on objects related to the investigation.

`investigation_<name>_owner`: Shall get permission to manage access permissions on the investigation.

Here `<name>` is replaced by the investigation name.

Rules: Variant 1, per Investigation Rules

Simple way to setup access rules: create a set of rules for each investigation.

Rule

crudFlags: CRUD

what: `Datafile <-> Dataset <-> Investigation[name='<name>']`

grouping: `investigation_<name>_writer`

Rule

crudFlags: CRUD

what: `UserGroup <-> Grouping[name='investigation_<name>_writer']`

grouping: `investigation_<name>_owner`

Rules: Variant 1, per Investigation Rules

- Simple.
- Works fine in test installation.
- Drawback: excessive number of rules, 28 rules per investigation, more then a half of all objects in the ICAT are rules.
- Will this scale reasonably in a production size deployment?

Rules: Variant 2, Static Rules Based on "Magic Names"

Try to achieve the same result with a limited number of generic rules:

Rule

crudFlags: CRUD

what:
```
SELECT df FROM Datafile df
JOIN df.dataset ds JOIN ds.investigation i
JOIN Grouping g JOIN g.userGroups ug
JOIN ug.user u
WHERE g.name =
CONCAT('investigation_',i.name,'_writer')
AND u.name = :user
```

grouping: NULL

Rules: Variant 2, Static Rules Based on "Magic Names"

Rule

crudFlags: CRUD

what:
```
SELECT aug FROM UserGroup aug
JOIN aug.grouping ag
JOIN Investigation i
JOIN Grouping g JOIN g.userGroups ug
JOIN ug.user u
WHERE ag.name =
CONCAT('investigation_',i.name,'_writer)
AND g.name =
CONCAT('investigation_',i.name,'_owner')
AND u.name = :user
```

grouping: NULL

Rules: Variant 2, Static Rules Based on "Magic Names"

- Works in principle in test installation.
- Only fixed set of static rules.
- Drawback: incredible slow! Seven minutes to answer a simple query on a test ICAT having about 700 investigations.

Rules: Variant 2, Static Rules Based on "Magic Names"

Why is it so slow?

Query

```
SELECT df FROM Datafile df
JOIN df.dataset ds JOIN ds.investigation i
JOIN Grouping g JOIN g.userGroups ug
JOIN ug.user u
WHERE g.name = CONCAT('investigation_',i.name,'_writer')
AND u.name = :user
```

Missing relation between Investigation and Grouping
$\Rightarrow$ Need to evaluate string expression on full cartesian product.

Complexity: $\mathcal{O}(n^2)$ in the number of investigations.

InvestigationGroup

Possible solution: Add the missing relation. Add to ICAT schema:

InvestigationGroup

Many to many relationship between investigation and grouping
Uniqueness constraint: grouping, investigation

Relationships:

Card	Class	Field	Cascaded
1,1	Investigation	investigation	No
1,1	Grouping	grouping	No

Other fields:

Field	Type
role	String [255]

Rules: Variant 3, InvestigationGroup

Add relations between Investigation and Groups:

InvestigationGroup

investigation: Investigation `<name>`

grouping: Grouping `investigation_<name>_writer`

role: writer

and accordingly for reader and owner.
Add rules:

Rule

crudFlags: CRUD

what:
```
Datafile <-> Dataset <-> Investigation <->
InvestigationGroup [role='writer'] <->
Grouping <-> UserGroup <-> User [name=:user]
```

grouping: `NULL`

Rules: Variant 3, InvestigationGroup

Rule

crudFlags: CRUD

what:
```
SELECT tug FROM UserGroup tug
JOIN tug.grouping tg
JOIN tg.investigationGroups tig
JOIN tig.investigation i
JOIN i.investigationGroups uig
JOIN uig.grouping ug
JOIN ug.userGroups uug JOIN uug.user u
WHERE (tig.role = 'writer'
OR tig.role = 'reader')
AND uig.role = 'owner' AND u.name = :user
```

grouping: NULL

Rules: Variant 3, InvestigationGroup

- Solves the issue: only one fixed set of static rules. Only three Grouping and three InvestigationGroup per Investigation.
- Provides all the flexibility.
- Should scale reasonably, at least no obvious reason why it should not.
- Requires a change in the ICAT schema.
- But: this change is limited to the addition of the new type. Already existing types are not altered $\Rightarrow$ no compatibility issues. Sites not using it should not be affected in any way.

Discussion

Thank you for your attention!

Comments? Discussion?

Which Objects to Setup Rules for?

Writers get CRUD permission on:

- `Sample <-> Investigation`,
- `Dataset <-> Investigation`,
- `Datafile <-> Dataset <-> Investigation`,
- `InvestigationParameter <-> Investigation`,
- `SampleParameter <-> Sample <-> Investigation`,
- `DatasetParameter <-> Dataset <-> Investigation`,
- `DatafileParameter <-> Datafile <-> Dataset <-> Investigation`,
- `Shift <-> Investigation`,
- `Keyword <-> Investigation`,
- `Publication <-> Investigation`,
- `InvestigationInstrument <-> Investigation`,

they get RU permission on `Investigation`, and R permission on `InvestigationUser <-> Investigation`.

Mantid & ICAT

About me

{
  “name” : “Jay Rainey”,
  “affiliation” : “ISIS, Mantid”,
  “email” : “jay.rainey@stfc.ac.uk”,
  “github” : “@jawrainey”
}

Agenda

1. Mantid
2. ICAT within Mantid
3. IDS within Mantid
4. Future plans
5. Questions

Mantid

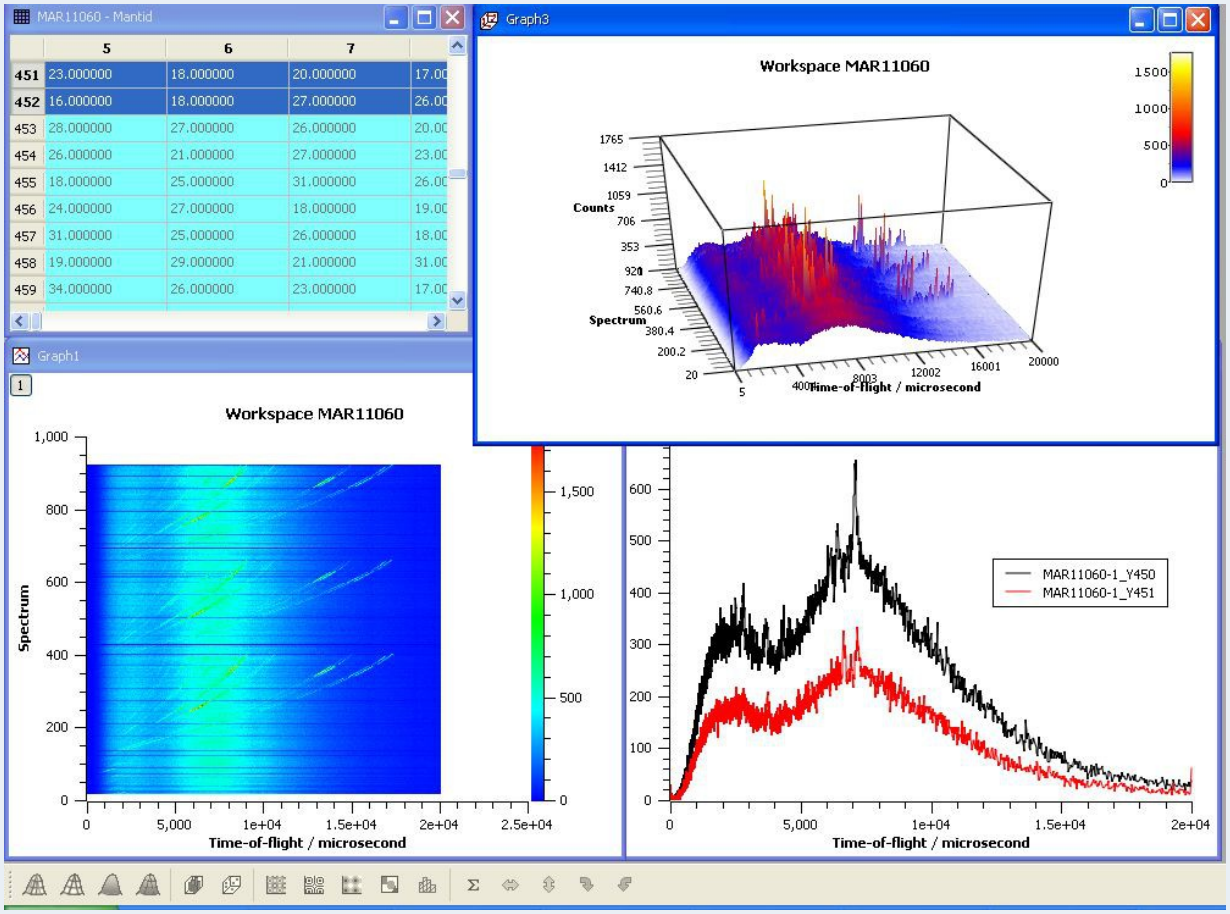
- What is Mantid?
  - Framework that facilitates HPC & visualization and analysis of *neutron* and *muon* data
  - Collaboration between ISIS and SNS
  - Supports various techniques
  - Inelastic (Direct/Indirect)
  - Diffraction
  - Muon
  - Disordered Materials

Mantid

- Features include:
  - Flexible: Instrument/Facility/Technique independent
  - Cross platform (UNIX, Win, OS X)
  - Open source (@mantidproject)
  - Ease of use (Expert -> Visitor)

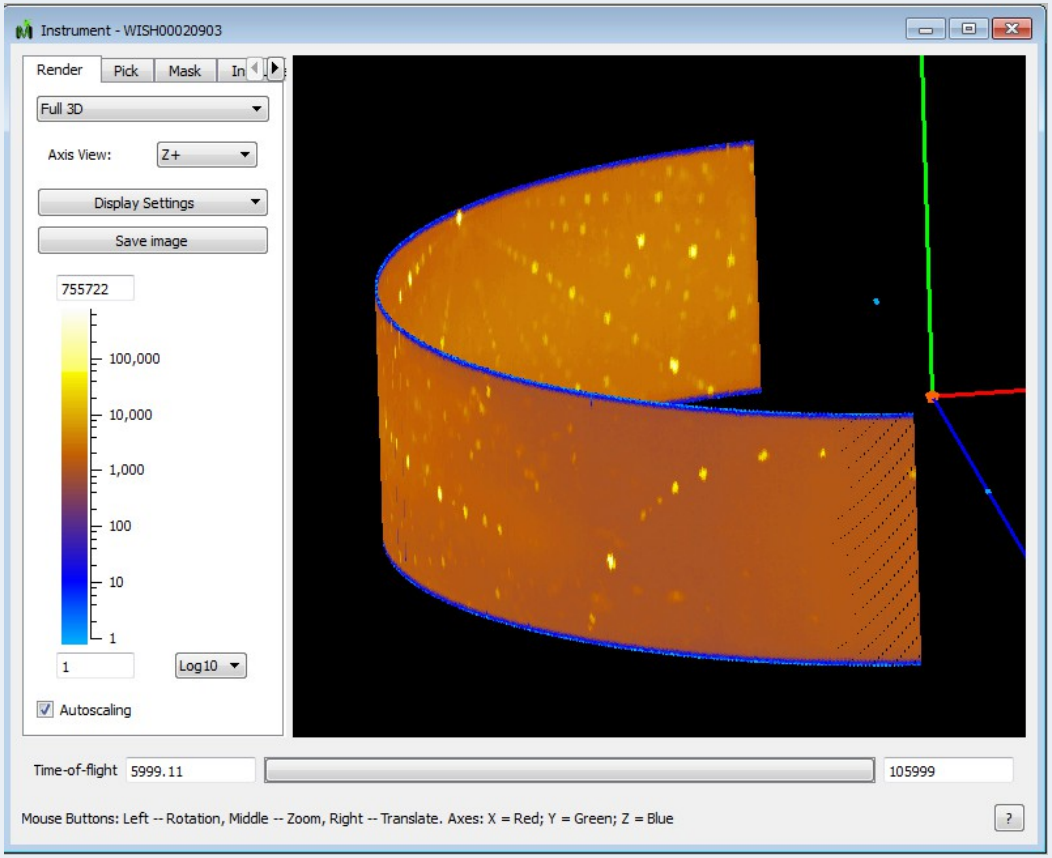
Mantid - visualisations

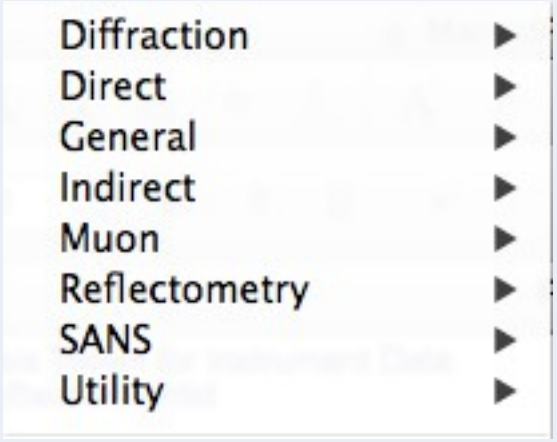
- MantidPlot



Mantid - visualisations

- Custom interfaces





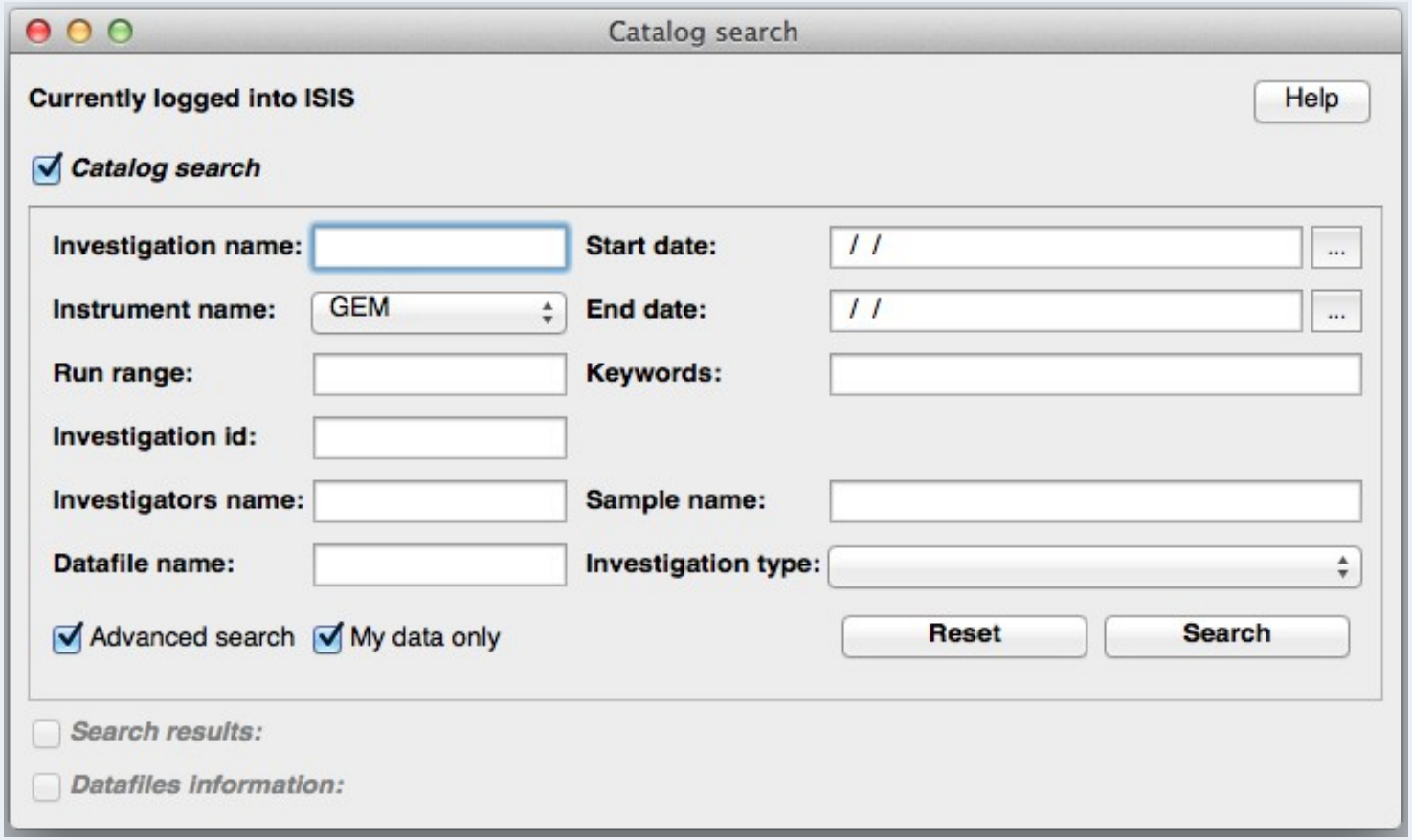
ICAT within Mantid

- Why is it used?
  - Load data directly from archives
  - Easily access catalogued data

