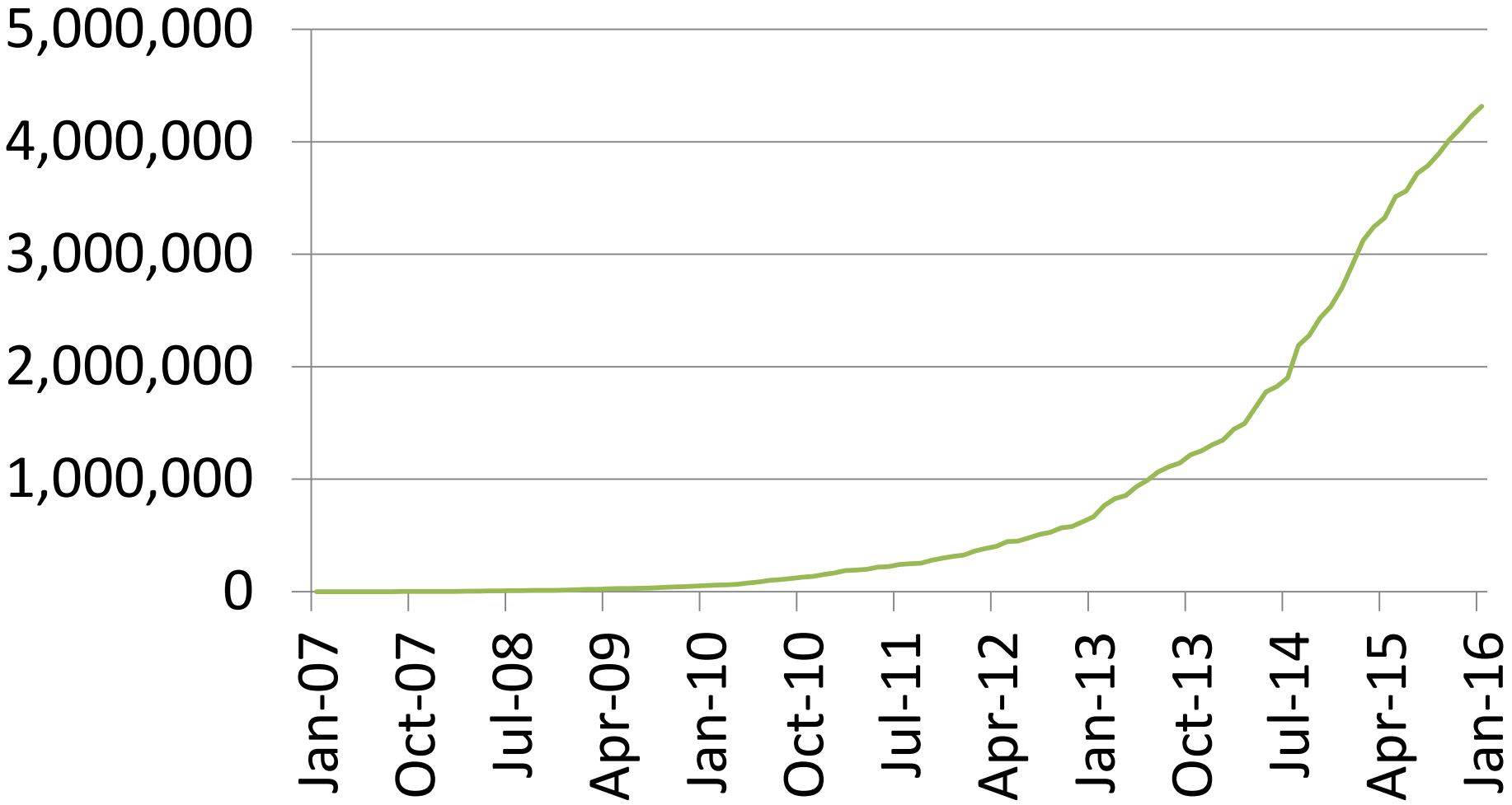
Remote Data Analysis Service for Photon Science

Calypso+ - JRA2

Diamond (and STFC)

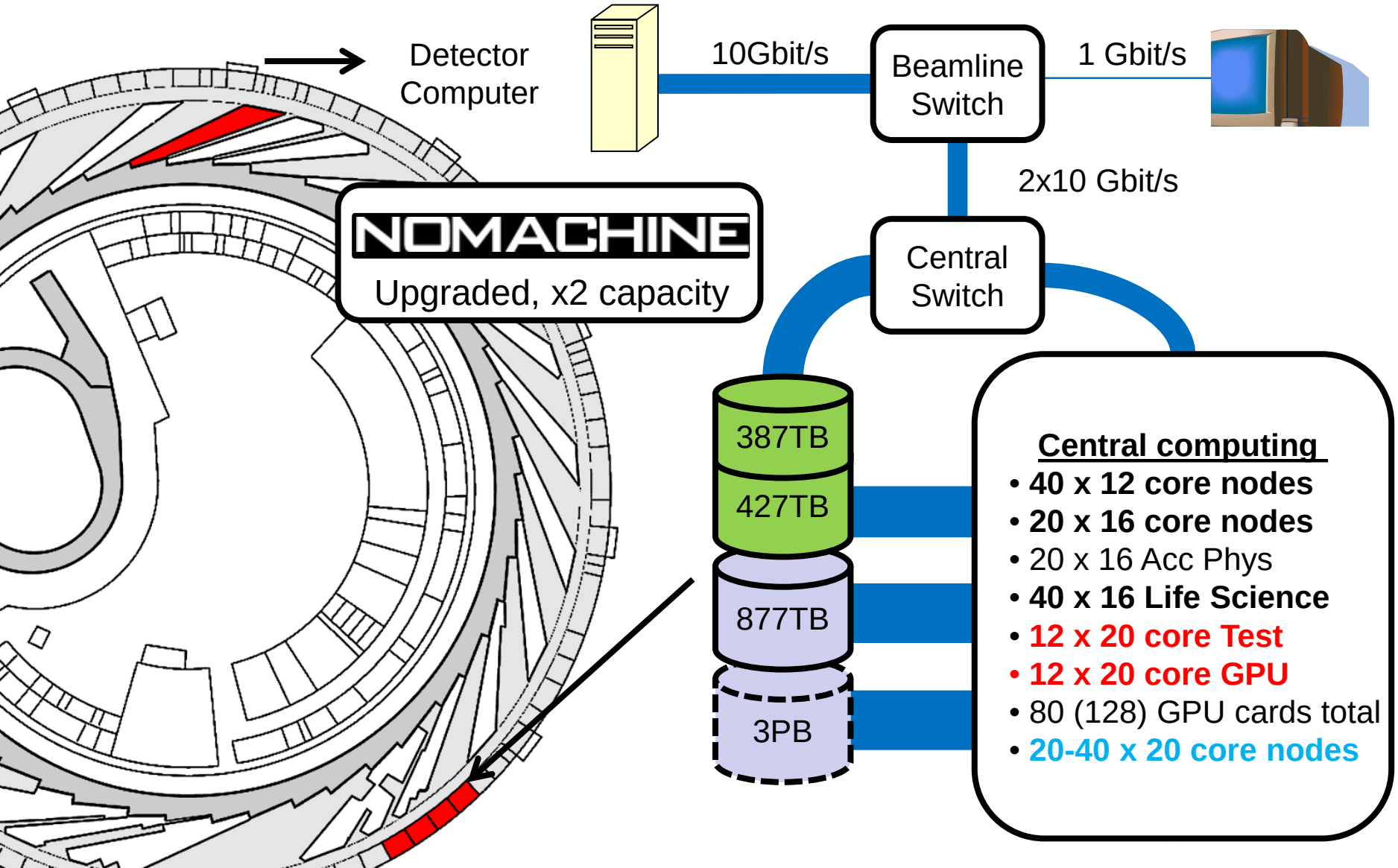
DLS Total Data Volumes (GB)

