

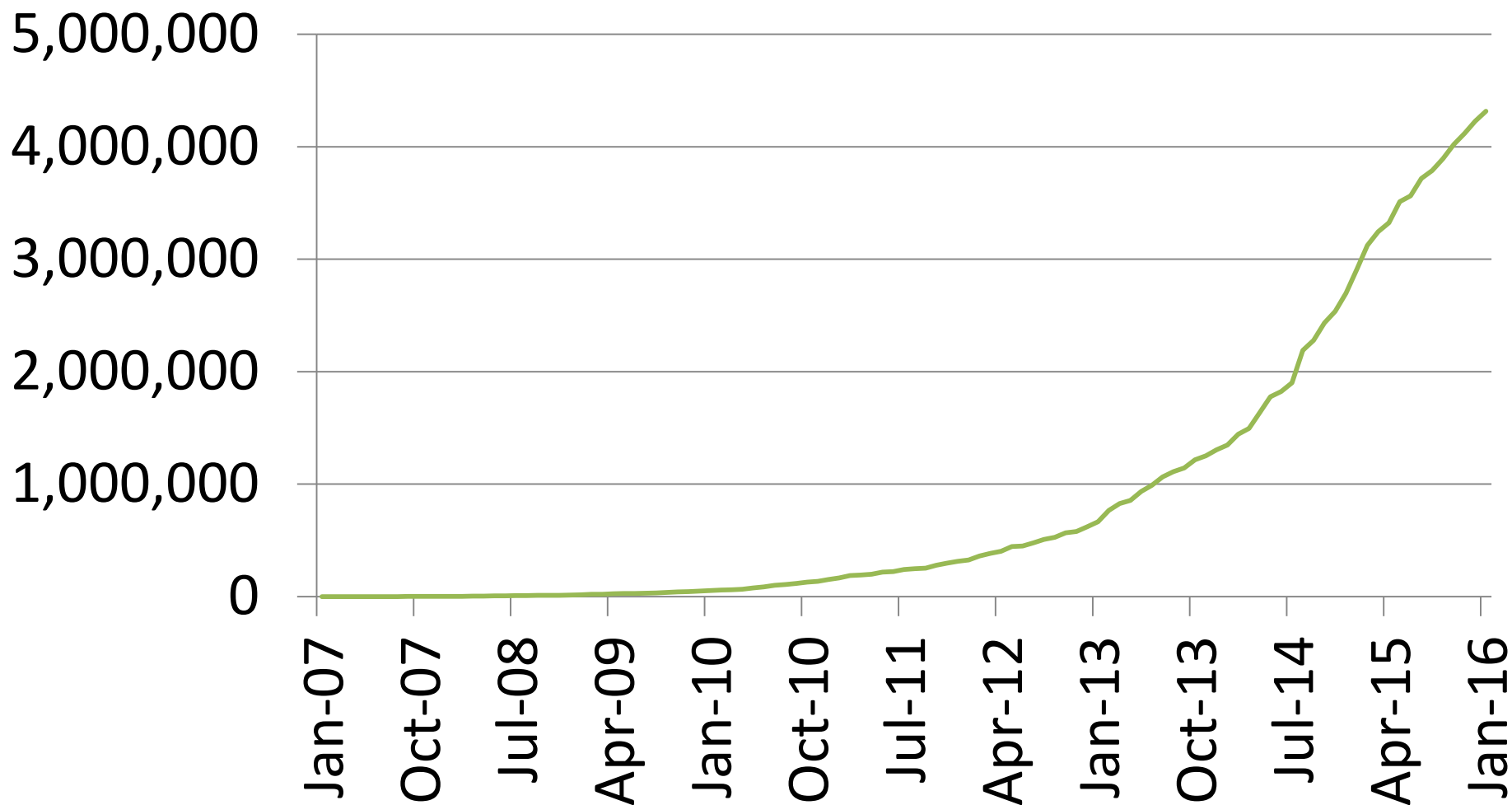
Remote Data Analysis Service for Photon Science