- How is it used?
  - **New** Catalog == **new** adaptor

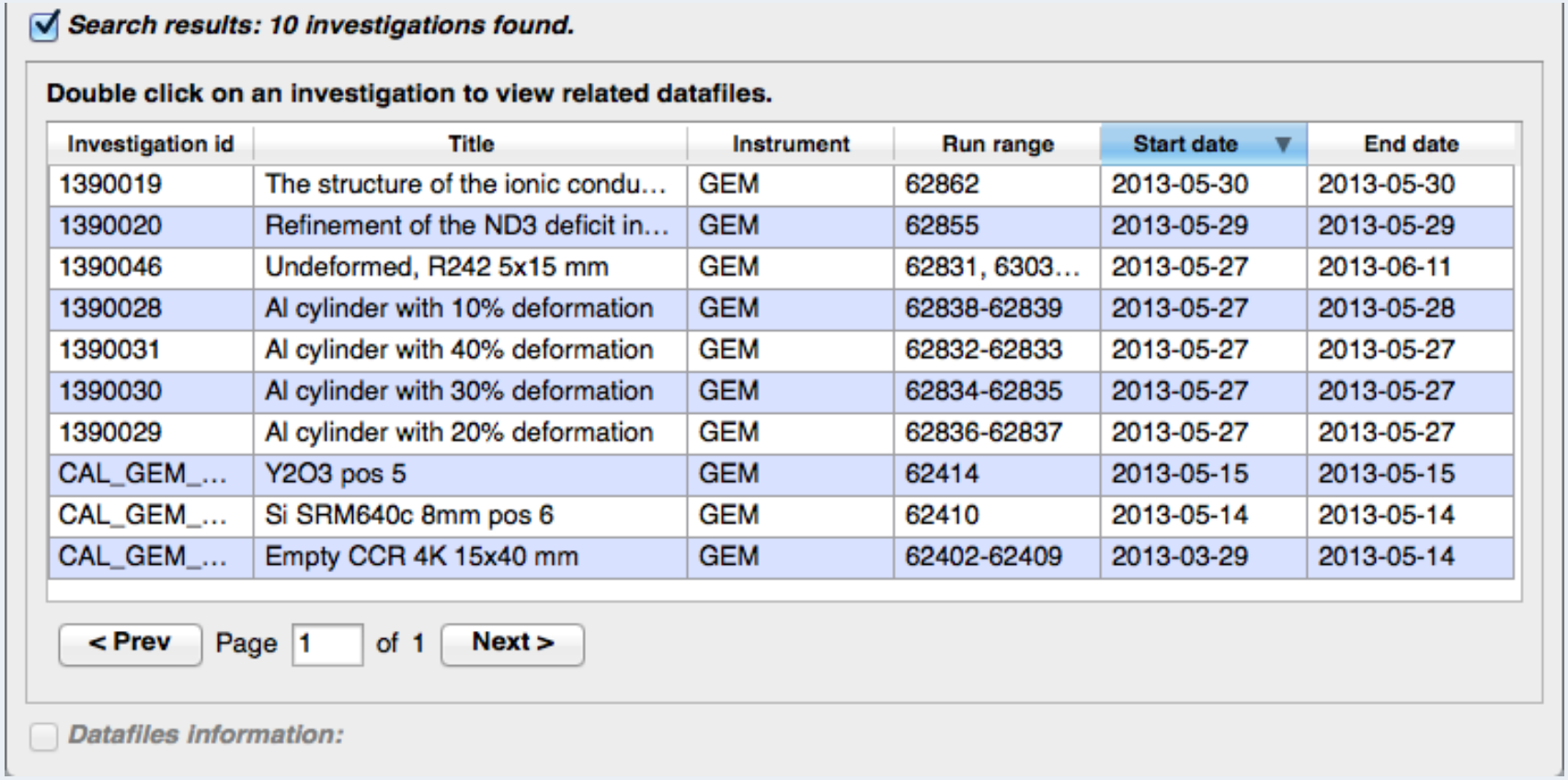
ICAT within Mantid

- Search parameters



ICAT within Mantid

- Search results



Search results: 10 investigations found.

Double click on an investigation to view related datafiles.

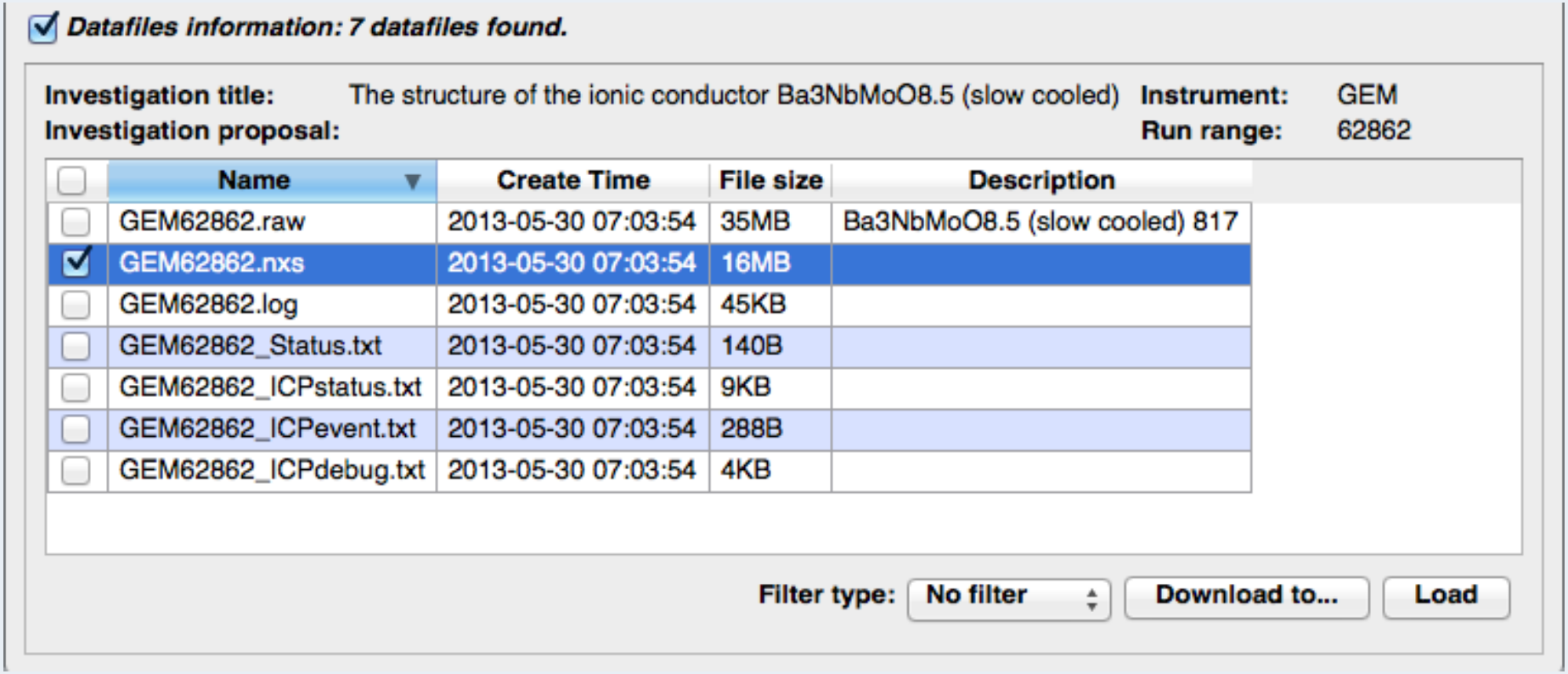
Investigation id	Title	Instrument	Run range	Start date	End date
1390019	The structure of the ionic condu…	GEM	62862	2013-05-30	2013-05-30
1390020	Refinement of the ND3 deficit in…	GEM	62855	2013-05-29	2013-05-29
1390046	Undeformed, R242 5x15 mm	GEM	62831, 6303…	2013-05-27	2013-06-11
1390028	Al cylinder with 10% deformation	GEM	62838-62839	2013-05-27	2013-05-28
1390031	Al cylinder with 40% deformation	GEM	62832-62833	2013-05-27	2013-05-27
1390030	Al cylinder with 30% deformation	GEM	62834-62835	2013-05-27	2013-05-27
1390029	Al cylinder with 20% deformation	GEM	62836-62837	2013-05-27	2013-05-27
CAL_GEM_…	Y2O3 pos 5	GEM	62414	2013-05-15	2013-05-15
CAL_GEM_…	Si SRM640c 8mm pos 6	GEM	62410	2013-05-14	2013-05-14
CAL_GEM_…	Empty CCR 4K 15x40 mm	GEM	62402-62409	2013-03-29	2013-05-14

< Prev Page 1 of 1 Next >

Datafiles Information:

ICAT within Mantid

- Investigation datafiles



IDS within Mantid

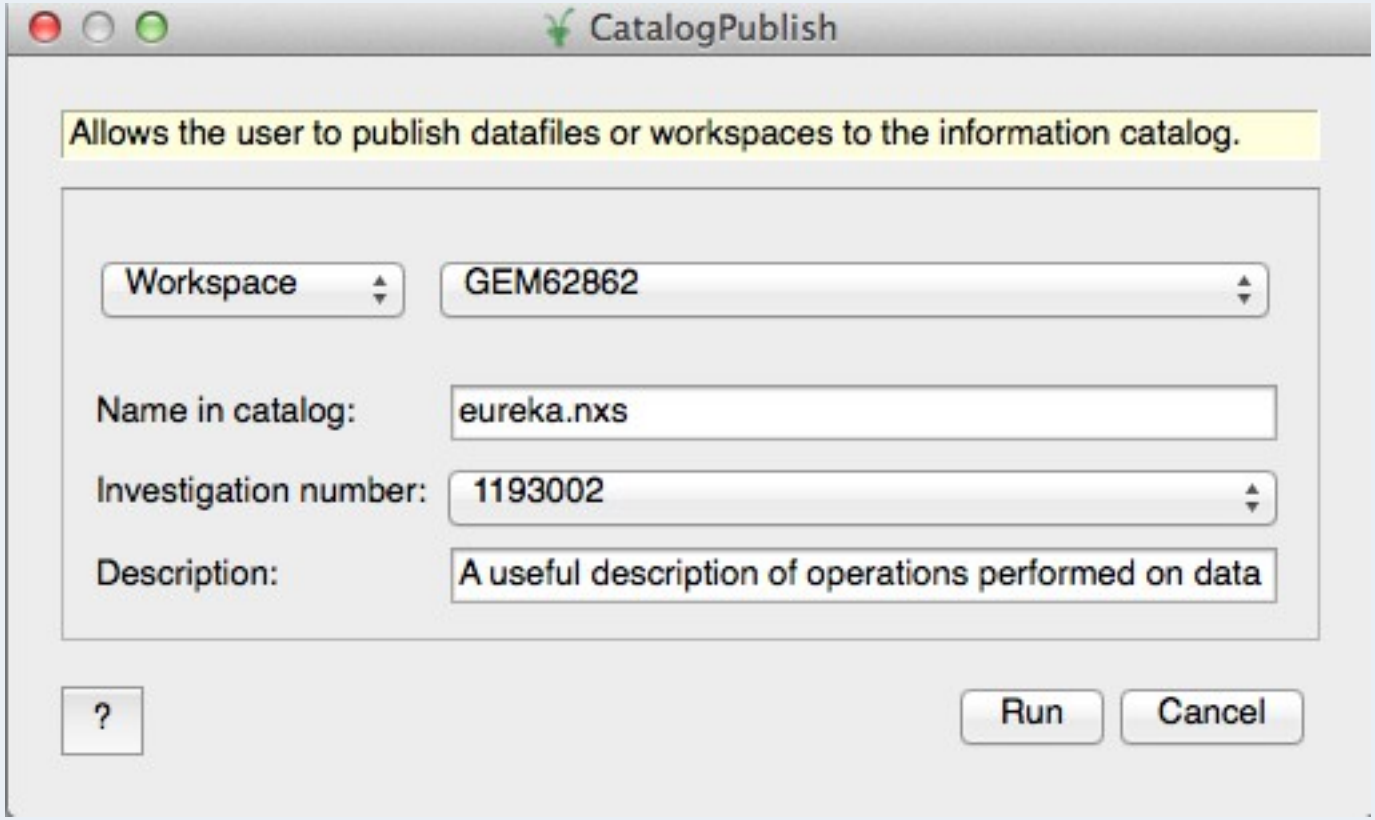
- Downloading (GET)
  - Off-site access
  - Local archive access not required
  - Access latest datafiles

IDS within Mantid

- Publishing (PUT)
  - Share analysed data
  - Automatically upload workspace history

IDS within Mantid

- Publishing GUI



Future plans

- Cross facility querying (optional)
  - One query to be sent to each logged in catalog.
  - Results collated on search GUI

Future plans

- DOIs generated for published *public* data

Questions?

IRO Viewer (Data Journal)

Preservation, Context and Linking

Brian Matthews and Antony Wilson
STFC

ICAT Meeting
Dublin, 25

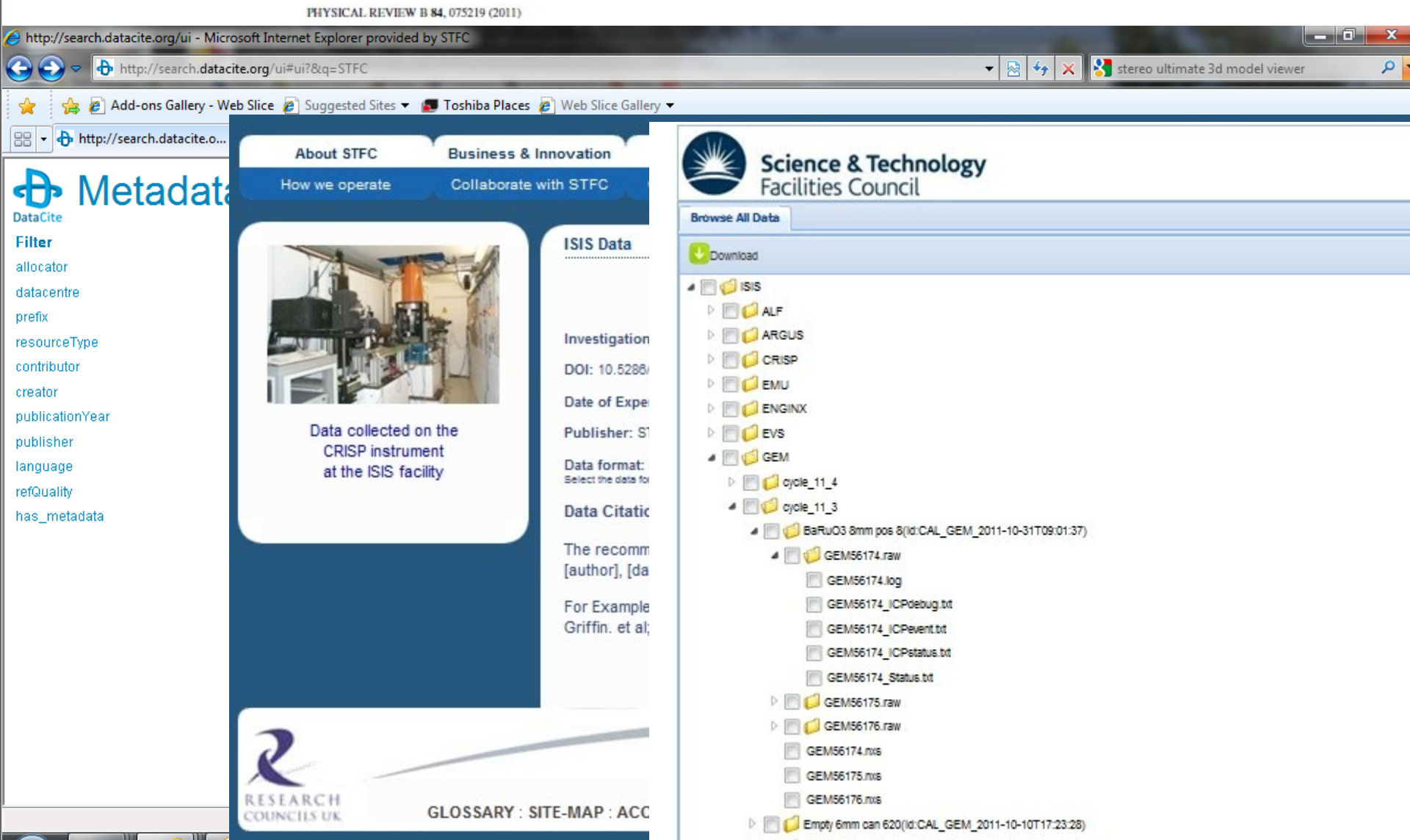
Overview

- Facility data and research objects
- Components
- Demo?
- Preservation
- Summary

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Facility Data and Research Objects

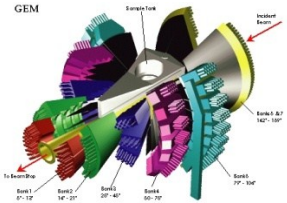
DOI Data Access Process

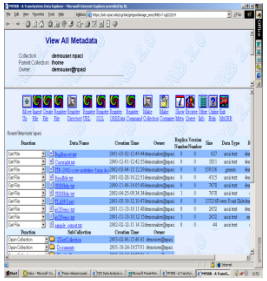


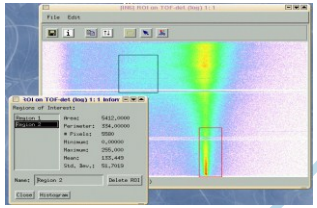
Is this enough?