(data from archive)



diamond

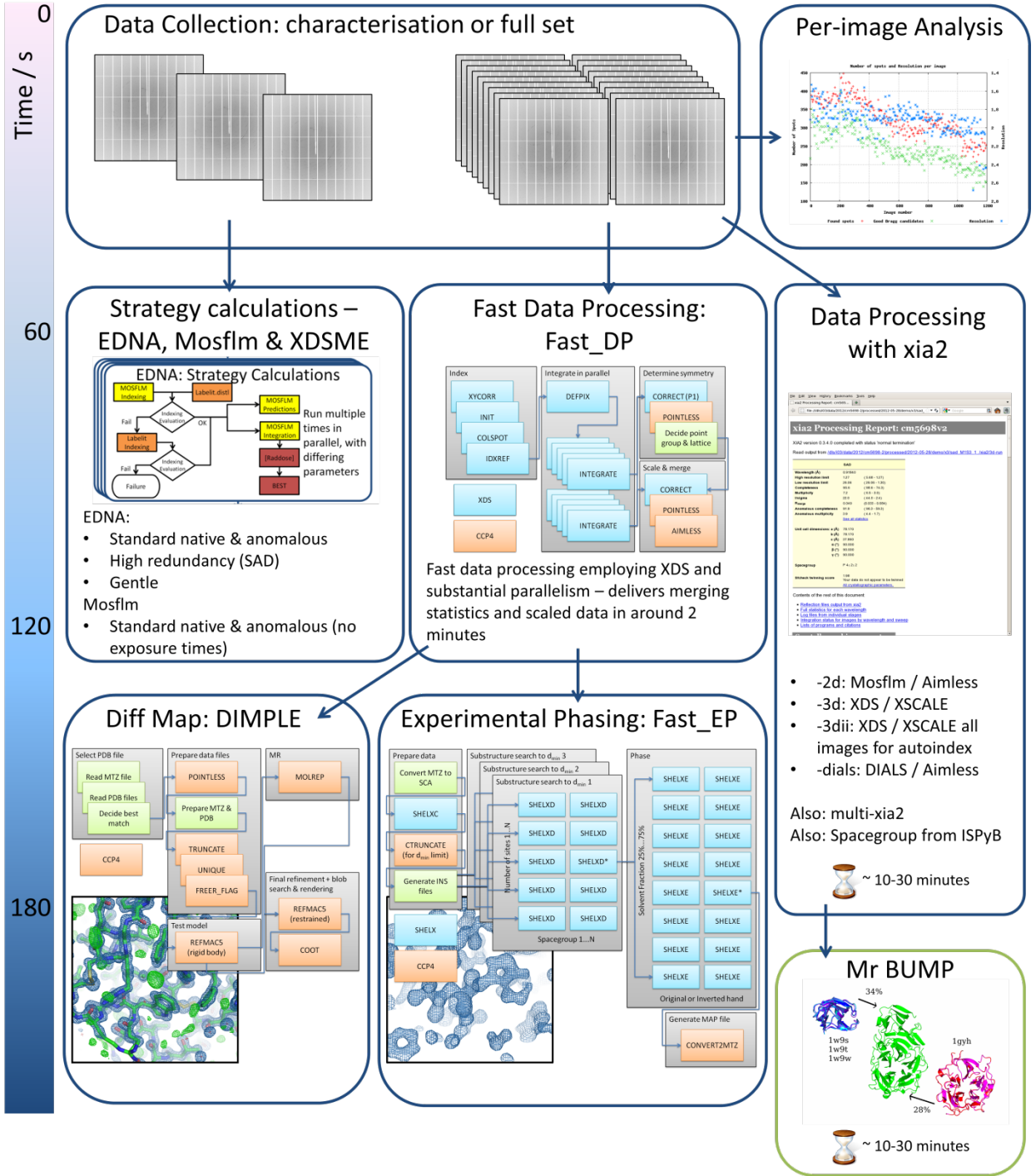
Outline of current computing resources



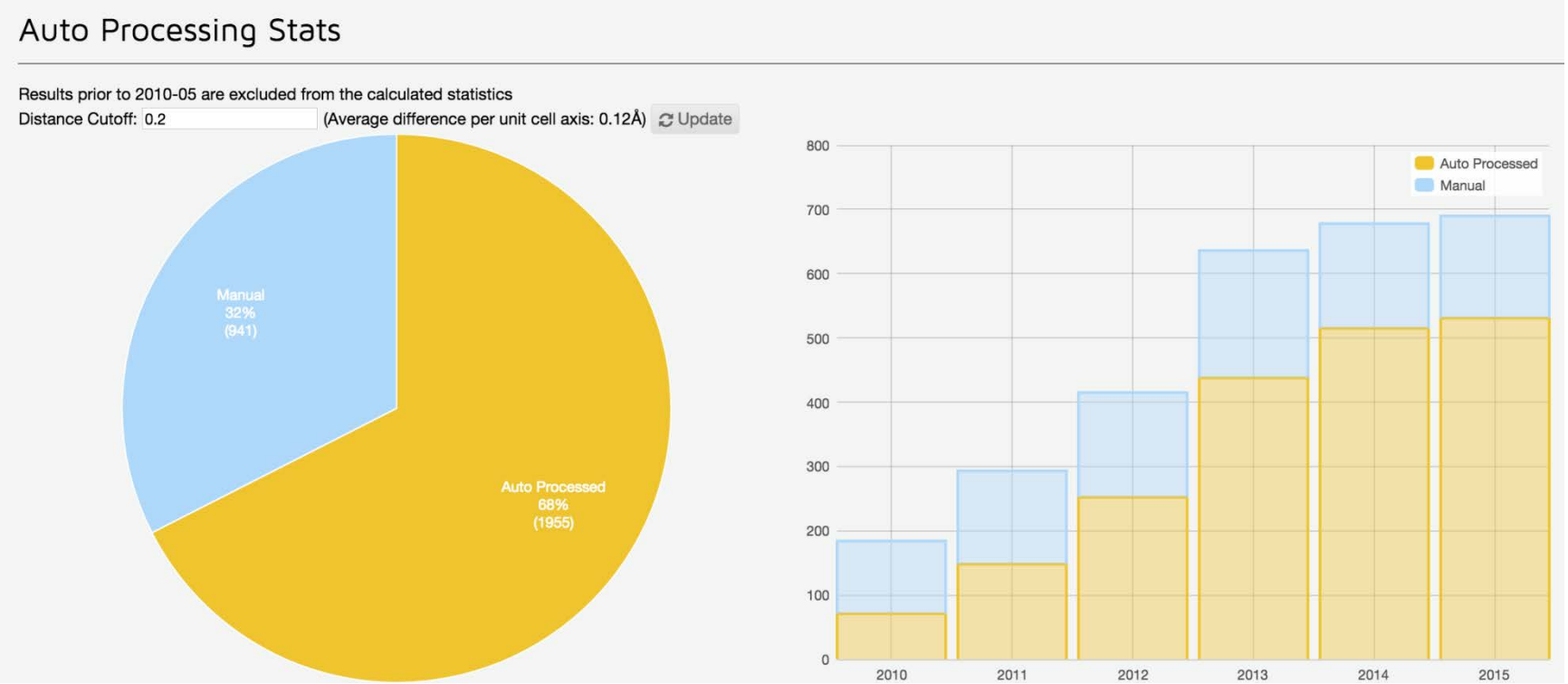
Data collection and experiment monitoring