Calypso+ - JRA2

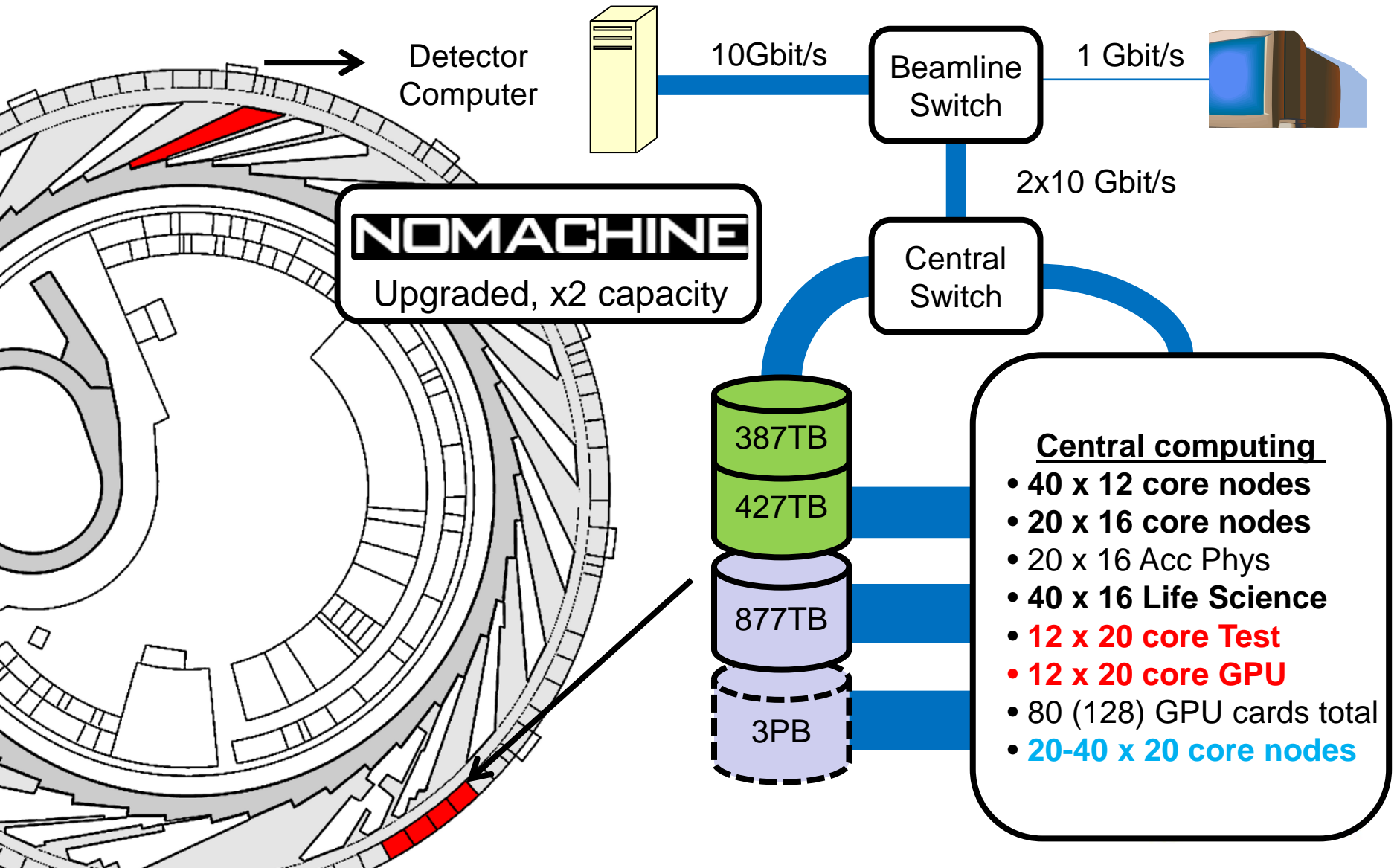
Diamond (and STFC)

DLS Total Data Volumes (GB)

(data from archive)