- What we have so far is good for:
  - us to manage data
  - users to access their own data
  - citation of raw data
- But
  - Publication and exchange of Investigations
    - Traceability and Validation?
    - Reuse of the data?
  - Preservation of Investigations
- Need to make context more explicit
  - Focussing on the dataset is the wrong subject of discourse

Facilities Data Lifecycle







Proposal

Approval

Scheduling

Experiment/ Investigation

Data storage

Data analysis

Record Publication

Scientist submits application for beamtime

Facility committee approves application

Facility registers, trains, and schedules scientist's visit

Scientists visits facility, run's experiment

Raw data filtered, and stored

Tools for processing made available

Subsequent publication registered with facility

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Publishing Investigations

- So what we want is a record of EXPERIMENTS not data.

- Thus want the record of the context
  - The experimental intention and actors
  - The instruments and configurations used
  - The sample
  - The environmental parameters and context
  - The Raw Data
- Thus we want to publish a record of the whole *INVESTIGATION*
  - Can get most of this this from what we have

- The Investigation becomes a “first class” research object
  - Published
  - Identified and treated as a single entity
  - Cited and credited
  - Record of the output of the *facility*

- Analogous to a Journal Article
  - Investigation as the unit of discourse for scientific facilities.

- But also as an access point for validation and reuse
  - Because we have a record of what actually happened.

Our DataCite entries are in fact Investigations

(red is for “data” notion, and green is for “investigation”)

Data collected on the GEM instrument at the ISIS facility

ISIS Data

RB920025.

Investigation title: Crystal and magnetic structures of EuWO1+xN2-x.

Creator: *Kusmartseva,A*
Creator: *Rodgers,J A*
Creator: *Attfield,J P*

DOI: 10.5286/ISIS.E.24071239

Date of Experiment: Tue Aug 04 14:38:23 BST 2009

Publisher: STFC ISIS Facility

Data format: RAW/Nexus
Select the data format above to find out more about it.

DOWNLOAD
download the dataset

Data Citation

The recommended format for citing this dataset in a research publication is as:
[author], [date], [title], [publisher], [doi]

For Example:
Kusmartseva,A. et al; (2009): 920025, STFC ISIS Facility, doi:10.5286/ISIS.E.24071239

Abstract

Eu2+ d0- transition metal perovskites are of interest as potential multiferroics when undoped, or as CMR materials. EuWO1+xN2-x is a new magnetoresistive material and exists over a broad range of x = -0.2 to 0.5. It has a ferromagnetic ordering transition at TC = 12 K. Neutron diffraction is needed to determine the I112/m monoclinic superstructure evidenced by TEM that arises from O/N ordering and octahedral tilting, and the magnetic order. This may include a coexistence of antiferromagnetic/ ferromagnetic orders (as found in a previous GEM study of the analogueEuNbO2N). 2 days on GEM are needed to study 2 samples with different x values (one stoichiometric x =0, the other highly doped x = 0.5) because of high absorption by Eu.

Research Objects

- Represent the “investigation” as a Research Object
  - Research Objects (ROs) are semantically rich aggregations of resources that bring together data, methods and people in scientific investigations. Their goal is to create a class of artefacts that can encapsulate our digital knowledge and provide a mechanism for sharing and discovering assets of reusable research and scientific knowledge
    - www.researchobject.org and elsewhere

- Use RDF, Semantic Web and Linked Data Technologies to support this
  - And we can experiment with these tools for our data

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Building an Investigation Research Object

:investigator

Investigation #n
DOI:STFC.xxx.n

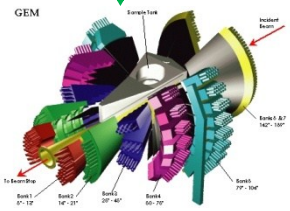
:dataset

:application

:relatedDataset

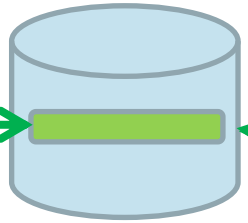
:instrument

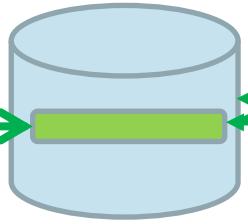
:sample

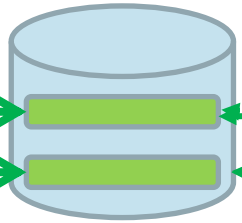


:publication

:publication

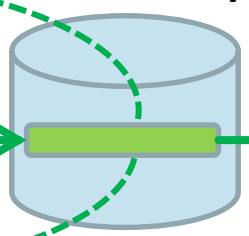




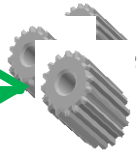


:inputDataset

Software Repository



:outputDataset



Software Package 1

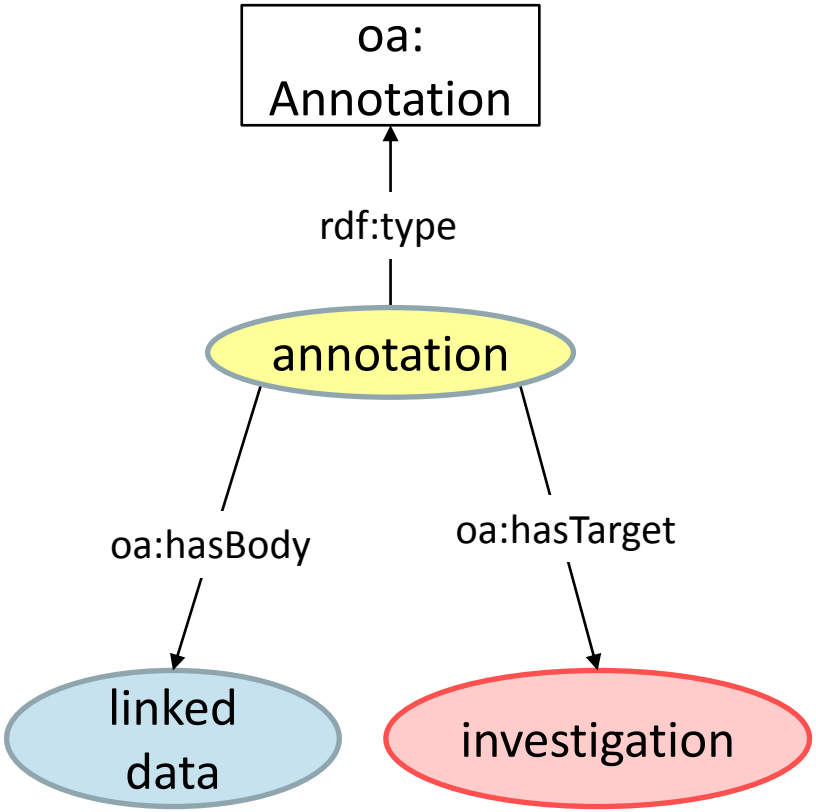
cito:cites

cito:cites

SCAPE

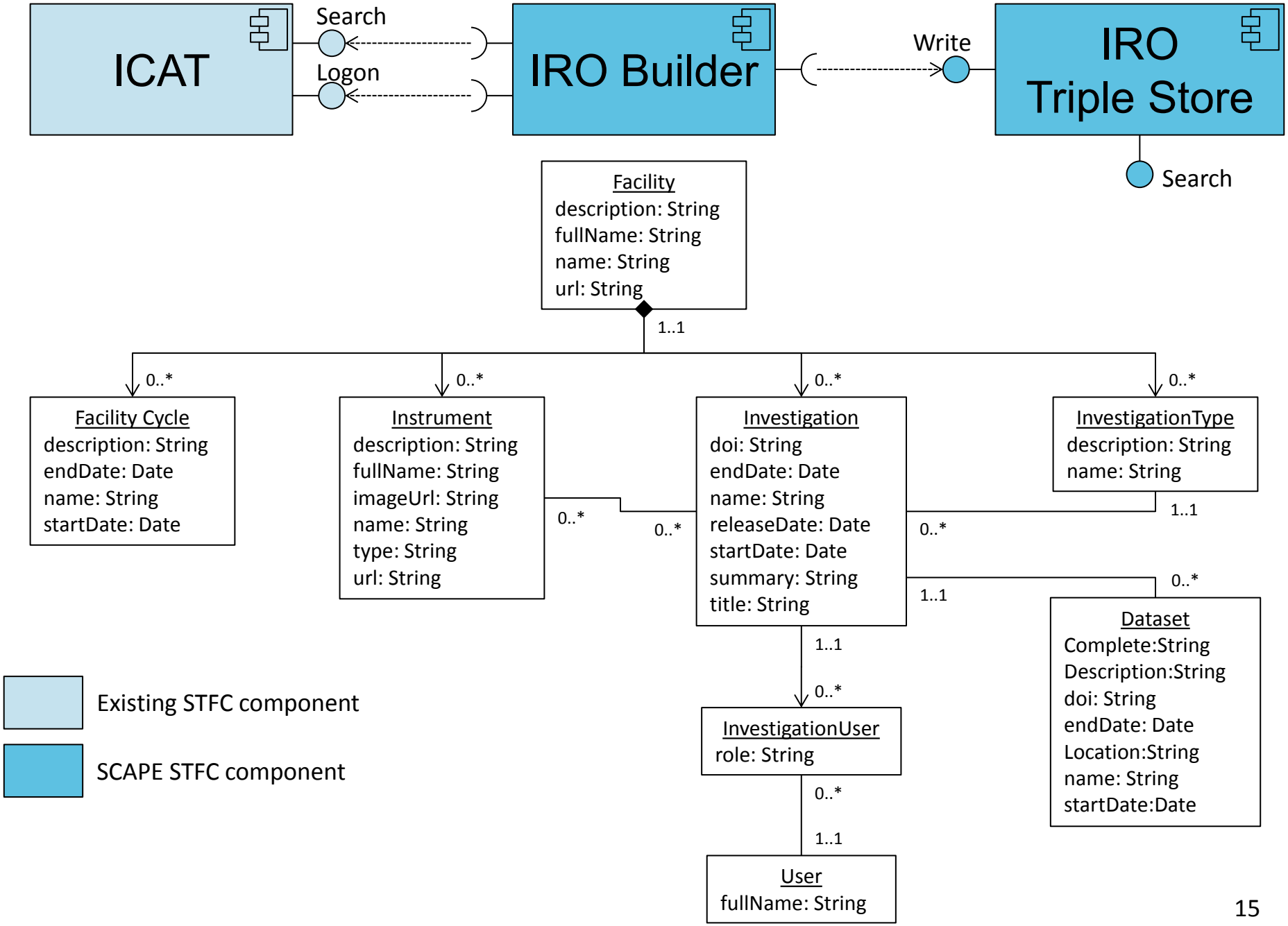
Annotations

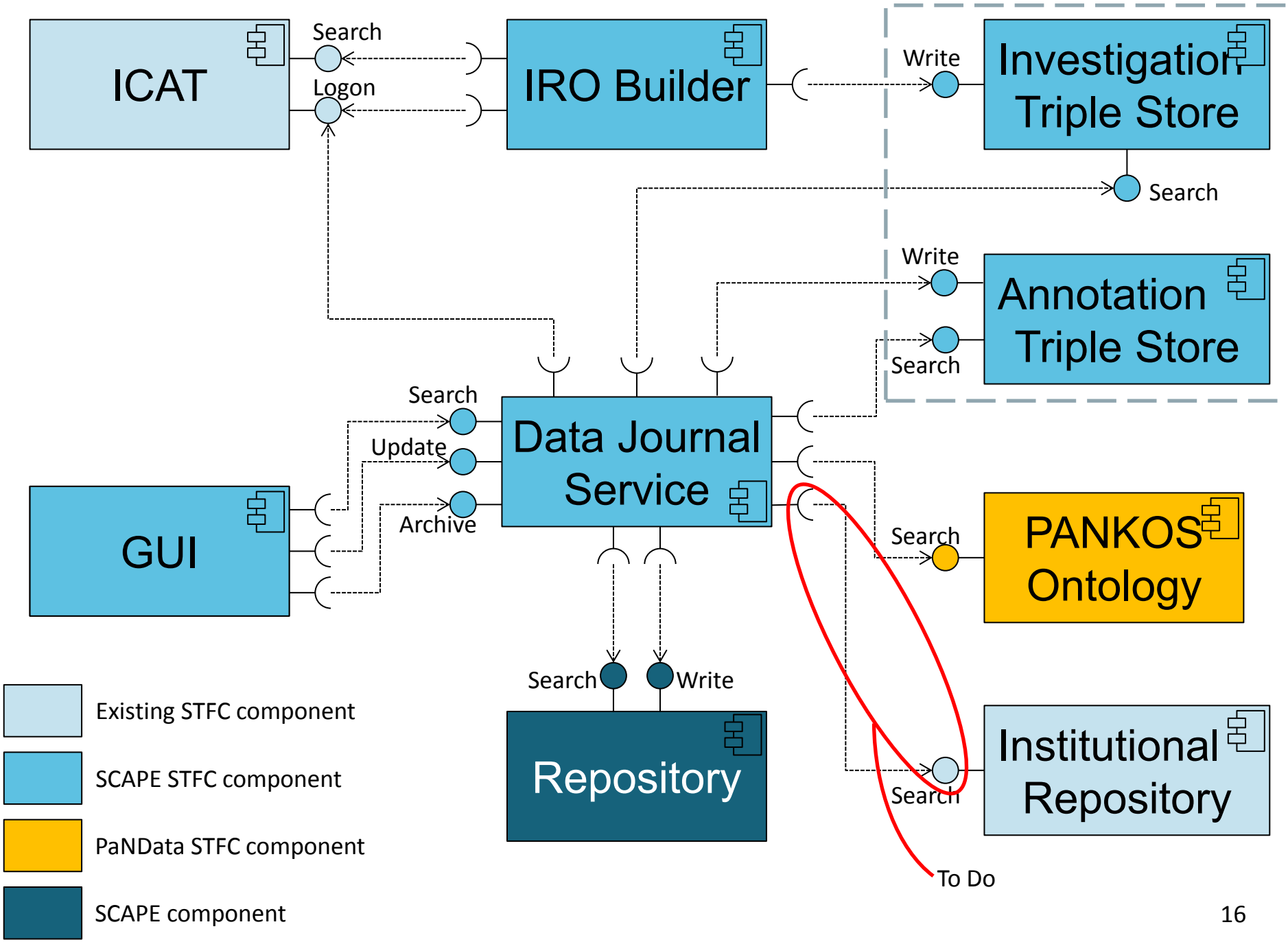
- Annotations used to construct a Research Object
  - http://www.w3.org/community/openannotation/
  - More on this later



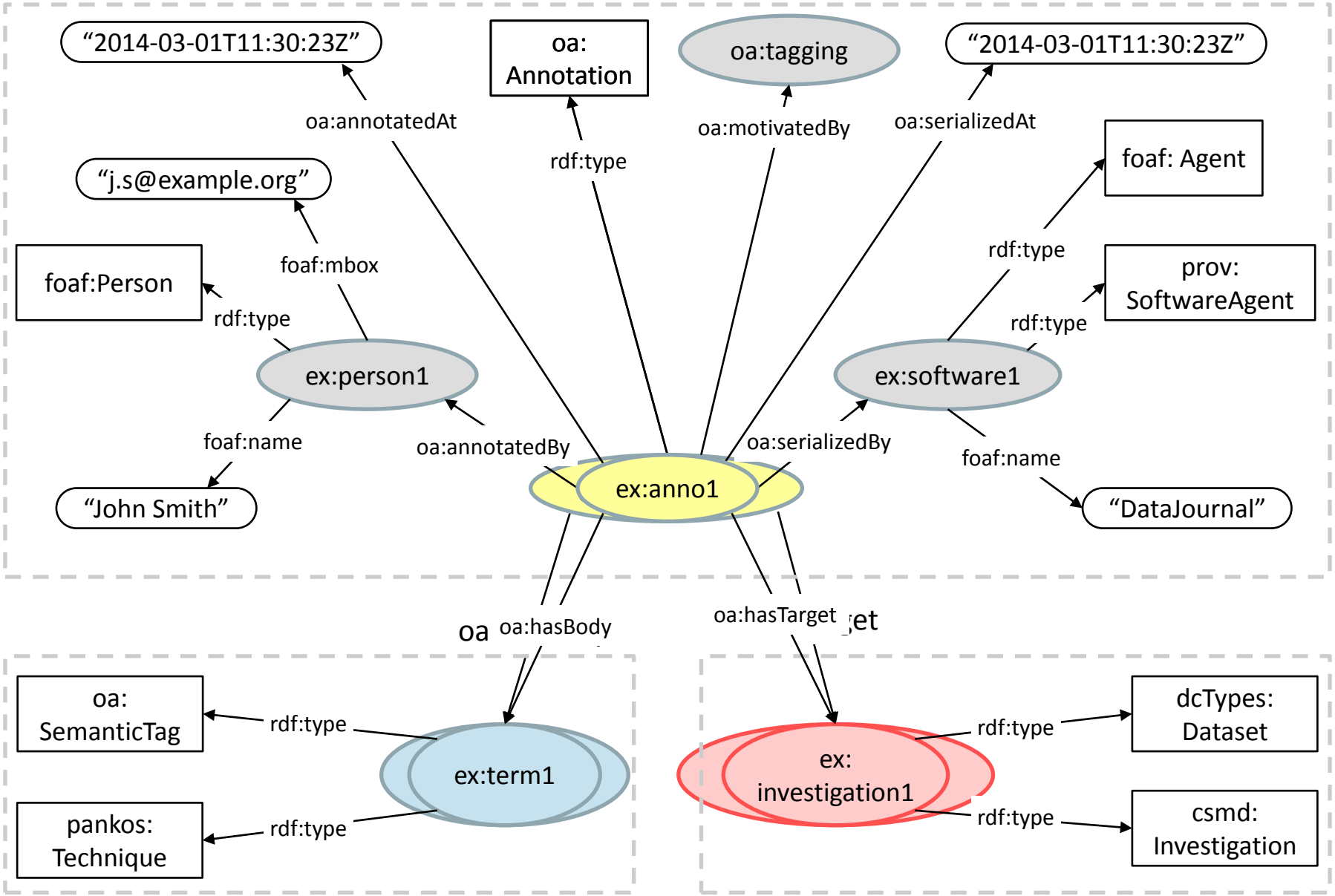
This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Components