Automated data reduction/analysis

Automated structure solution



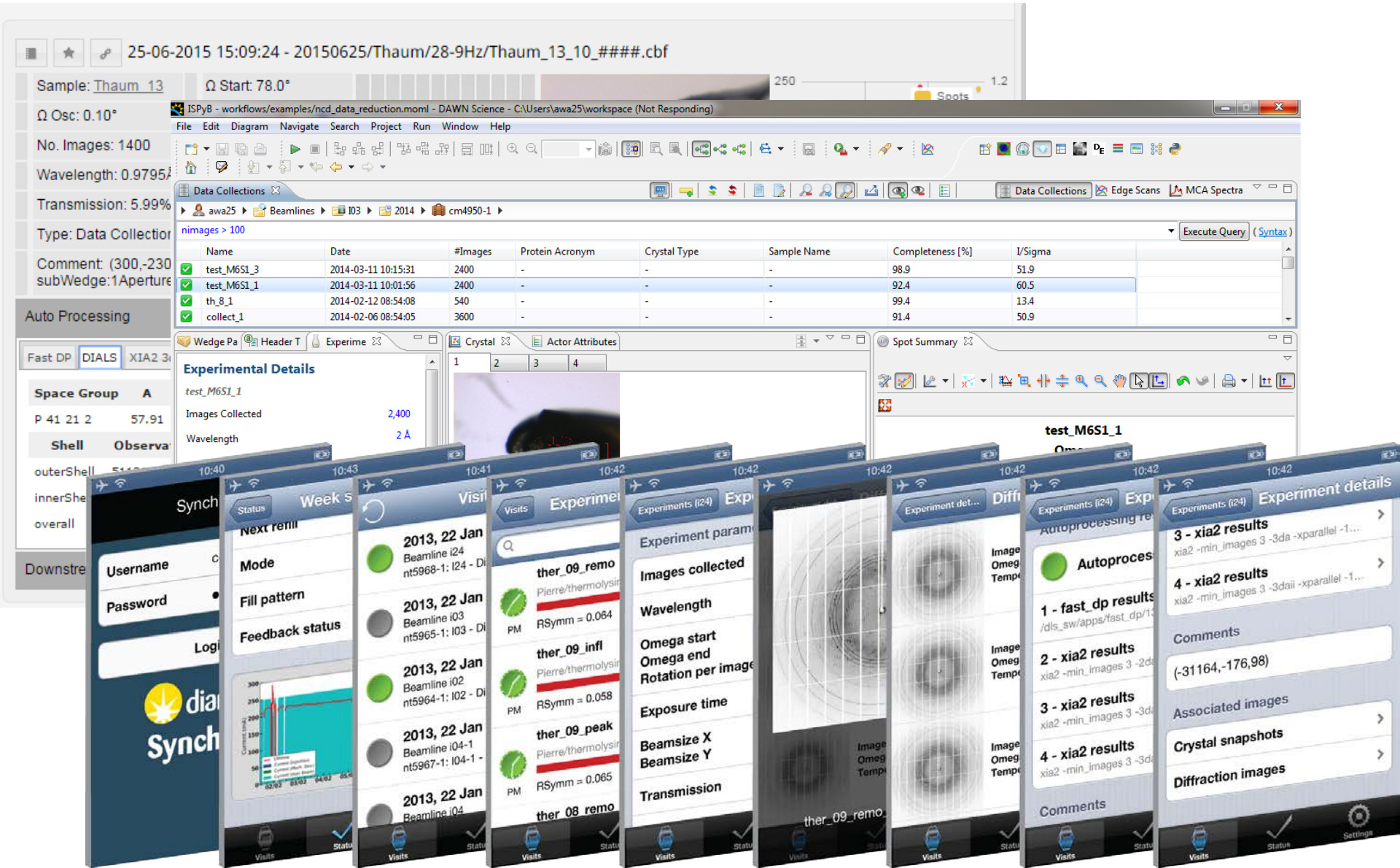
Value of Automatic Processing



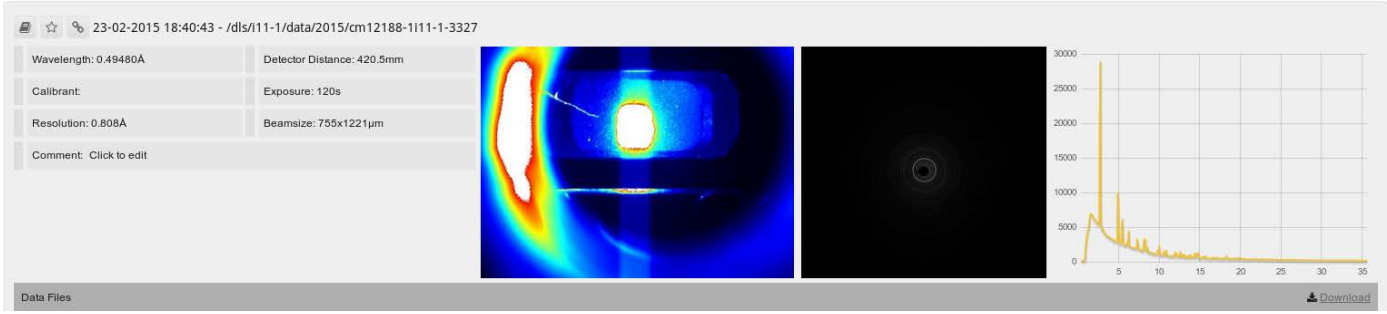
- 0.2 < distance metric = sqrt(sum((a-a*)^2+(b-b*)^2 …) for a,b,c,alpha,beta,gamma
- distance of 0.2 = average difference in each cell parameter of 0.0816A

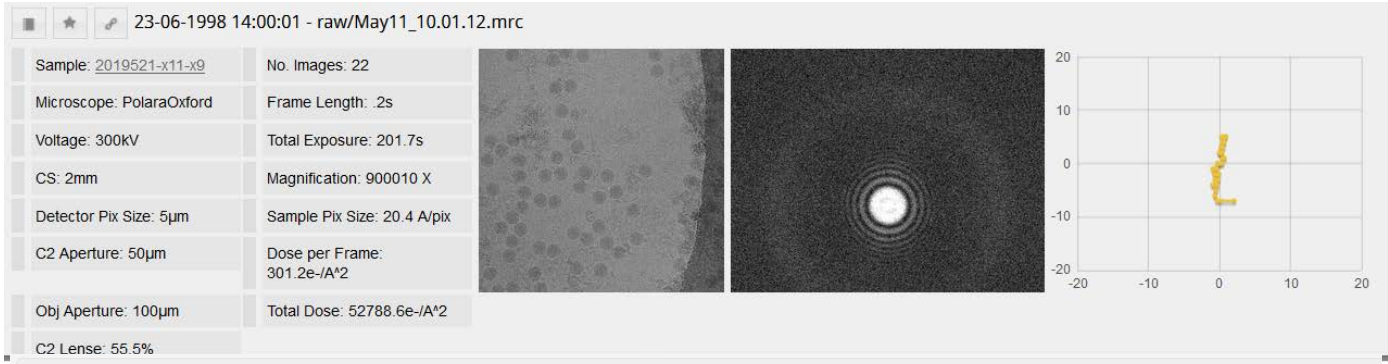
diamond

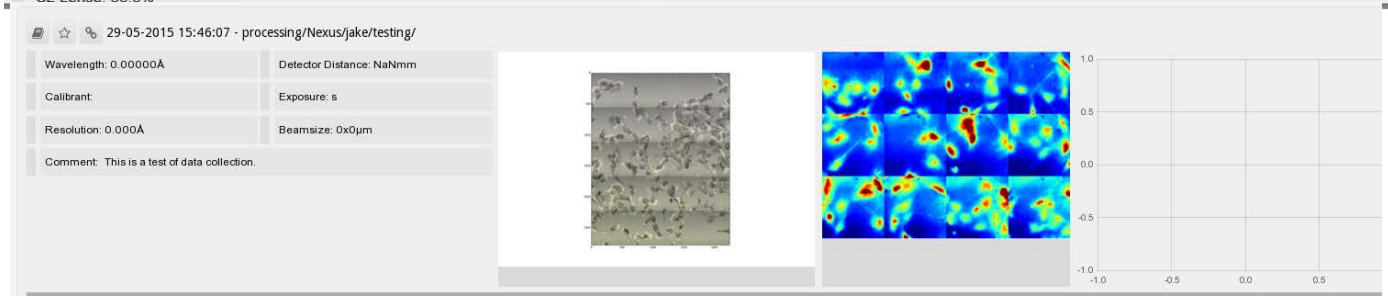
Interfaces


25-06-2015 15:09:24 - 20150625/Thaum/28-9Hz/Thaum_13_10_####.cbf
Sample: Thaum_13
Ω Start: 78.0°
Ω Osc: 0.10°
No. Images: 1400
Wavelength: 0.9795
Transmission: 5.99%
Type: Data Collection
Comment: (300,-230
subWedge:1Aperture
Auto Processing
Fast DP
DIALS
Space Group
P 41 21 2
57.91
Shell
outerShell
innerShe
overall
Downstre
ISPyB - workflows/examples/ncd_data_reduction.moml - DAWN Science - C:\Users\awa25\workspace (Not Responding)
File Edit Diagram Navigate Search Project Run Window Help
Data Collections
Edge Scans
MCA Spectra
awa25 Beamlines I03 2014 cm4950-1