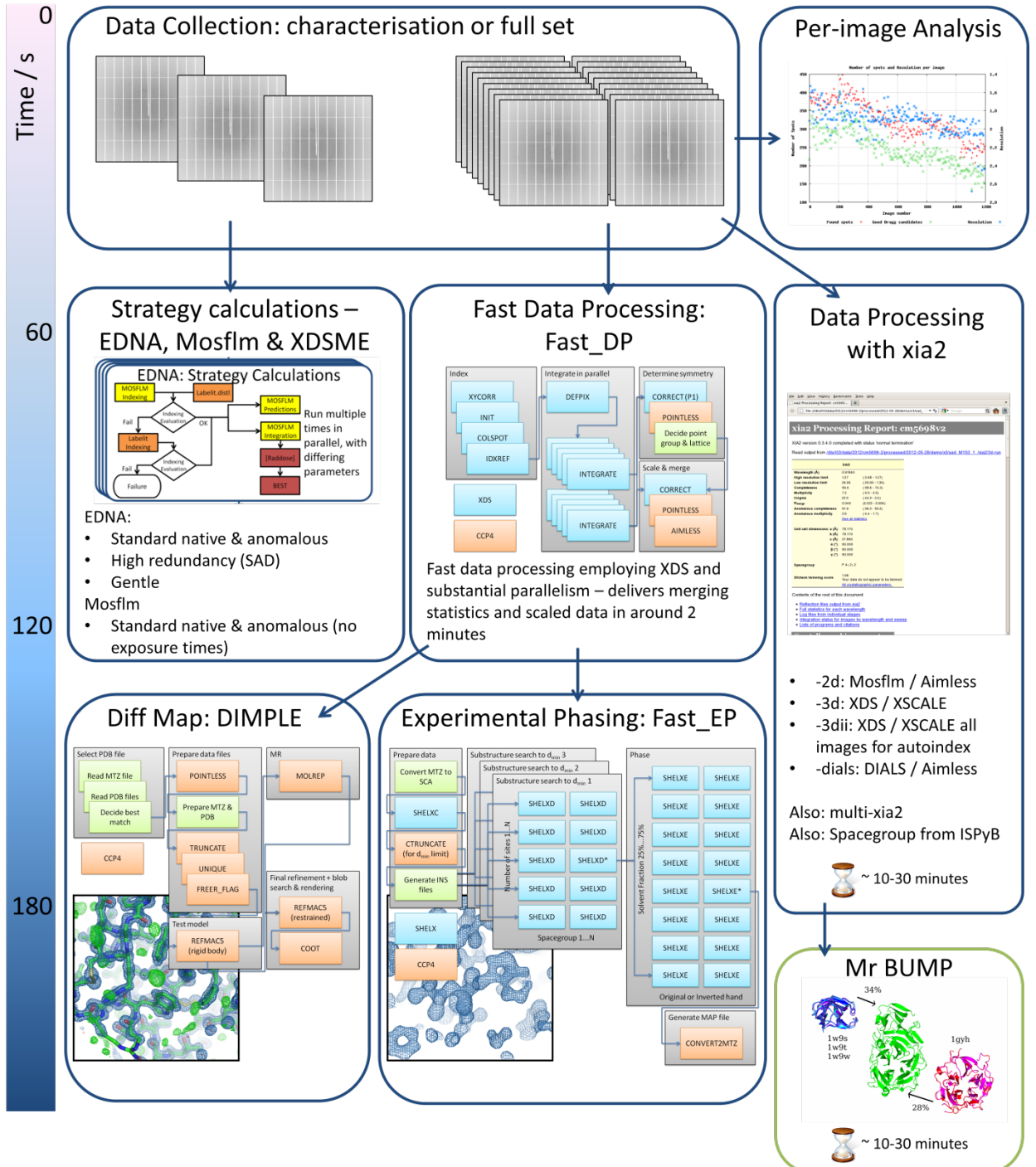
Outline of current computing resources



Data collection and experiment monitoring

Automated data reduction/analysis

Automated structure solution

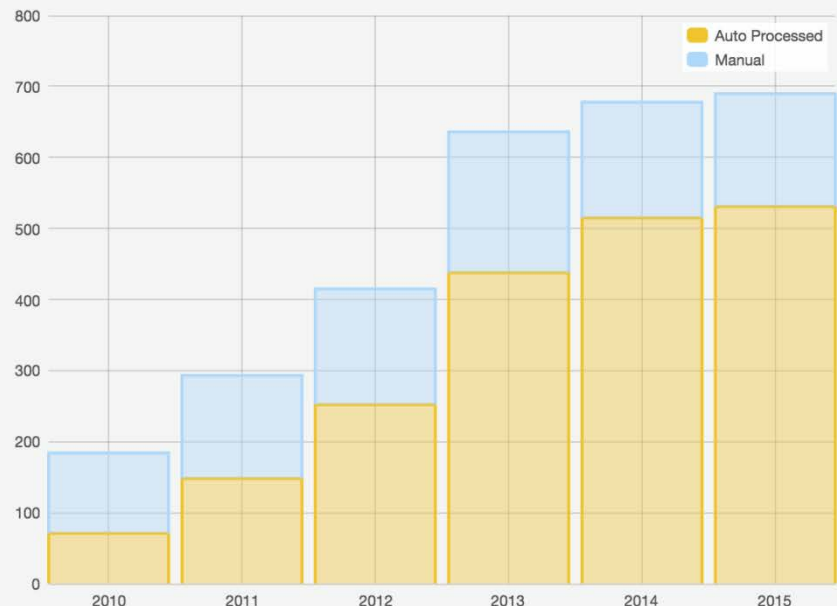
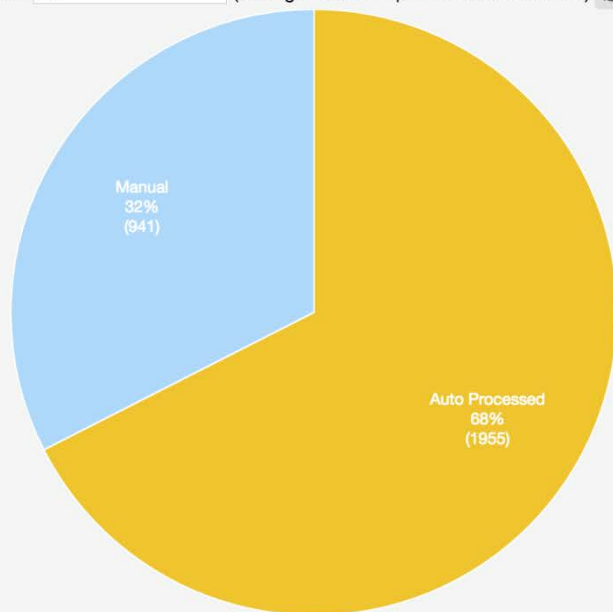


Value of Automatic Processing

Auto Processing Stats

Results prior to 2010-05 are excluded from the calculated statistics

Distance Cutoff: 0.2 (Average difference per unit cell axis: 0.12Å) [Update](#)



- $0.2 < \text{distance metric} = \sqrt{\sum((a-a^*)^2 + (b-b^*)^2 \dots)}$
for a,b,c,alpha,beta,gamma
- distance of 0.2 = average difference in each cell parameter of 0.0816Å

Interfaces

25-06-2015 15:09:24 - 20150625/Thaum/28-9Hz/Thaum_13_10_####.cbf

Sample: Thaum 13 Ω Start: 78.0° 250 Spots 1.2

Ω Osc: 0.10°

No. Images: 1400

Wavelength: 0.9795 Å

Transmission: 5.99%

Type: Data Collector

Comment: (300,-230) subWedge:1Aperture

Auto Processing

Fast DP **DIALS** XIA2 3

Space Group **A**

P 41 21 2 57.91

Shell **Observa**

outerShell 511

innerShell 511

overall

Downstre

ISPyB - workflows/examples/ncd_data_reduction.moml - DAWN Science - CAUsers\awa25\workspace (Not Responding)