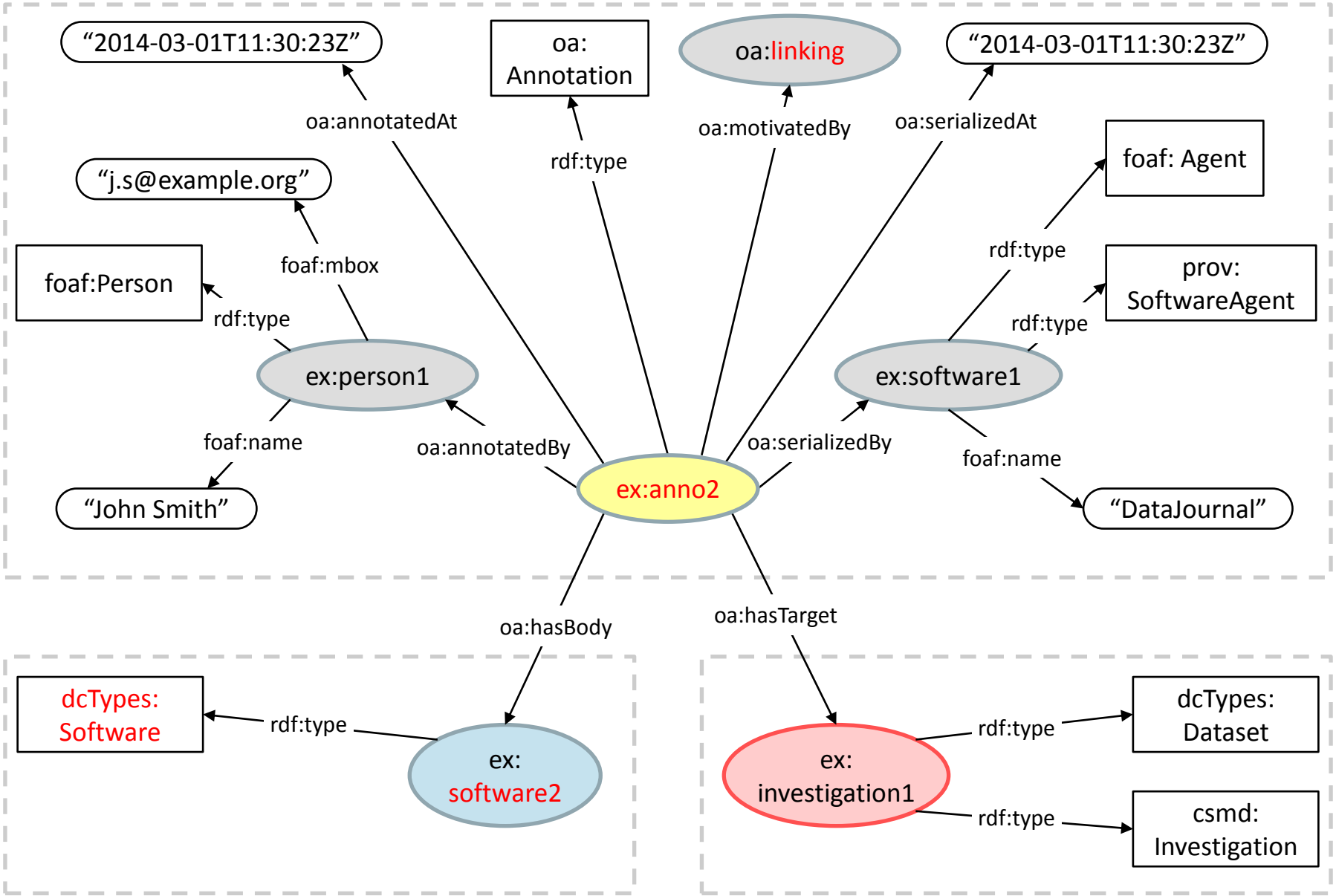

ICAT
Search
Logon
IRO Builder
Write
IRO Triple Store
Search
Facility
description: String
fullName: String
name: String
url: String
1..1
0..*
0..*
0..*
0..*
Facility Cycle
description: String
endDate: Date
name: String
startDate: Date
Instrument
description: String
fullName: String
imageUrl: String
name: String
type: String
url: String
0..*
0..*
Investigation
doi: String
endDate: Date
name: String
releaseDate: Date
startDate: Date
summary: String
title: String
InvestigationType
description: String
name: String
1..1
0..*
1..1
0..*
Dataset
Complete:String
Description:String
doi: String
endDate: Date
Location:String
name: String
startDate:Date
1..1
0..*
InvestigationUser
role: String
0..*
1..1
User
fullName: String
Existing STFC component
SCAPE STFC component


ICAT
Search
Logon
IRO Builder
Write
Investigation Triple Store
Search
Write
Annotation Triple Store
Search
Search
Update
Archive
Data Journal Service
GUI
Search
PANKOS Ontology
Search
Write
Repository
Search
Institutional Repository
To Do
Existing STFC component
SCAPE STFC component
PaNData STFC component
SCAPE component

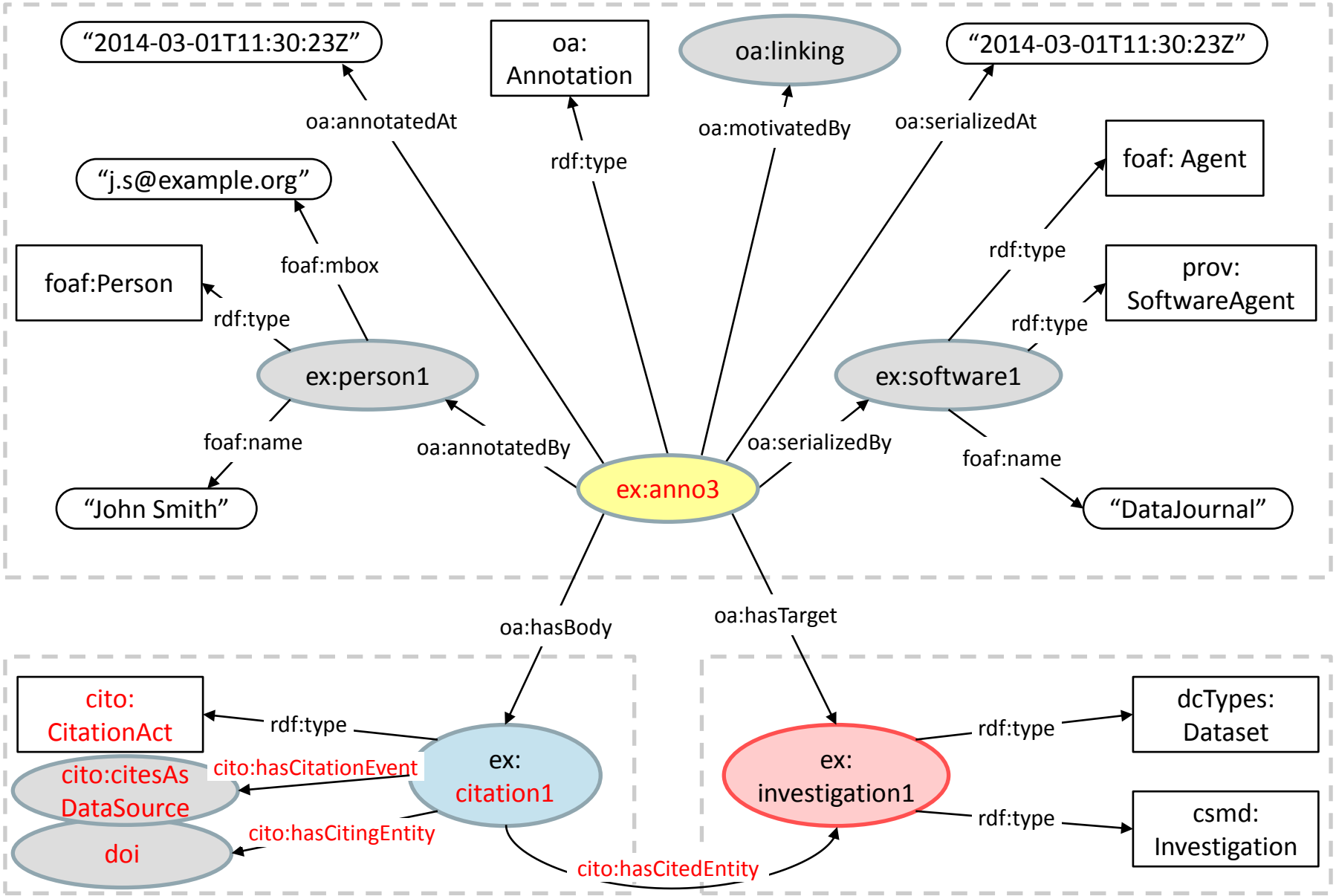
Annotation - Semantic Tagging



Annotation – Linking Software



Annotation – Citations



Demo

Science & Technology Facilities Council

ISIS Data Journal

The archive for ISIS research data

ISIS

Cycles | Investigation | Edit | Archived Versions

This is the main index for the ISIS Data Journal.

There is an entry in the journal for every investigation carried out at ISIS that has a DOI or is no longer in the embargo period. The investigations are grouped into cycles, where a cycle is a period of time during which ISIS was running. The references for the investigation are in the form [Author], [Date], [Title/RB Number], [Publisher], [DOI], where RB Number is a unique number allocated to the investigation. The investigations are sorted by start date. Long running investigations may appear in more than one cycle.

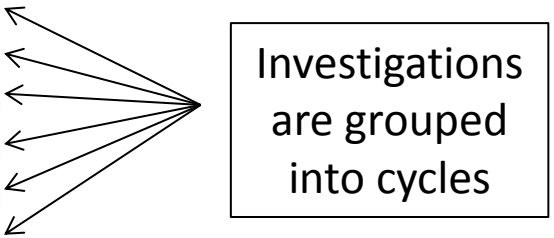
Filters

▶ **Investigation Type**

▶ **Instrument**

Apply

- cycle_13_4 (11/11/2013 - 31/01/2013)
- cycle_13_3 (01/09/2013 - 10/11/2013)
- cycle_13_2 (01/07/2013 - 31/08/2013)
- cycle_13_1 (02/05/2013 - 30/06/2013)
- cycle_12_5 (02/02/2013 - 01/05/2013)
- cycle_12_4 (11/11/2012 - 01/02/2013)
- cycle_12_3 (01/09/2012 - 10/11/2012)
- cycle_12_2 (01/07/2012 - 31/08/2012)
- cycle_12_1 (16/04/2012 - 30/06/2012)
- cycle_11_5 (16/01/2012 - 15/04/2012)
- cycle_11_4 (11/11/2011 - 15/01/2012)
- cycle_11_3 (11/08/2011 - 10/11/2011)
- cycle_11_2 (01/07/2011 - 10/08/2011)
- cycle_11_1 (17/05/2011 - 19/06/2011)
- cycle_10_3 (01/03/2011 - 16/05/2011)
- cycle_10_2 (22/06/2010 - 12/08/2010)
- cycle_10_1 (20/04/2010 - 28/05/2010)
- cycle_09_5 (16/02/2010 - 25/03/2010)
- cycle_09_4 (10/11/2009 - 21/12/2009)



Science & Technology Facilities Council

ISIS Data Journal

The archive for ISIS research data

ISIS

Cycles | Investigation | Edit | Archived Versions

This is the main index for the ISIS Data Journal.

There is an entry in the journal for every investigation carried out at ISIS that has a DOI or is no longer in the embargo period. The investigations are grouped into cycles, where a cycle is a period of time during which ISIS was running. The references for the investigation are in the form [Author], [Date], [Title/RB Number], [Publisher], [DOI], where RB Number is a unique number allocated to the investigation. The investigations are sorted by start date. Long running investigations may appear in more than one cycle.

Filters

- **Investigation Type**
- **Instrument**

Apply

- cycle_13_4 (11/11/2013 - 31/01/2013)
- cycle_13_3 (01/09/2013 - 10/11/2013)
  - Dr Winfried Kockelmann et al; (2013): 1230027, STFC ISIS Facility.
  - ISIS; (2013): CAL_LOQ_2013-08-29T16:41:42, STFC ISIS Facility.
  - ISIS; (2013): CAL_PEARL_2013-08-12T18:54:55, STFC ISIS Facility.
  - ISIS; (2013): CAL_SANS2D_2013-08-28T15:26:21, STFC ISIS Facility.
  - ISIS; (2013): CAL_OFFSPEC_2013-09-06T16:01:53, STFC ISIS Facility.
  - ISIS; (2013): CAL_EMU_2013-08-07T12:08:16, STFC ISIS Facility.
  - Dr Dirk Visser et al; (2013): 1262009, STFC ISIS Facility, doi:10.5286/ISIS.E.24089719.
  - Dr Hazel Sparkes et al; (2013): 1310281, STFC ISIS Facility, doi:10.5286/ISIS.E.24090420.
  - Dr Sylvia McLain et al; (2012): 1200001, STFC ISIS Facility, doi:10.5286/ISIS.E.24089926.
  - Dr Tom Fennell et al; (2012): 1210213, STFC ISIS Facility, doi:10.5286/ISIS.E.24089599.
  - Dr James Lord et al; (2012): 1220055, STFC ISIS Facility, doi:10.5286/ISIS.E.24089615.
  - Dr Radu Coldea et al; (2012): 1210345, STFC ISIS Facility, doi:10.5286/ISIS.E.24088980.

investigations

- cycle_13_2 (01/07/2013 - 31/08/2013)
- cycle_13_1 (02/05/2013 - 30/06/2013)
- cycle_12_5 (02/02/2013 - 01/05/2013)
- cycle_12_4 (11/11/2012 - 01/02/2013)
- cycle_12_3 (01/09/2012 - 10/11/2012)

Science & Technology Facilities Council

ISIS Data Journal

The archive for ISIS research data

ISIS

Cycles | Investigation | Edit | Archived Versions

This is the main index for the ISIS Data Journal.

There is an entry in the journal for every investigation carried out at ISIS that has a DOI or is no longer in the embargo period. The investigations are grouped into cycles, where a cycle is a period of time during which ISIS was running. The references for the investigation are in the form [Author], [Date], [Title/RB Number], [Publisher], [DOI], where RB Number is a unique number allocated to the investigation. The investigations are sorted by start date. Long running investigations may appear in more than one cycle.

Filters

Investigation Type

- Disordered Materials Published Data
- calibration
- commercial_experiment
- engineering
- experiment
- measurement
- simulation
- test

Instrument

Apply

- cycle_13_4 (11/11/2013 - 31/01/2013)
- cycle_13_3 (01/09/2013 - 10/11/2013)
  - Dr Dirk Visser et al; (2013): 1262009, STFC ISIS Facility, doi:10.5286/ISIS.E.24089719.
  - Dr Hazel Sparkes et al; (2013): 1310281, STFC ISIS Facility, doi:10.5286/ISIS.E.24090420.
  - Dr Sylvia McLain et al; (2012): 1200001, STFC ISIS Facility, doi:10.5286/ISIS.E.24089926.
  - Dr Tom Fennell et al; (2012): 1210213, STFC ISIS Facility, doi:10.5286/ISIS.E.24089599.
  - Dr James Lord et al; (2012): 1220055, STFC ISIS Facility, doi:10.5286/ISIS.E.24089615.
  - Dr Radu Coldea et al; (2012): 1210345, STFC ISIS Facility, doi:10.5286/ISIS.E.24088980.
- cycle_13_2 (01/07/2013 - 31/08/2013)
- cycle_13_1 (02/05/2013 - 30/06/2013)
- cycle_12_5 (02/02/2013 - 01/05/2013)
- cycle_12_4 (11/11/2012 - 01/02/20
- cycle_12_3 (01/09/2012 - 10/11/20
- cycle_12_2 (01/07/2012 - 31/08/20
- cycle_12_1 (16/04/2012 - 30/06/20
- cycle_11_5 (16/01/2012 - 15/04/2012)
- cycle_11_4 (11/11/2011 - 15/01/2012)
- cycle_11_3 (11/08/2011 - 10/11/2011)
- cycle_11_2 (01/07/2011 - 10/08/2011)

filtered results

Filters are used to restrict results

Investigations are presented in the style of references

cycle_09_5 (16/02/2010 - 25/03/2010)

Professor John Evans; (2009): Magnetic and Structural Studies on Oxychalcogenides, STFC ISIS Facility.