nimages > 100
Execute Query
Syntax
Name
Date
#Images
Protein Acronym
Crystal Type
Sample Name
Completeness [%]
I/Sigma
test_M6S1_3 2014-03-11 10:15:31 2400 98.9 51.9
test_M6S1_1 2014-03-11 10:01:56 2400 92.4 60.5
th_8_1 2014-02-12 08:54:08 540 99.4 13.4
collect_1 2014-02-06 08:54:05 3600 91.4 50.9
Experimental Details
test_M6S1_1
Images Collected 2,400
Wavelength 2 Å
Crystal
Actor Attributes
Spot Summary
test_M6S1_1
Synch
Username
Password
Mode
Fill pattern
Feedback status
Week
Visits
Experiments (i24)
Experiment details
Images collected
Wavelength
Omega start
Omega end
Rotation per image
Exposure time
Beamsize X
Beamsize Y
Transmission
ther_09_remo
ther_09_infl
ther_09_peak
ther_08_remo
RSymm = 0.064
RSymm = 0.058
RSymm = 0.065
2013, 22 Jan
Beamline i24
Beamline i03
Beamline i02
Beamline i04-1
Autoproces
1 - fast_dp results
2 - xia2 results
3 - xia2 results
4 - xia2 results
Comments
(-31164,-176,98)
Associated images
Crystal snapshots
Diffraction images
Settings