File Edit Diagram Navigate Search Project Run Window Help

Data Collections

awa25 > Beamlines > I03 > 2014 > cm4950-1

nimages > 100

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

Execute Query (Syntax)

Wedge Pa Header T Experi Crystal Actor Attributes

Spot Summary

test_M6S1_1

Experimental Details

test_M6S1_1

Images Collected 2,400

Wavelength 2 Å

1 2 3 4

10:40 10:43 10:41 10:42 10:42 10:42 10:42 10:42

Synchrotron

Username

Password

Log

dia

Synch

Status

Week s

Next refill

Mode

Fill pattern

Feedback status

Visits

2013, 22 Jan

Beamline i24

nt5968-1: i24 - Di

2013, 22 Jan

Beamline i03

nt5965-1: i03 - Di

2013, 22 Jan

Beamline i02

nt5964-1: i02 - Di

2013, 22 Jan

Beamline i04-1

nt5967-1: i04-1 -

2013, 22 Jan

Beamline i04

ther_09_remo

Pierre/thermolysir

RSymm = 0.064

ther_09_infl

Pierre/thermolysir

RSymm = 0.058

ther_09_peak

Pierre/thermolysir

RSymm = 0.065

ther 08 remo

Experiment param

Images collected

Wavelength

Omega start

Omega end

Rotation per image

Exposure time

Beamsize X

Beamsize Y

Transmission

Image Omega Temp

Image Omega Temp

Image Omega Temp

Image Omega Temp

Experiment det...

Diff

Experiments (24) Exp

Autoprocessing re

Autoproces

1 - fast_dp results

/dis_sw/apps/fast_dp/1

2 - xia2 results

xia2 -min_images 3 -2d

3 - xia2 results

xia2 -min_images 3 -3d

4 - xia2 results

xia2 -min_images 3 -3d

Comments

(-31164,-176,98)