Dr Diane Holland et al; (2009): Disorder in substituted beta-tricalcium phosphate for waste immobilisation, STFC ISIS Facility.

Dr Stuart Clarke; (2009): calcite water interface, STFC ISIS Facility.

Dr David Lennon et al; (2009): Evolutionary developments in the use of INS to investigate heterogeneous catalysts., STFC ISIS Facility.

Dr Ondrej Muransky et al; (2009): Continuation of Proposal (RB820197): The elasto-plastic response of dual-phase Zr-2.5Nb alloy using in-situ ND diffraction & EPSC modelling, STFC ISIS Facility.

Dr Ali Zarbakhsh et al; (2009): Surfactants at the metal-oil interface, STFC ISIS Facility.

Dr Gabriel Cuello; (2009): Absolute normalisation of spectra in DINS experiments:exploring the route to obtain scattering cross-sections, STFC ISIS Facility.

Dr Mark Johnson et al; (2009): A search for transverse phonons in DNA, STFC ISIS Facility.

Dr Matthew Tucker; (2009): Pressure-temperature phase diagram of of the Jahn-Teller transition in LaMnO3 , STFC ISIS Facility.

Dr Devashi Adroja et al; (2009): Inelastic neutron scattering investigations of the high temperature phase transition in YbAl3C3, STFC ISIS Facility.

Dr Abbie McLaughlin et al; (2009): The Crossover Between Giant Magnetoresistance and Superconductivity in RuSr2Gd1.8-xY0.2CexCu2O10-d , STFC ISIS Facility, doi:10.5286/ISIS.E.24078932.

Dr Howard Stone et al; (2009): Inter-phase load partitioning in a directionally-solidified Cr-Cr3Si eutectic, STFC ISIS Facility, doi:10.5286/ISIS.E.24078918.

Dr Winfried Kockelmann; (2009): Crystallographic determination of the minting of coins in Tudor Britain : an study of the Mary Rose collection. , STFC ISIS Facility.

Dr Matthias Gutmann et al; (2009): Diffuse scattering study of single crystal PrBa2Cu3O7-x, STFC ISIS Facility.

Dr Winfried Kockelmann; (2009): Combined imaging and diffraction studies of Renaissance bronzes , STFC ISIS Facility.

Dr Devashi Adroja et al; (2009): Experimental evidence for quadrupolar order in the heavy fermion compound Ce3Pd20Si6, STFC ISIS Facility.

Dr Devashi Adroja et al; (2009): Excitations in an orbitally dimerized spin S=1 honeycomb antiferromagnet, STFC ISIS Facility, doi:10.5286/ISIS.E.24077575.

unknown; (2009): 28/9-01 D2O + Pb 10t 80K, STFC ISIS Facility.

Dr Christophe Candolfi; (2009): Structural investigations of type-I clathrate compounds, STFC ISIS Facility.

unknown; (2009): μSR STUDY OF SPIN-FLUCTUATIONS IN ITINERANT METAMAGNETIC LACO9SI4, STFC ISIS Facility.

Professor Jon Goff et al; (2009): Divacancy superstructures and enhanced thermopower in NaxCayCoO2, STFC ISIS Facility, doi:10.5286/ISIS.E.24076930.

Dr Maria Paula Marques et al; (2009): THE CONFORMATIONAL BEHAVIOUR OF ANTIOXIDANT CHROMONES, STFC ISIS Facility.

Dr Graeme Blake et al; (2009): Crystal and magnetic structures of the anionogenic ferromagnet Ba0.85K0.15O2, STFC ISIS Facility, doi:10.5286/ISIS.E.24073785.

Dr Edward Oliver et al; (2009): Mg_sample5_-150deg_a;s-5;e0.00;p34.74, STFC ISIS Facility, doi:10.5286/ISIS.E.24071501.

Dr Matthias Gutmann; (2009): Neutron Diffraction Studies on CaCrFe3As3 compound, STFC ISIS Facility.

Science & Technology Facilities Council

ISIS Data Journal

The archive for ISIS research data

ISIS

Cycles | Investigation | Edit | Archived Versions

Previous Investigation | Next Investigation | Archive | Download

RB920302

Investigation title: Divacancy superstructures and enhanced thermopower in NaxCayCoO2

Release Date: 25-03-2013

Creator: *Professor Jon Goff*
Creator: *Professor Alan Tennant*
Creator: *Mr Manoj Soundhira Pandiyan*
Creator: *Dr Sivaperumal Uthayakumar*

DOI: 10.5286/ISIS.E.24076930

Date of Experiment: 22-09-2009 - 25-03-2010

Facility: ISIS Pulsed Neutron & Muon Source
Publisher: ISIS Data Journal

Data format: RAW/Nexus
Select the data format above to find out more about it.

Data Citation
The recommended format for citing this dataset in a research publication is as:
[author], [date], [title], [publisher], [doi]

For Example:
Professor Jon Goff et al; (2009): Divacancy superstructures and enhanced thermopower in NaxCayCoO2, STFC ISIS Facility, doi:10.5286/ISIS.E.24076930.

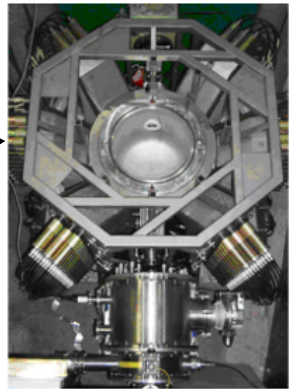
Abstract
Replacing Na ions in NaxCoO2 with divalent Ca ions results in the formation of ordered arrays of divacancy clusters, rather than the trivacancy clusters found for the pure compound at high x. Our previous experiment on SXD showed that Na0.7Ca0.1CoO2 adopts one of the superstructures with a low Coulombic ground state energy. We now wish to explore the ground states of the system doped with divalent Ca ions as a function of concentration in order to compare directly with theory. Doping with divalent ions results in a much lower hole concentration in the cobalt layers, and this is of fundamental interest since it will lead to new electronic and magnetic properties, and technological importance because it is expected to improve thermoelectric performance. We shall determine the ground states of NaxCayCoO2 with x+y=7/9, 11/13 and 13/15, where commensurate superstructures are predicted.

▶ **Additional Parameters**

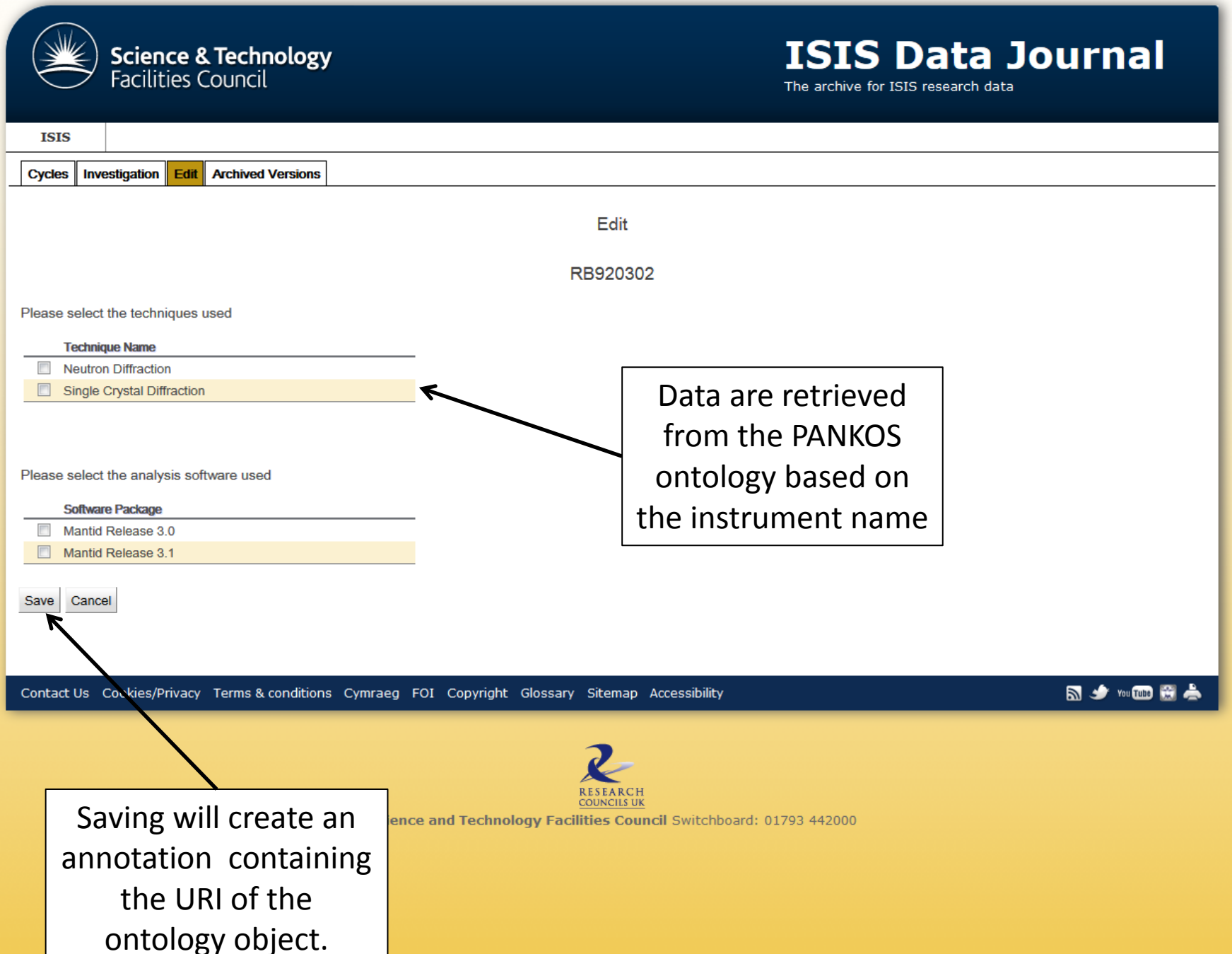
▶ **Data Sets**

Adding context

Details about an investigation are retrieved from the triple store



Data collected on the SXD instrument at the ISIS Pulsed Neutron & Muon Source facility


Science & Technology Facilities Council
ISIS Data Journal
The archive for ISIS research data
ISIS
Cycles
Investigation
Edit
Archived Versions
Edit
RB920302
Please select the techniques used
Technique Name
Neutron Diffraction
Single Crystal Diffraction
Please select the analysis software used
Software Package
Mantid Release 3.0
Mantid Release 3.1
Save
Cancel
Contact Us
Cookies/Privacy
Terms & conditions
Cymraeg
FOI
Copyright
Glossary
Sitemap
Accessibility
RESEARCH COUNCILS UK
ence and Technology Facilities Council Switchboard: 01793 442000
Data are retrieved from the PANKOS ontology based on the instrument name
Saving will create an annotation containing the URI of the ontology object.

Science & Technology Facilities Council

ISIS Data Journal

The archive for ISIS research data

ISIS

Cycles | Investigation | Edit | Archived Versions

Previous Investigation | Next Investigation | Archive | Download

RB920302

Investigation title: Divacancy superstructures and enhanced thermopower in NaxCayCoO2

Release Date: 25-03-2013

Creator: *Professor Jon Goff*
Creator: *Professor Alan Tennant*
Creator: *Mr Manoj Soundhira Pandiyan*
Creator: *Dr Sivaperumal Uthayakumar*

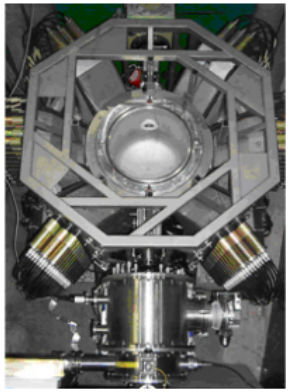
Creating SIP

DOI: 10.5286/ISIS.E.24076930

Date of Experiment: 22-09-2009 - 25-03-2010

Facility: ISIS Pulsed Neutron & Muon Source
Publisher: ISIS Data Journal

Data format: RAW/Nexus
Select the data format above to find out more about it.



Data collected on the SXD instrument at the ISIS Pulsed Neutron & Muon Source facility

Data Citation
The recommended format for citing this dataset in a research publication is as:
[author], [date], [title], [publisher], [doi]

For Example:
Professor Jon Goff et al; (2009): Divacancy superstructures and enhanced thermop … i:10.5286/ISIS.E.24076930.

Now contains merged data from the ontology

Abstract
Replacing Na ions in NaxCoO2 with divalent Ca ions results in the formation of orde … the trivacancy clusters found for the pure compound at high x. Our previous experiment on SXD showed that Na0.7Ca0.1CoO … w Coulombic ground state energy. We now wish to explore the ground states of the system doped with divalent Ca ions as a fu … rectly with theory. Doping with divalent ions results in a much lower hole concentration in the cobalt layers, and this is of fundamental interest since it will lead to new electronic and magnetic properties, and technological importance because it is expected to improve thermoelectric performance. We shall determine the ground states of NaxCayCoO2 with x+y=7/9, 11/13 and 13/15, where commensurate superstructures are predicted.