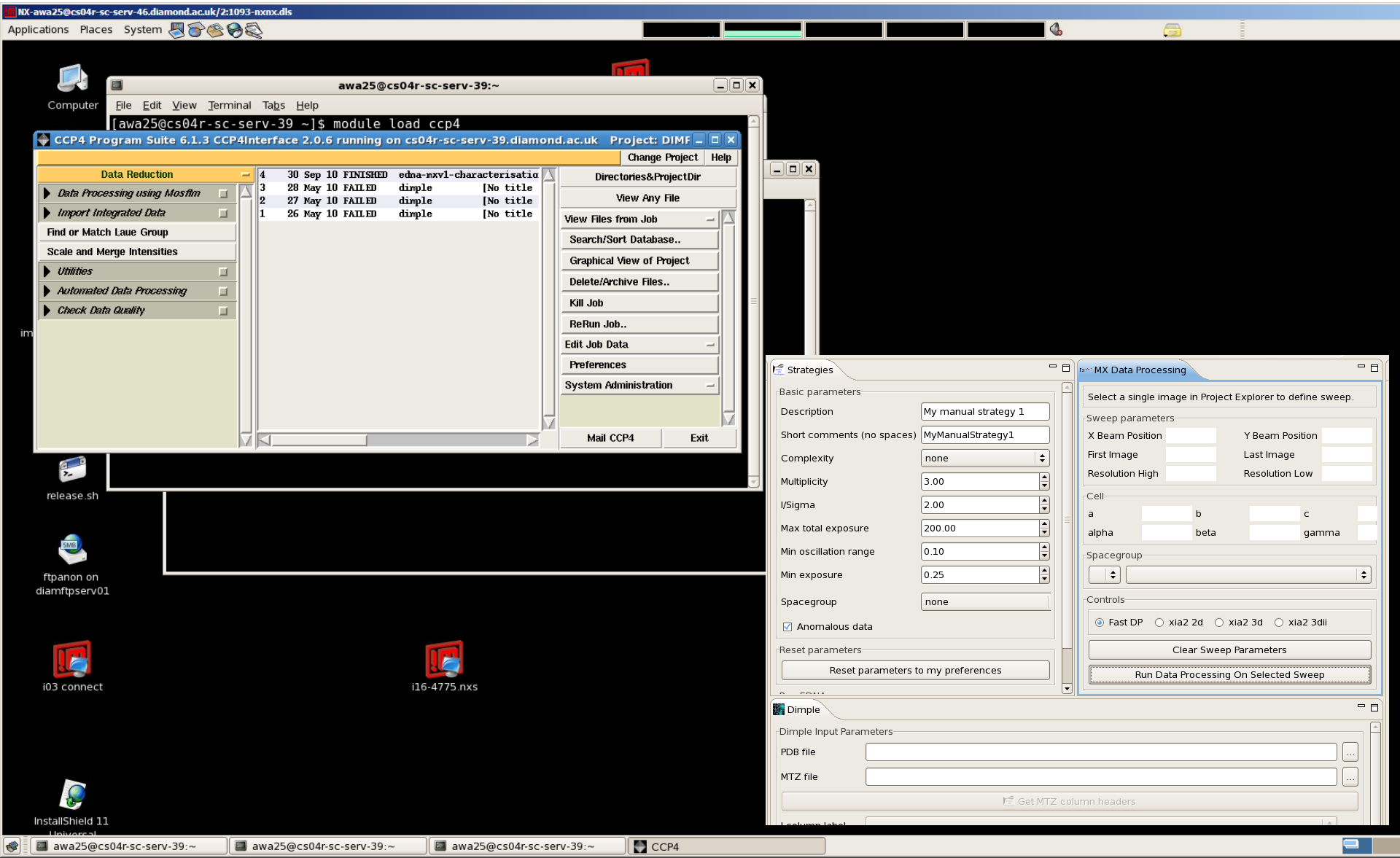
- Powder Diffraction (LDE)
- Tomography
- EM
- Infrared



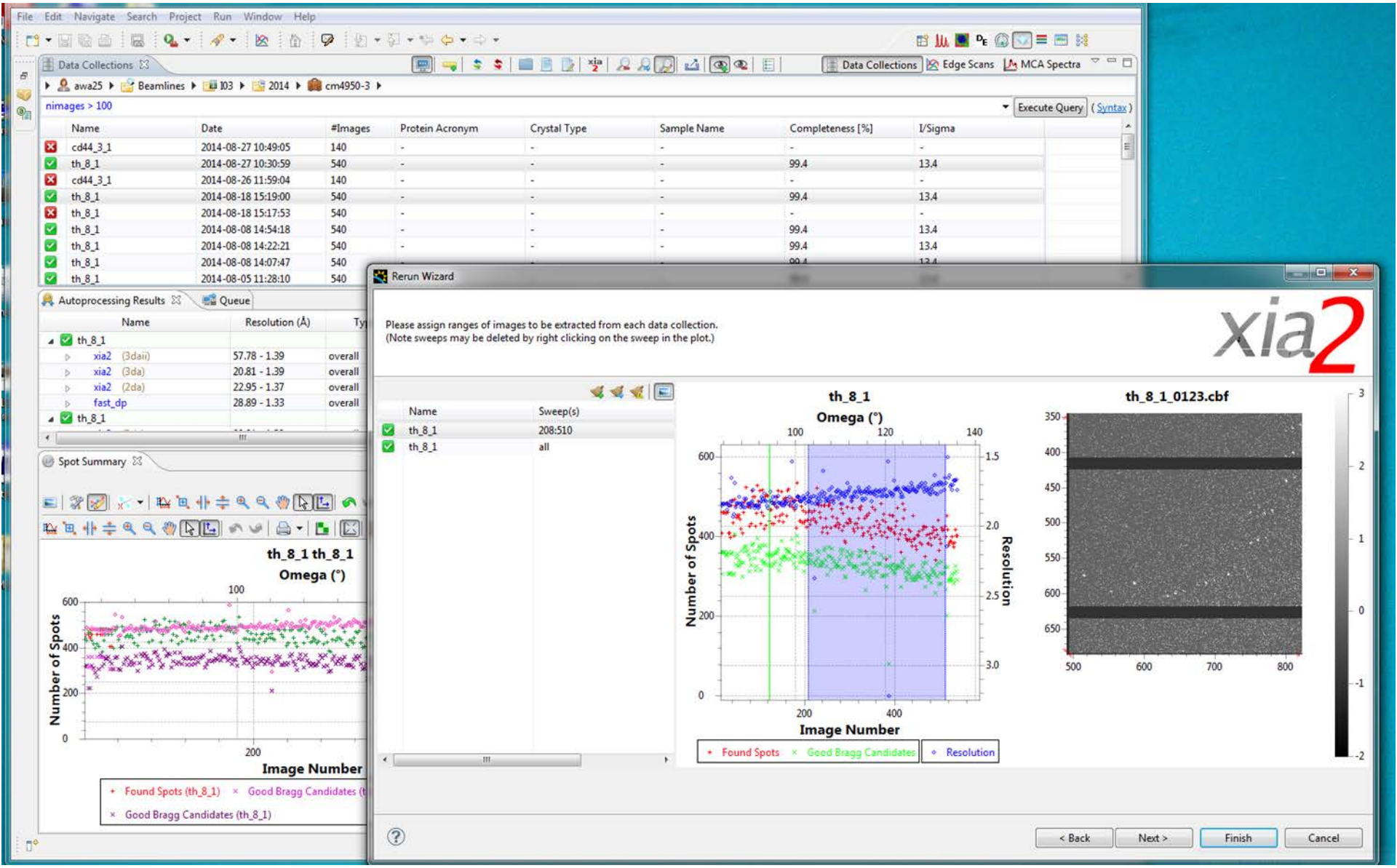




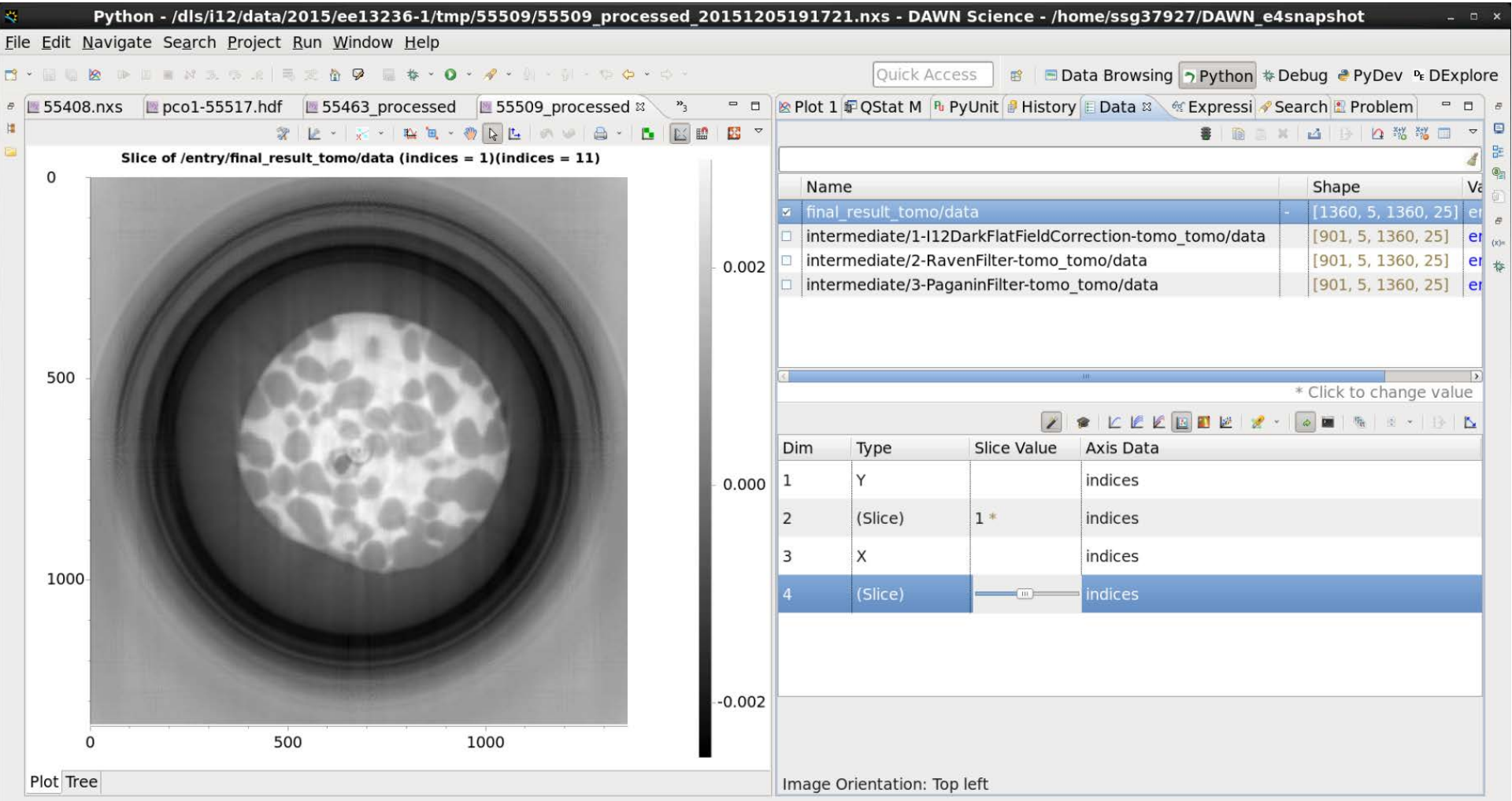
Remote Reprocessing



ISPyB– MX LIMS interface (also in SynchWeb)

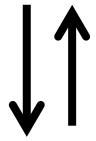


Reconstructed data (in collaboration with STFC and ISIS)