Techniques
Single Crystal Diffraction

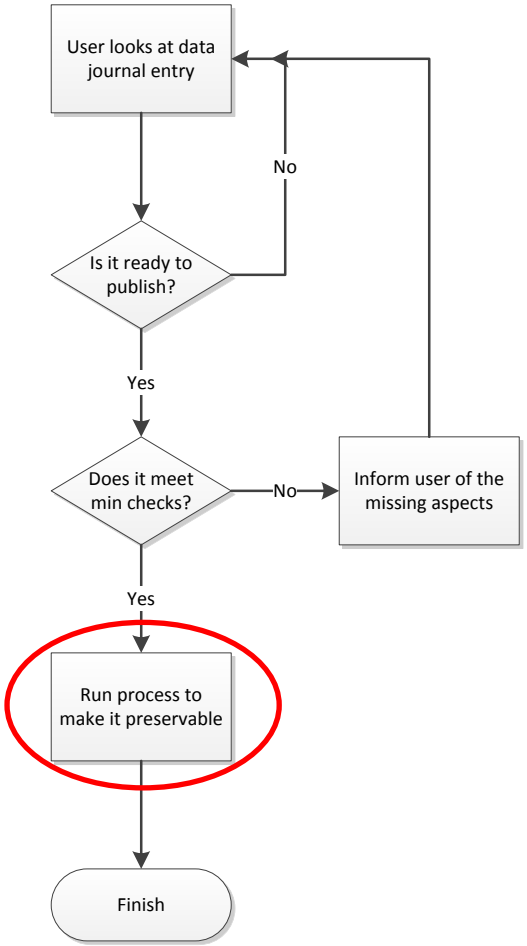
▶ **Additional Parameters**

▶ **Data Sets**

Preservation

User Initiated Preservation

This is the question that the creator of the experiment will ask

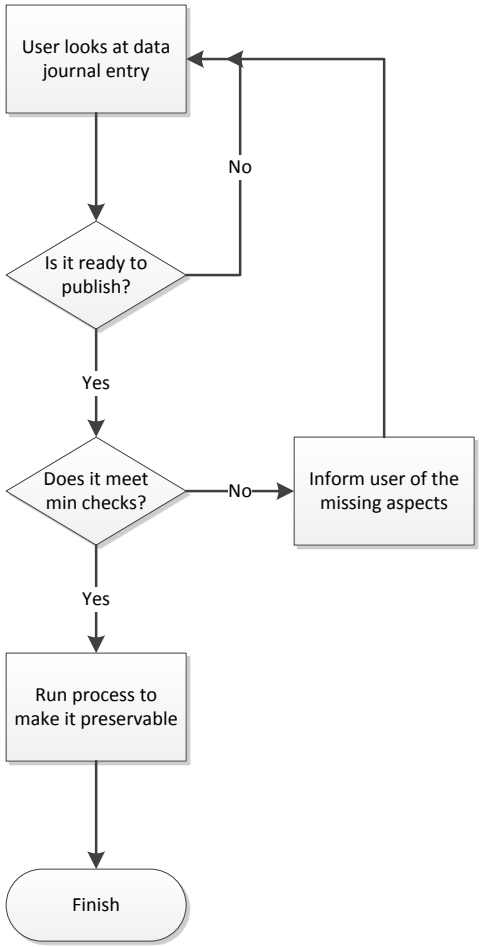


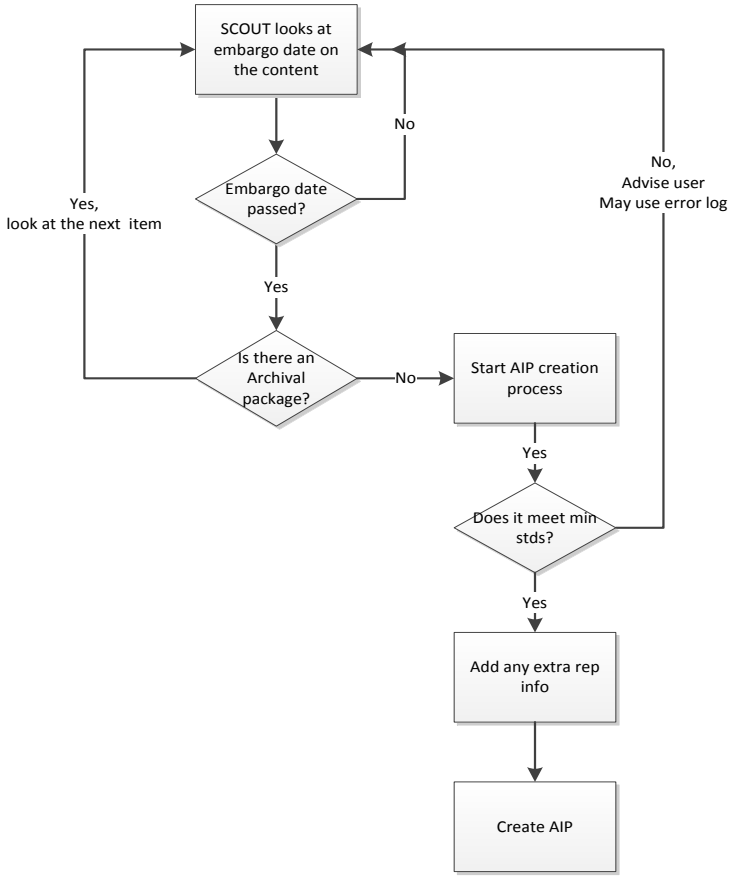
This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

```
<ns14:IRO>
    <ns14:uri>https://icatisis.esc.rl.ac.uk/investigation/24076930</ns14:uri>
    <ns14:facilityName>ISIS Pulsed Neutron &amp;amp; Muon Source</ns14:facilityName>
    <ns14:name>920302</ns14:name>
    <ns14:title>Divacancy superstructures and enhanced thermopower in  NaxCayCoO2</ns14:title>
    <ns14:releaseDate>2013-03-25T15:05:05.000Z</ns14:releaseDate>
    <ns14:startDate>2009-09-22T09:05:07.000+01:00</ns14:startDate>
    <ns14:endDate>2010-03-25T15:05:05.000Z</ns14:endDate>
    <ns14:principalAuthor>Professor Jon Goff</ns14:principalAuthor>
    <ns14:additionalAuthors>Mr Manoj Soundhira Pandiyan</ns14:additionalAuthors>
    <ns14:additionalAuthors>Professor Alan Tennant</ns14:additionalAuthors>
    <ns14:additionalAuthors>Dr Sivaperumal Uthayakumar</ns14:additionalAuthors>
    <ns14:summary>Replacing Na ions in NaxCoO2 with divalent Ca ions results in the formation of …. </ns14:summary>
    <ns14:doi>10.5286/ISIS.E.24076930</ns14:doi>
    <ns14:dataCitation>Professor Jon Goff et al; (2009): 920302, ISIS Pulsed Neutron &amp;amp; Muon Source,
        doi:10.5286/ISIS.E.24076930</ns14:dataCitation>
    <ns14:instrumentName>SXD</ns14:instrumentName>
    <instrumentTechniques>
        <ns14:ontology>
            <dc:dublin-core>
                <dc:contributor>Brian Matthews</dc:contributor>
                <dc:creator>Holly Zhen</dc:creator>
                <dc:description>This ontology describes various neutron and synchrotron facilities from all over Europe, with
                    information regarding their instruments and the techniques used.</dc:description>
                <dc:identifier>1.0</dc:identifier>
                <dc:publisher>Science and Technology Facility Council</dc:publisher>
                <dc:title>Proton and Neutron Knowledge Organisation System</dc:title>
            </dc:dublin-core>
        </ns14:ontology>
        <ns14:technique>
            <ns14:uri>http://www.purl.org/pankos#SingleCrystalDiffraction</ns14:uri>
            <ns14:preferredName>Single Crystal Diffraction</ns14:preferredName>
        </ns14:technique>
    </instrumentTechniques>
</ns14:IRO>
```

Automated Preservation

This is the question that the creator of the experiment will ask





This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Hard Stuff….

- When is a Investigation Research Object Complete?
  - Does this change over time?
  - Does it changes depending on who you are?
- Preserving links – how much trust do you have in others?
  - What does this mean for the preserved object
- Is all data of the same value?
  - For ISIS data in SCAPE : should there be different processes for different investigation types, samples etc?

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Summary

- DONE
  - Initial IRO built from ICAT data
  - Links to ISIS web site automatically added
  - User may add links to PaNData ontology
  - User may add links to software packages
  - User may initiate archive process
  - IRO archived via SCAPE connector to Fedora repo
- What next
  - Add more data to archived IRO
  - Use SCOUT to automate triggering of archive process
  - Link in data from institutional repository

This work was partially supported by the SCAPE Project.
The SCAPE project is co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137).

Science & Technology Facilities Council

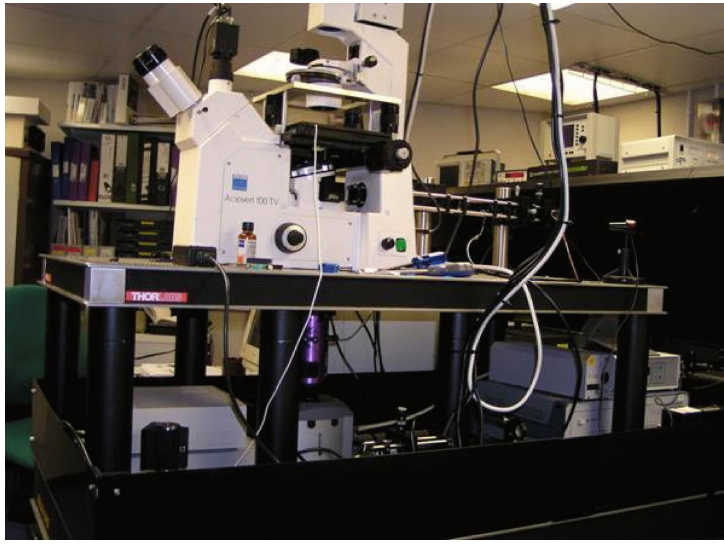
ICAT Job Portal

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

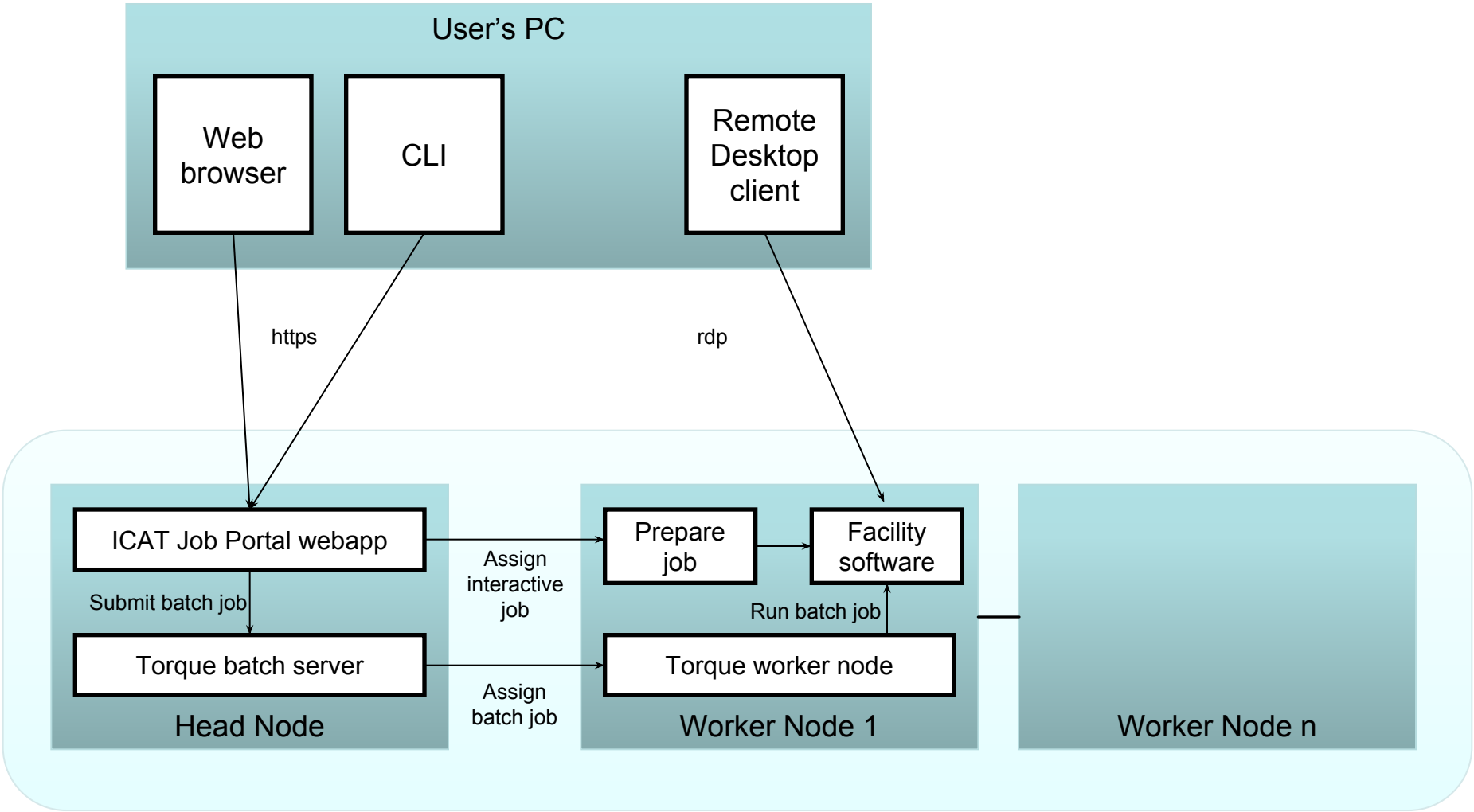


- LSF operate OCTOPUS imaging cluster: lasers coupled to interconnected microscopy stations.
  - a large number of data files
  - applications to process and visualise them
  - interactive program with an easy to use GUI to offer lists of raw and processed datasets and offer the ability to process those 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.

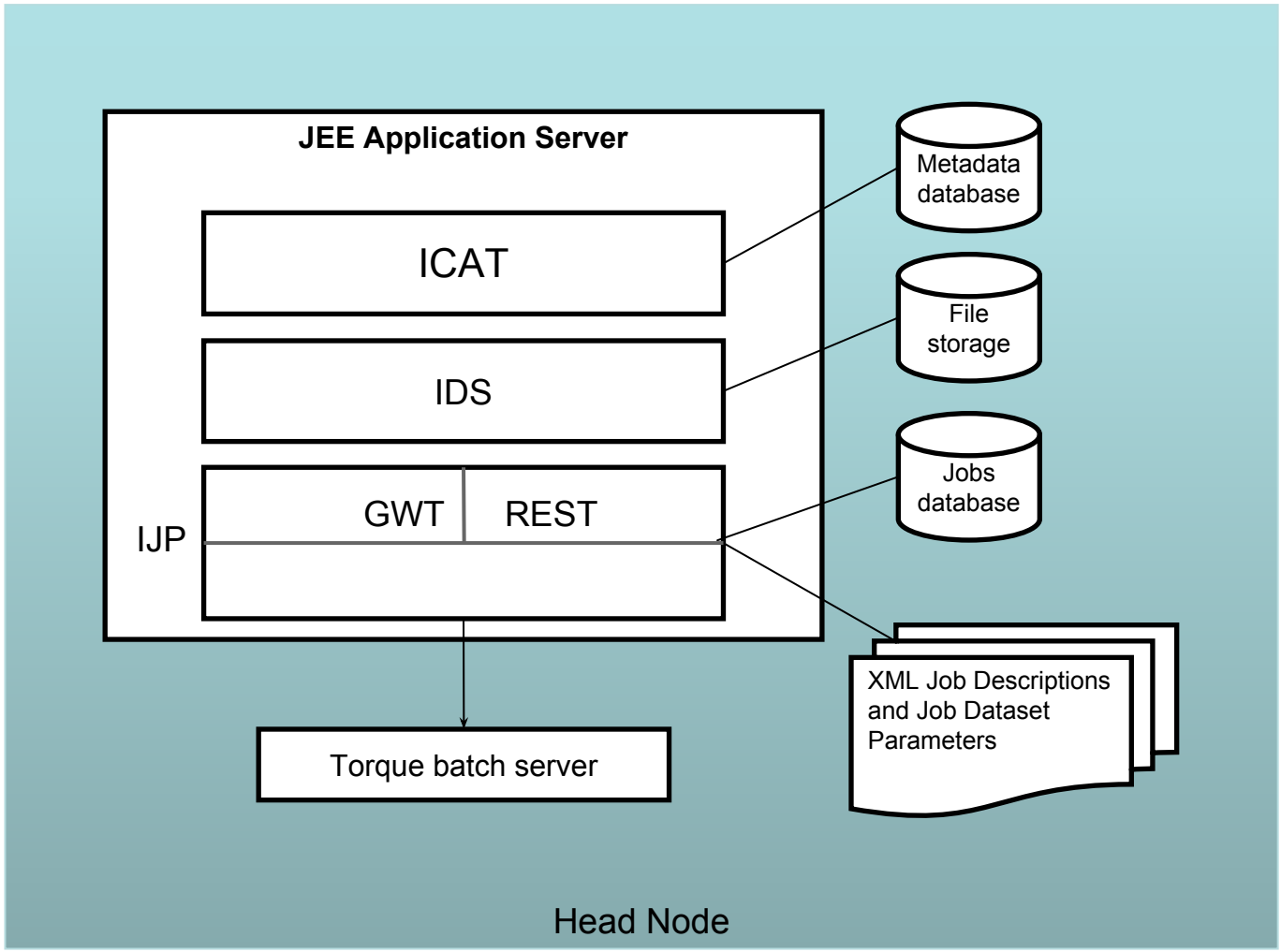
A solution

- 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

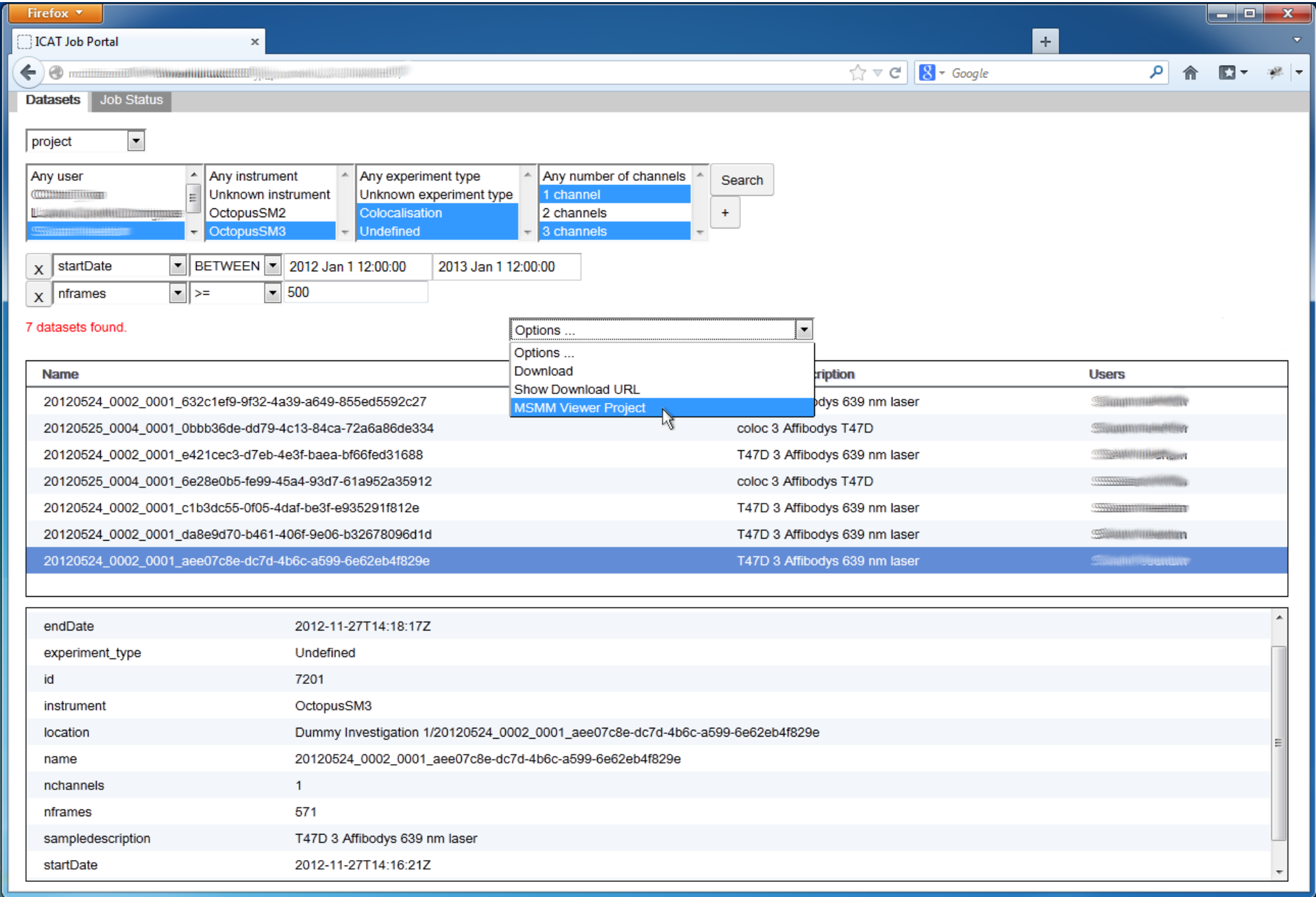
Architecture Overview



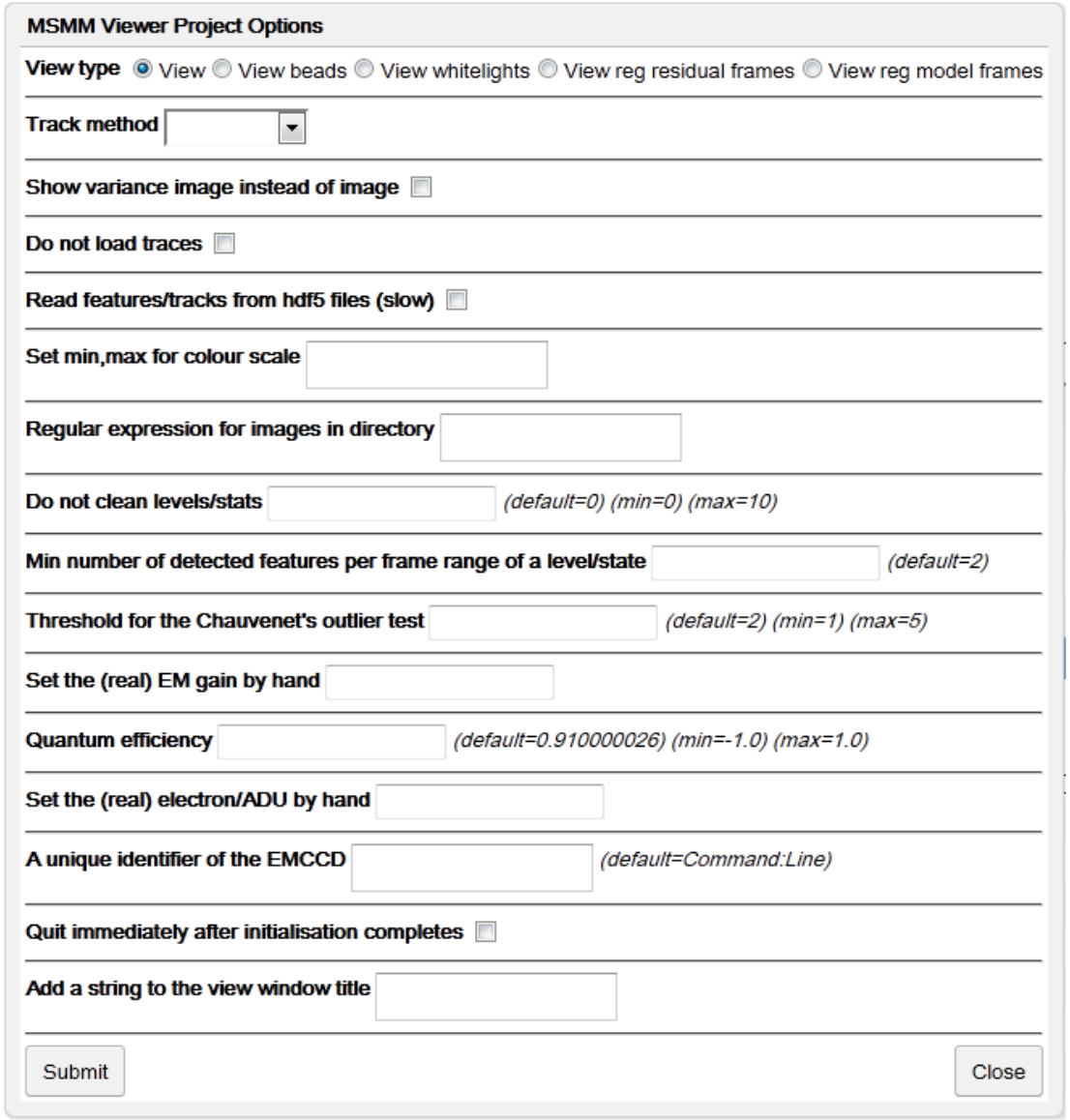
Head Node Architecture



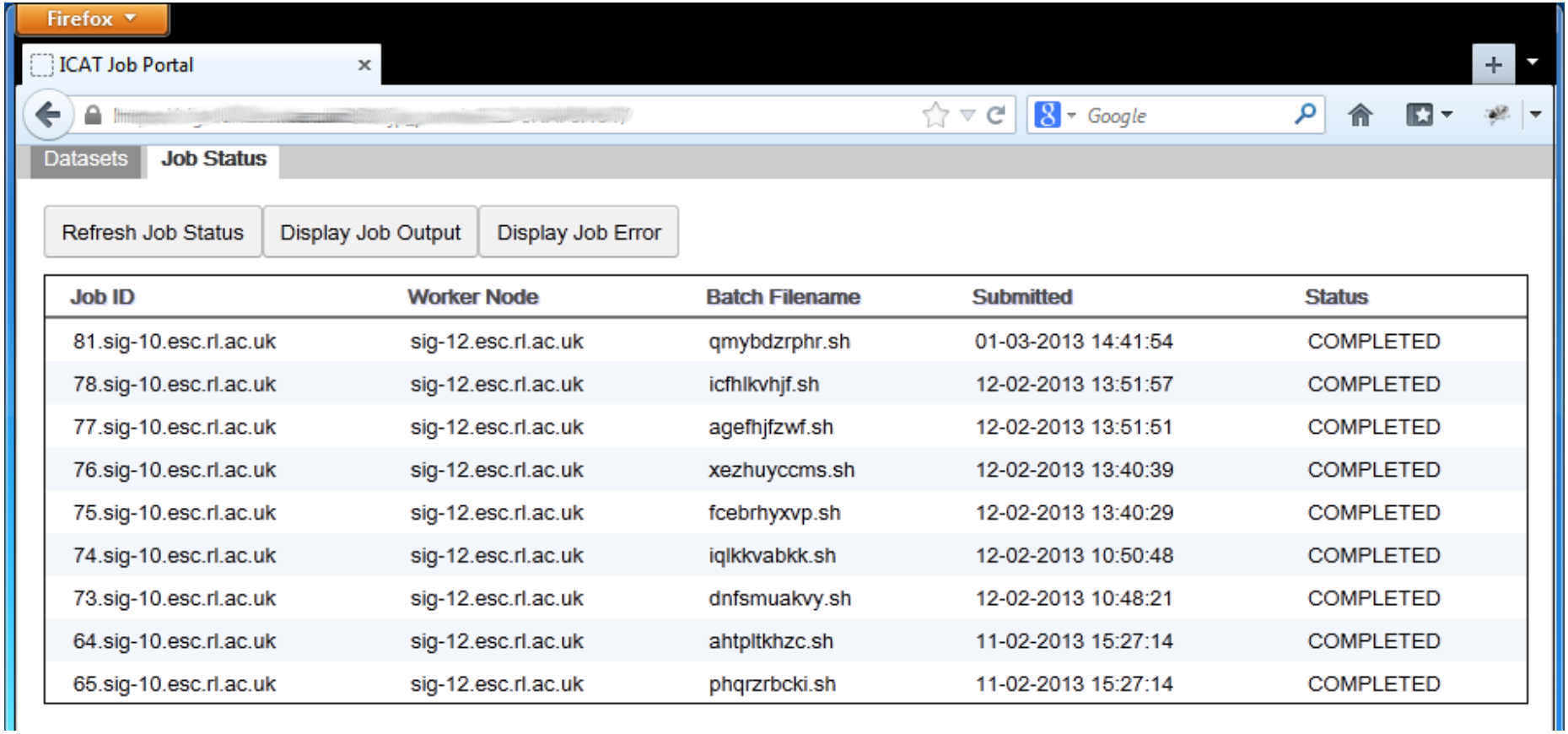
Job Portal Main Panel (Datasets)



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
Submit
Close

Job Status Panel

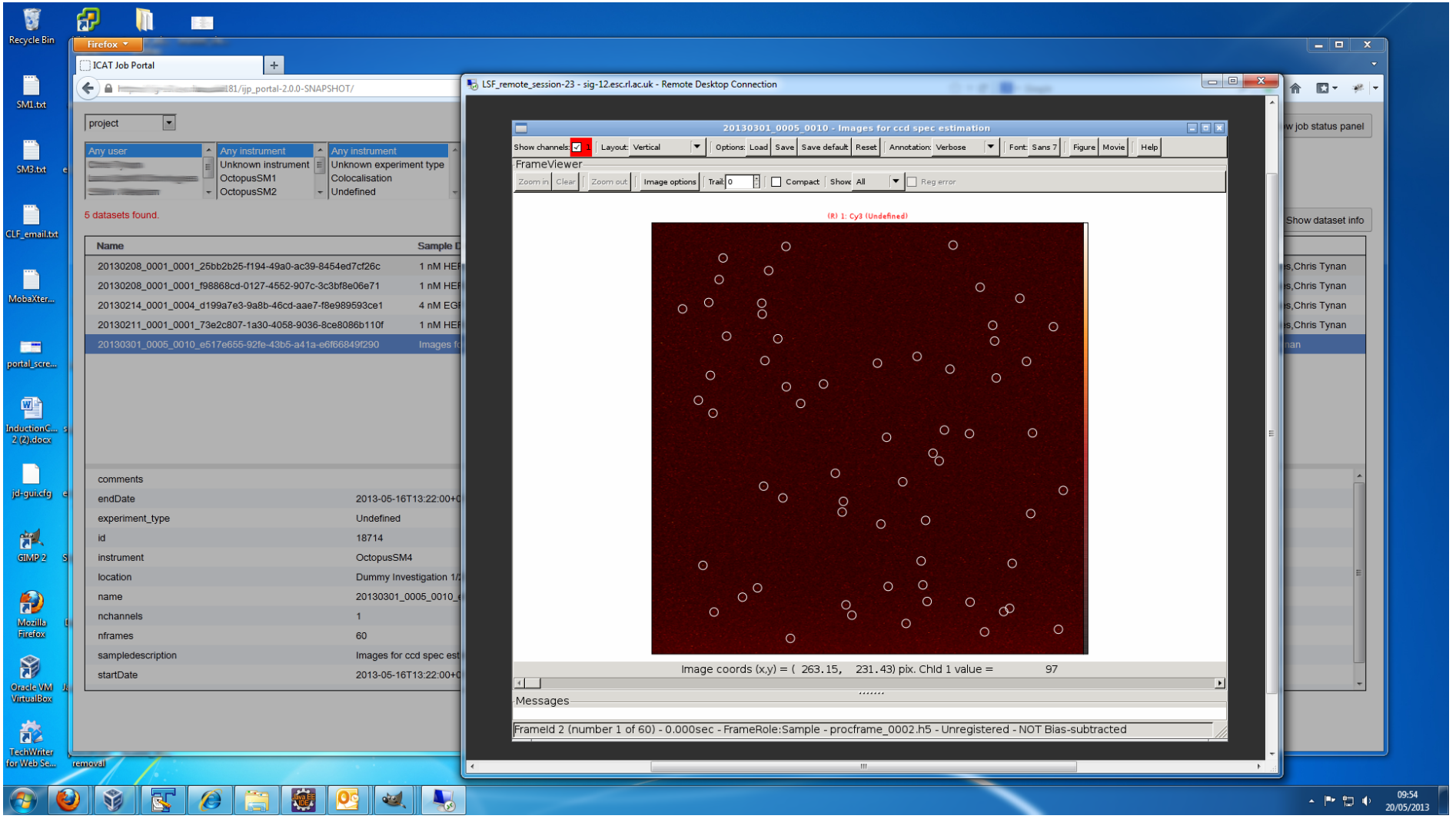


Job ID	Worker Node	Batch Filename	Submitted	Status
81.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	qmybdzrphr.sh	01-03-2013 14:41:54	COMPLETED
78.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	icfhlkvhjf.sh	12-02-2013 13:51:57	COMPLETED
77.sig-10.esc.rl.ac.uk	sig-12.esc.rl.ac.uk	agefhjfzwf.sh	12-02-2013 13:51:51	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 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

```
<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>Do not clean levels/stats</name>
        <type>integer</type>
        <programParameter>--Levels.no-clean</programParameter>
        <defaultValue>0</defaultValue>
```

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>
    <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>Do not clean levels/stats</name>
        <type>integer</type>
        <programParameter>--Levels.no-clean</programParameter>
        <defaultValue>0</defaultValue>
```

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

Experience of the ICAT API and Documentation

MANTID

Agenda

1. ICAT API Experience
2. Documentation
3. Questions

ICAT API Experience

- C++ & gSOAP
- Very simple API
- Each call is (roughly) the same
  - One concrete example needed
- ICAT4 speed!
- Great return errors (IDS & ICAT)

Documentation

- Excellent IDS documentation
  - States what is required, and what is optional

Documentation

- Query Examples
  - No *complex* example of JQPL