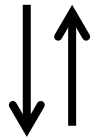


The framework: Savu

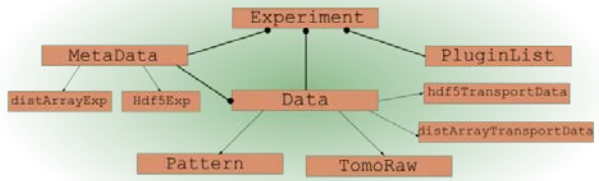
Data Layer

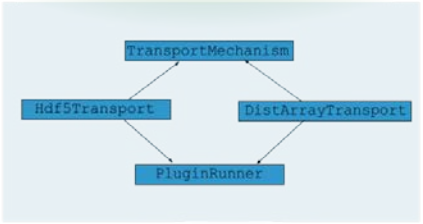


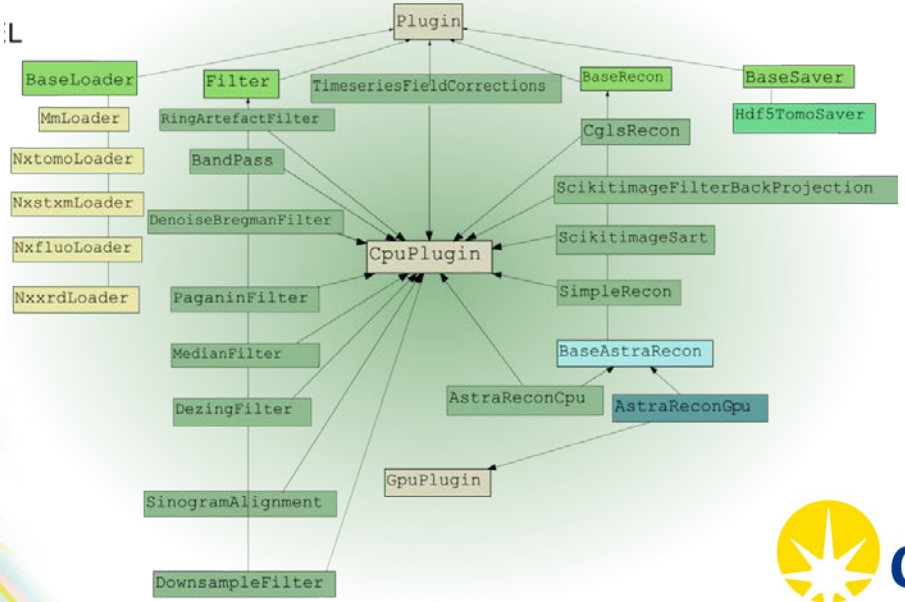
Control Layer



Plugin Layer





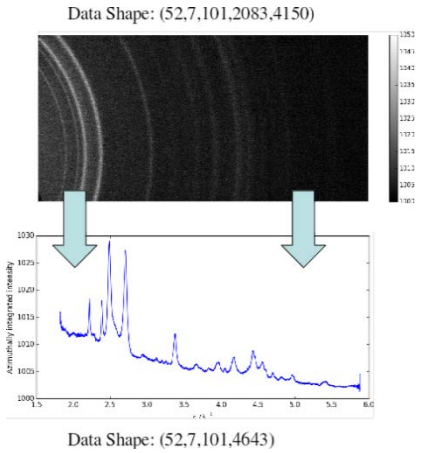


diamond

Multi-modal data processing

No other existing pipeline can perform the whole process.

XRD reduction:

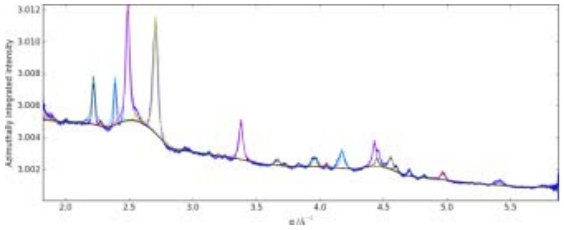


- 1.2 TB to ~2 GB
- 7 hours single threaded.
- 8 minutes using Savu with 72 cores (over 4 nodes).

Background removal:

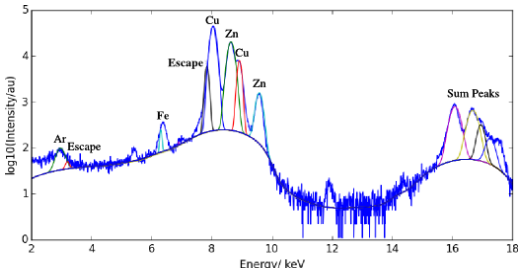
- 1.5 hours single threaded.
- 90s using Savu.

XRD peak fitting:



- 6-7 hours single threaded.
- 15 minutes using Savu.

XRF Peak fitting:

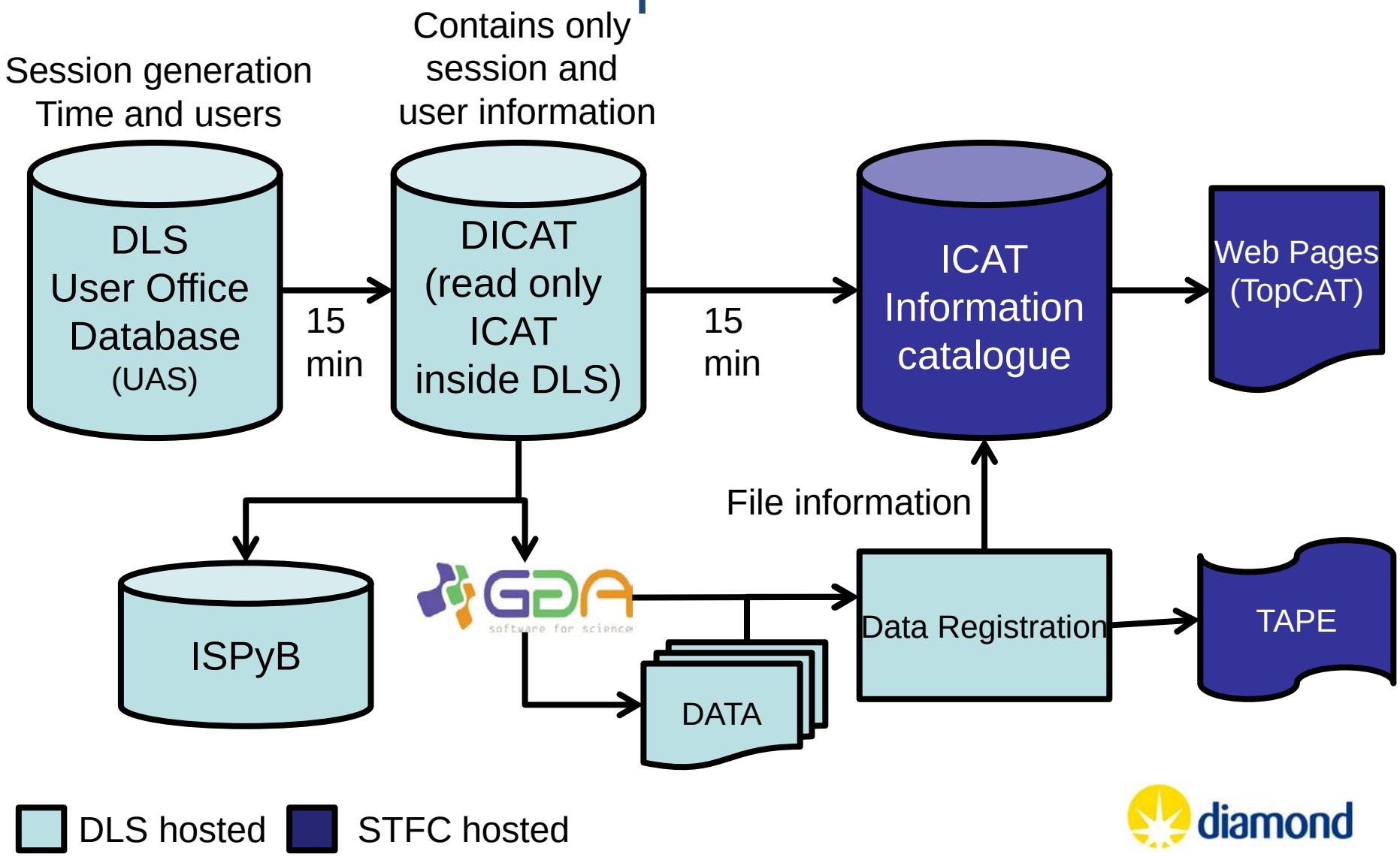


- 2 hours single threaded.
- 5 minutes using Savu.

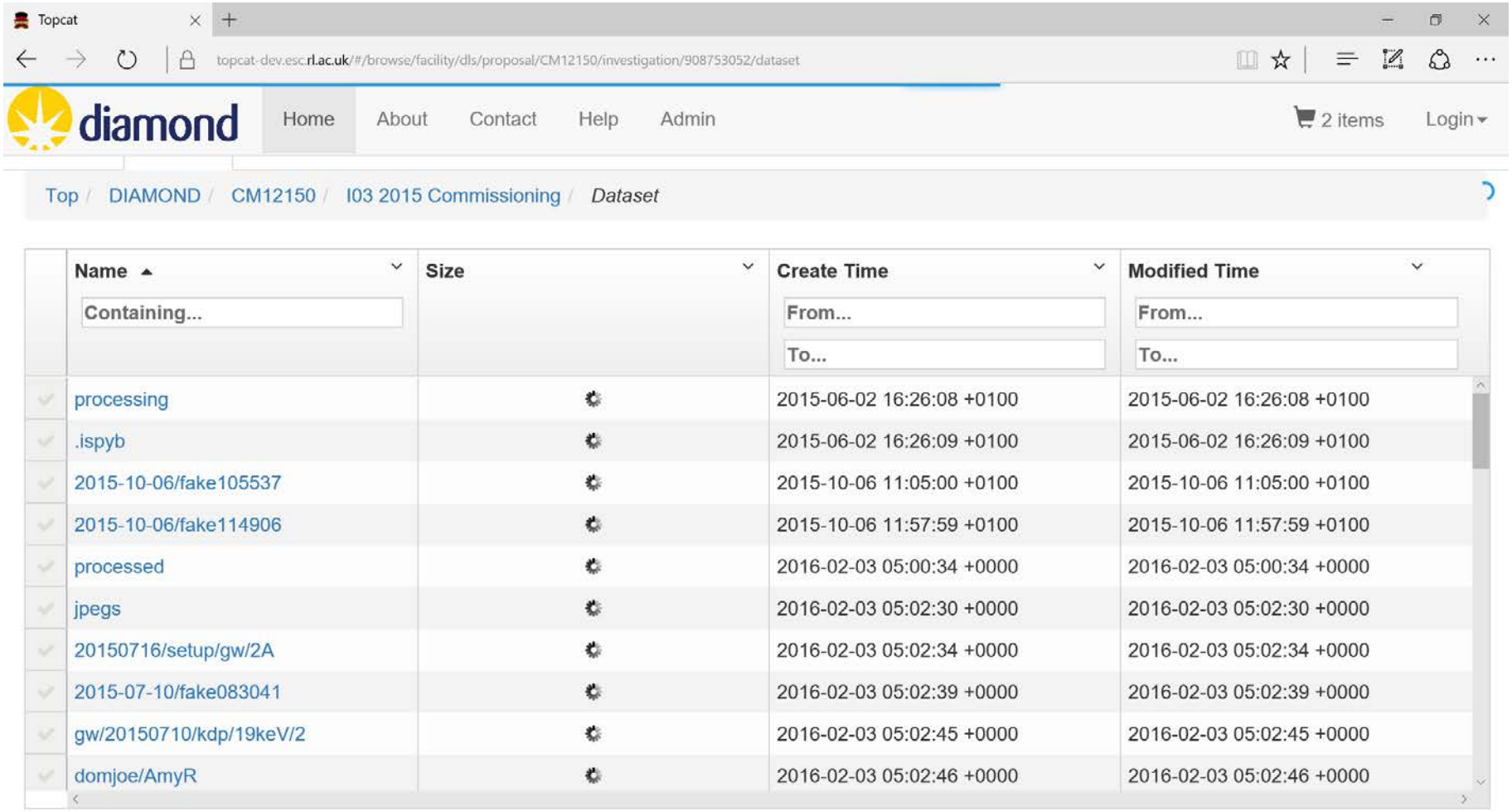
Computing/Software Support Groups

Group	Leader
Beamline Controls	Nick Rees
Data Acquisition (DAQ)	Rob Walton
Data Analysis (Scisoft)	Alun Ashton
User Office development	Bill Pulford & Sam Hough
Scientific Computing	Greg Mathews
STFC Scientific Computing	Tom Griffin and Brian Matthews
Resources that could be involved	…….

Data is captured from every stage of the process



Access to archived (http or Globus) or and re-staging @DLS or @STFC.



List of expertise needed

1. **Project Coordination + workshops** (PSI/ESRF obviously)
   - **Workshops welcome in DLS**

List of expertise needed

1. **Use** (**science**) **case** definition and collation (multiple disciplines, facilities and user/industry) (Scientists)

But don't these need to be Physical Sciences?

- **Lots here, e.g. contribution to PanDAAS including industry**
- **CCP4 cloud?**
- **ULTRA tomography**
- **EM**

Packaging for applications

- If this is the technology exploration then
  - STFC have much experience that DLS is drawing on.
    - (docker, Shifter, cvmfs, quattor, aquilon…..)
    - Docker or quattor and cvmfs would seem best candidates for this kind of prototype.

diamond

Cloud setup and deployment

1. STFC utilises OpenNebula with a transition planned to OpenStack (within the year)
    1. ~1000 cores and 750Tb currently
    2. Increase of 50% in progress.

Configuration of site and test sites

Port and **package applications** and example

- **CCP4 cloud?**
- **ULTRA tomography**
- **EM**
- **Catalysis hub**

Umbrella authentication (AAI+security)

1. **Ascent?**
    1. **UK has extensive experience (Jens and Stefan)**

User portal development (Web engineers)

- How long is a piece of string?!
  - Remote desktop, portal, icatportal, CCP4 portal…..

Definition, **development** and **distribution** of mini demonstrator platform to test sites

1. Actual deployment. Marketplace? Egi applications marketplace https://appdb.egi.eu/

2. Overlap with the portal?

Report on results and links to HNSciCloud + EOSC

1. STFC are involved