```
    SELECT DISTINCT inves
FROM Investigation inves
      JOIN inves.investigationInstruments invInst
      JOIN invInst.instrument inst
      JOIN inves.keywords keywords
WHERE inst.fullName = 'GEM’
AND keywords.name IN ('solid’)
ORDER BY inves.id DESC
```

Documentation

- ICAT Scheme documentation
  - Investigation name
  - Is the investigation ID (this is not documented)
- ICAT Client API documentation
  - Date format: Must be Y/M/D TIME (2011-12-15 00:00:00)

Questions?

Containers

24-25 March 2014

Steve Fisher
Rutherford Appleton Laboratory - STFC

Why not just Glassfish?

- Some sites are already running other containers and don't want to know about another one
- Application bugs might be revealed by testing in different containers
- Glassfish support from Oracle changed in November last year - pointed out to me by Andy Goetz.

Oracle said that it would:
"no longer release future major releases of Oracle GlassFish Server with commercial support — specifically Oracle GlassFish Server 4.x with commercial Java EE 7 support will not be released."

"Oracle recommends that existing commercial Oracle GlassFish Server customers begin planning to move to Oracle WebLogic Server."

What else?


ABANDON FISH!
MIGRATING FROM GLASSFISH TO JBOSS OR TOMEE
So long, and thanks for all the fish!
REBELLABS

Findings

Report indicates that there is not much to choose.

Differences:

- Clustering
- JPA

WildFly 8

Commercial support via RedHat (JBoss)

Use Hibernate as JPA implementation

Java EE7 Certified

TomEE+

Derived from OpenEJB

Collection of Apache Components - including OpenJPA

Commercial support via Tomitribe

Choices

- Do nothing
- Try another container
- If successful could routinely test ICAT components on both or could switch
- Try a third container?
  - Having done two then three should be easy ???

Proposal

- Try WildFly 8 and update the install scripts to be able to use either GlassFish or WildFly
- Add TOMEE+
- Decide on preferred container and do future testing only on that though bug reports on the other “ICAT supported” containers will be accepted and dealt with.
- Low on my priority list at the moment - but it doesn’t have to be done by me.

ICAT core, IDS and IJP Status and Roadmap

24-25 March 2014

Steve Fisher
Rutherford Appleton Laboratory - STFC

ICAT core - Status

Version 4.3 in October with subsequent bug fix releases

Emphasis now on stability and backwards compatibility

Main features of 4.3:

- The notification mechanism has been completely changed.
- Call logging has been added either to file or to a log table or both.
- Lucene free text search via the new searchText call.
- Various “small” schema changes
- Added new alternate query syntax (JQPL based)
  - If you are new to ICAT prefer the JPQL style
- Python setup script which works both on Linux and Windows.

Possible changes to the setup installation script

- Objections to
  - separate config step
  - reports of differences from example
- Proposal
  - avoid separate config step
  - `-vv` required to report differences from example
  - updated scripts will be packaged with new releases

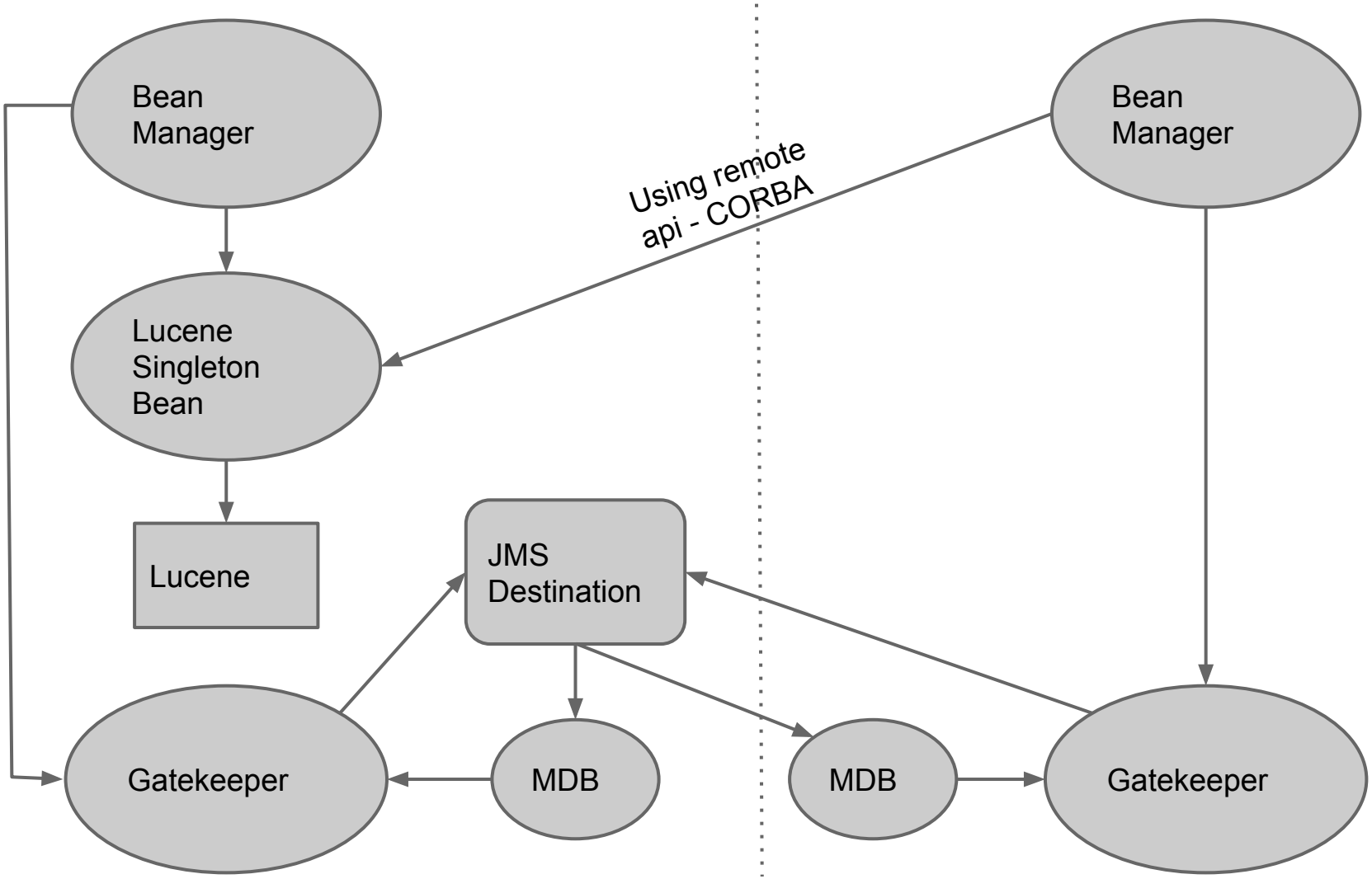
ICAT Core - Next Release (May 2014)

Clustered deployment for better performance

- Chose to avoid container based clustering:
  - Tied to chosen container
  - Glassfish solution not suitable

- Issues
  - Lucene access must be synchronized
  - Gatekeeper has state - cached:
    - Set of public tables (derived from rules)
    - Set of public steps
  - Must disable JPA caching

Clustered Deployment



ICAT Core - Next Release (May 2014)

- Improved API documentation – in particular queries and authorization
  - Necessary drudgery

ICAT Core - Next Release (May 2014) - but not in the road map

- A get function which takes an object with uniqueness constraint fields set
  - inv = get(sid, inv, “INCLUDE 1”)
- Java client was generated for each server release - now being done separately.
- From WSDL which has no comments. If do it from server code can include:
  - Full Javadoc from the entity documentation held by the server
  - Useful constructor for each object taking the uniqueness constraint fields
    - inv = new Investigation(facility, name, visitId)
- getApiVersion() returns a number held by the server
  - could add new call getVersion and deprecate the old call
- authz addition suggested by Rolf

ICAT core - RoadMap (Sep 2014)

- The provision of a JSON based RESTful web service
  - this should be a rather easy add-on

- Ability to migrate metadata/data between ICAT/IDS instances.
  - allow a file to be prepared of data to be fed into ICAT
  - old XMLIngest functionality will be included
  - will be able to deal with any ICAT data
    - not just a hierarchical projection of that data
  - It will also allow a new ICAT to be configured very rapidly.

ICAT core - RoadMap (Jan 2015)

- ICAT - study the use of a non-relational database to give better performance at large scale.

- Some kind of hybrid between relational and non-relational might be optimal.

Questions

?

Suggested changes to the roadmap

?

IDS - Status

First version recently released

Good feedback so far - but a few problems have shown up.

Structure of zip file:

```
"ids"/facility.name/investigation.name/investigation.visitId/dataset.name/datafile.name
```

- if any of the names have characters not suitable for chosen OS it does not work - the "/" is a special case
- in addition ESRF want to be able to control the zip file structure to give facility dependency which I had hoped to avoid.
- mapping the names can lead to duplicates

IDS - Proposed change

- Change the plugin interface to provide more information to the plugin
  - e.g. the Dataset.location field has been requested
- Add an extra call which the plugin must implement to compute the full file name of a zip file entry.
  - Could provide the existing algorithm as a default
  - Facility may choose to base the zip file entry on its own storage structure
    - trivially guarantees unique names

IDS - roadmap

Apr 2014

- The IDS computes checksums when data files are uploaded - use this information to provide background checking that all data can be correctly read to provide the expected checksum.

Aug 2014

- Add the ability to work well when deployed on a shared file system
  - upload and download without file movement
- Extend to provide FUSE file system authorized by ICAT

Questions

?

Suggested changes to the roadmap

?

IJP - Portability

Currently only works on a single platform

Our users want it to run on at least three very different platforms (different OS, different batch system etc.) - but to present a single GUI

This requires (at least)

1. abstracting the batch system
2. providing a very simple “meta-scheduler”

IJP - Other things

- Expect to make a series of small enhancements to meet user needs.
- Also want to make IJP available to others so we will provide a very simple demo configuration with jobs such as concatenating copying and deleting datafiles.

Questions

?

Suggested changes to the roadmap

?

Science & Technology
Facilities Council

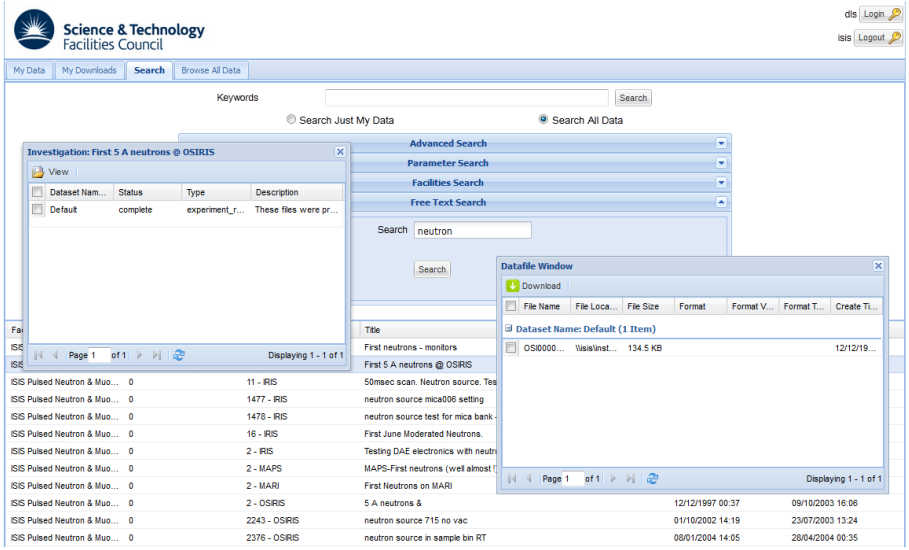
TopCAT

March 2014

Wayne Chung
wayne.chung@stfc.ac.uk

Science & Technology Facilities Council

TopCAT

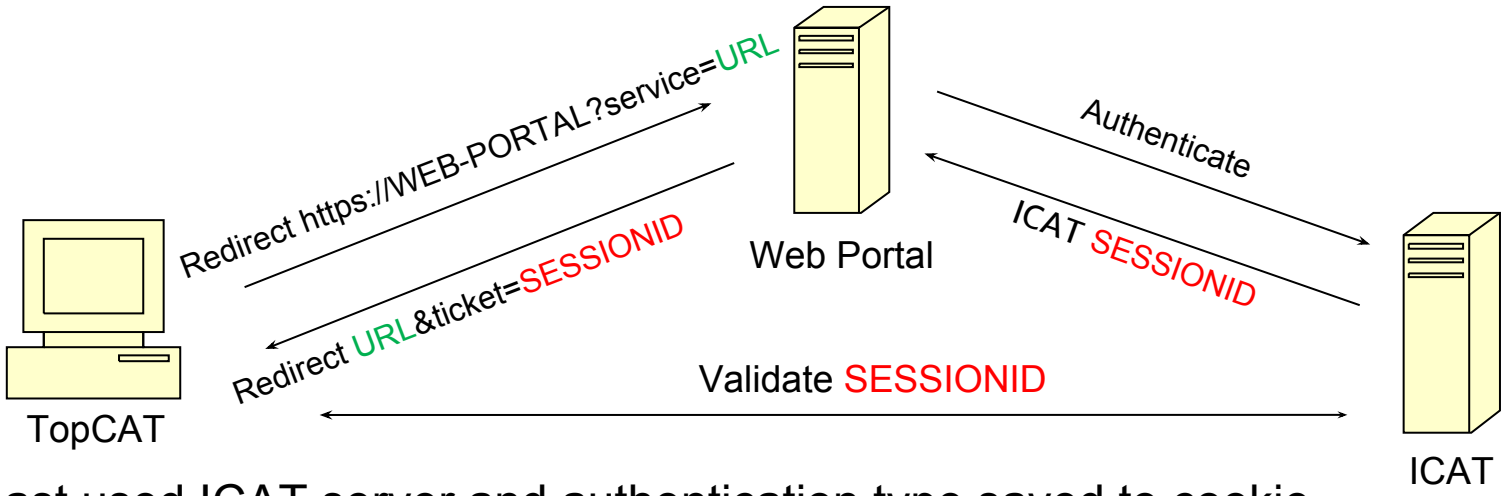


- Current release TopCAT 1.11.0
- Added support for the ICAT 4.3
- Added support for IDS 1.1.0
- Updated GWT 2.5.1
- updated GXT 2.31
- Added timed announcement message via TOPCATAdmin
- Added Free Text search (ICAT 4.3 only)

Next Release

April 2014

- Added External Redirect authentication type plugin

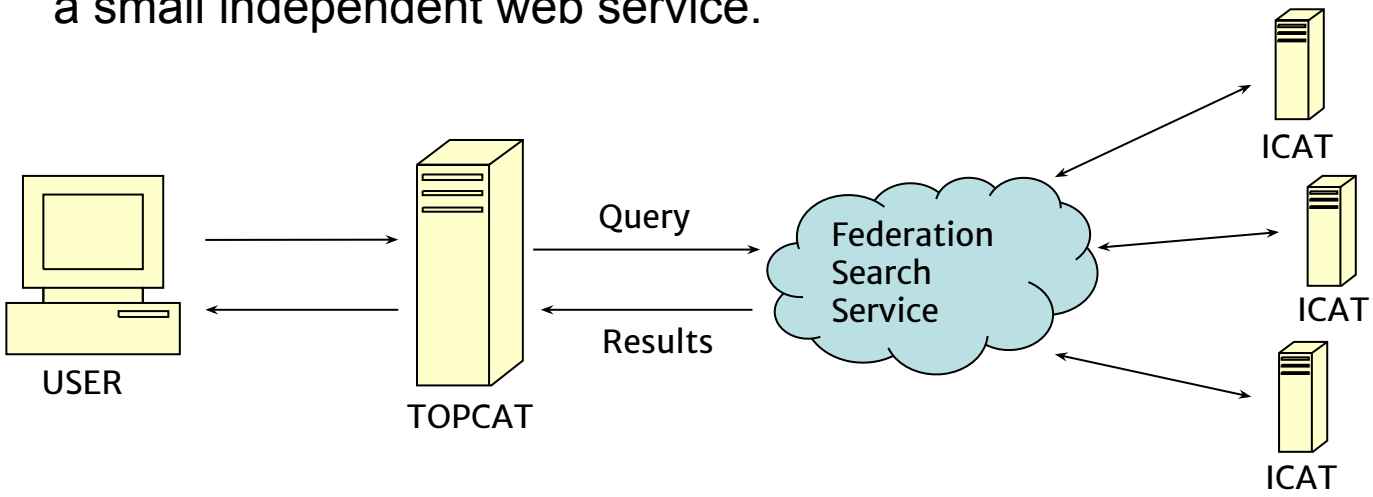


- Last used ICAT server and authentication type saved to cookie
- Provide data upload facilities via the IDS
- Remove support for ICAT below 4.2
- Bug fixes (https://code.google.com/p/topcat/issues/list)

Road Map

July 2014

- Add support for multiple facilities
- Rationalise and simplify the code to reduce dependencies
  - Result in more maintainable software
  - If plugin are desired, should not be compiled into distribution
- Desirable to separate off code for doing federation search, possibly into a small independent web service.

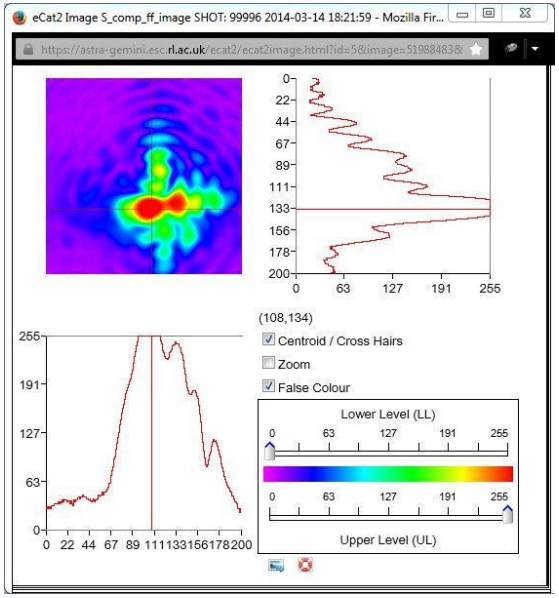


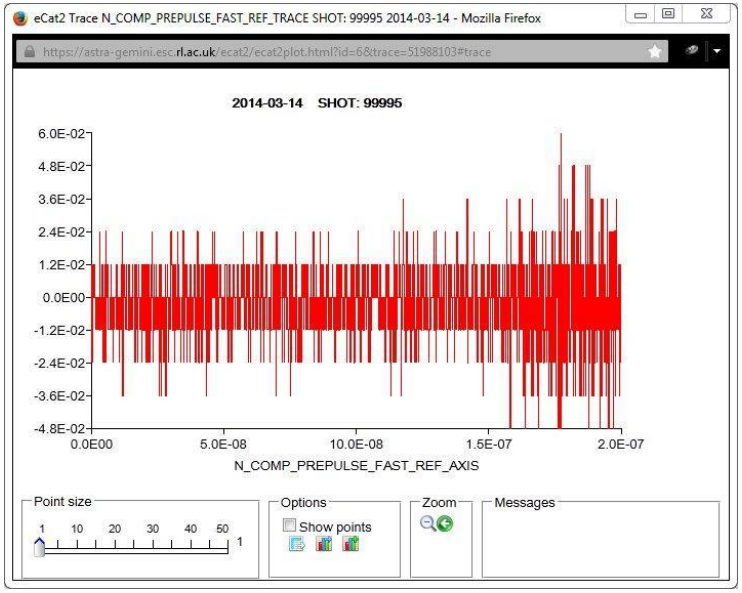
Science & Technology Facilities Council

Road Map

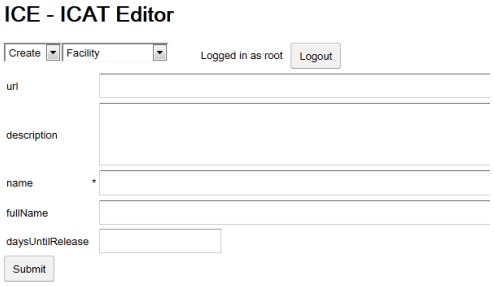
October 2014

- Add visualisation features (simple analysis of one and 2 dimensional data) by packaging such functionality already developed by the ICAT team at STFC for one of their facilities.





- Find replacement for GWT
- ICE (ICAT Editor) might be included in TopCAT



What Direction?

- What requirements do the community want for TopCAT?
- Is there really a use case for federation?
- Would it be better to provide a simple GUI for a single ICAT server?