Associated images

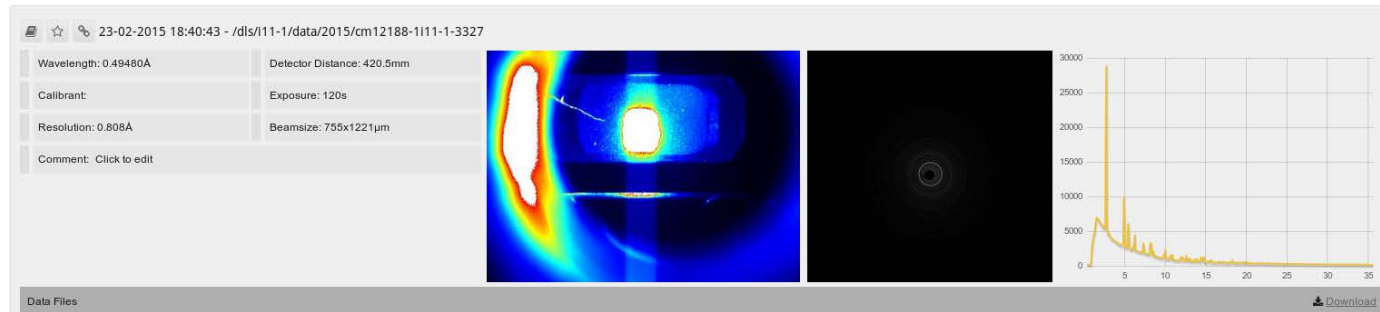
Crystal snapshots

Diffraction images

Comments

Settings

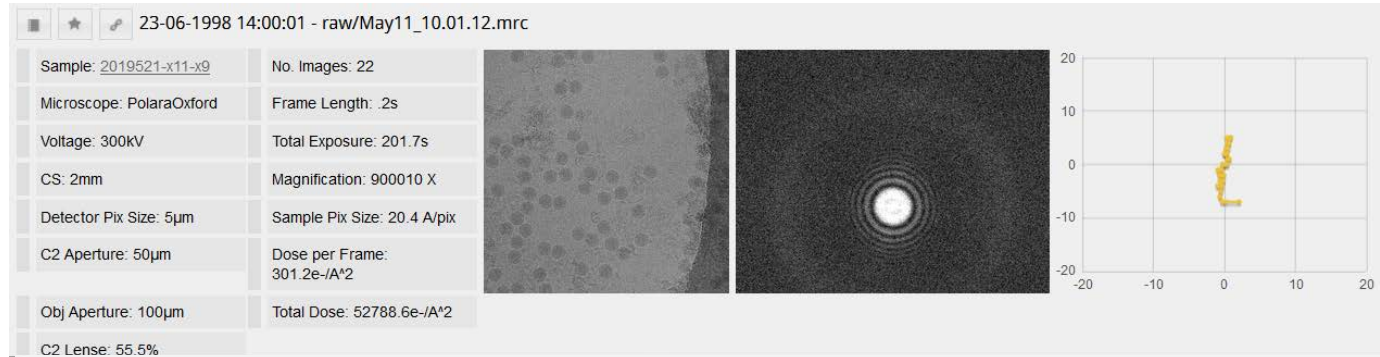
Powder Diffraction (LDE)



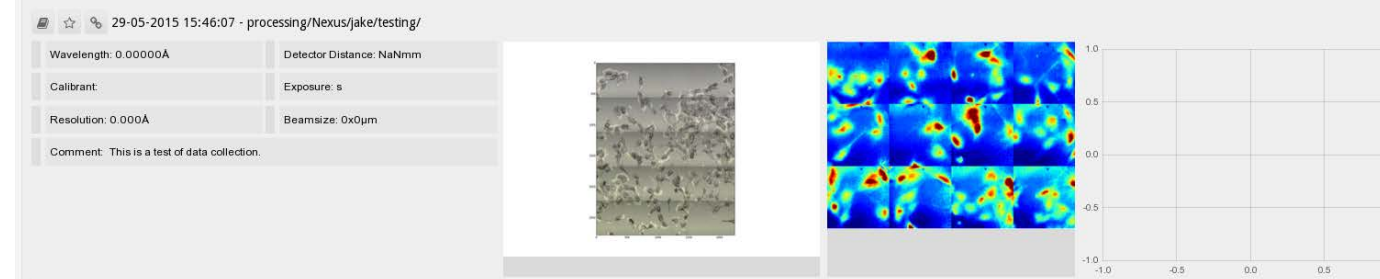
Tomography



EM



Infrared



Remote Reprocessing

The screenshot displays a remote desktop environment with the following components:

- Terminal Window:** Shows the command `module load ccp4` being executed in a shell.
- CCP4 Program Suite 6.1.3 CCP4Interface 2.0.6:** The main application window, titled "Project: DIMF". It features a sidebar with categories like "Data Reduction", "Data Processing using Mosflm", "Import Integrated Data", "Find or Match Laue Group", "Scale and Merge Intensities", "Utilities", "Automated Data Processing", and "Check Data Quality". The main area contains a table of job logs:

Job ID	Date	Status	Job Name	Title
4	30 Sep 10	FINISHED	edna-mxv1-characterisatio	
3	28 May 10	FAILED	dimple	[No title]
2	27 May 10	FAILED	dimple	[No title]
1	26 May 10	FAILED	dimple	[No title]

- Directories&ProjectDir:** A panel on the right side of the CCP4 interface with options like "View Files from Job", "Search/Sort Database..", "Graphical View of Project", "Delete/Archive Files..", "Kill Job", "ReRun Job..", "Edit Job Data", "Preferences", and "System Administration".
- Strategies:** A dialog box for defining processing strategies. It includes fields for "Description" (My manual strategy 1), "Short comments", "Complexity", "Multiplicity", "I/Sigma", "Max total exposure", "Min oscillation range", "Min exposure", and "Spacegroup". It also has a "Reset parameters to my preferences" button.
- MX Data Processing:** A dialog box for selecting a single image and defining sweep parameters. It includes fields for "X Beam Position", "Y Beam Position", "First Image", "Last Image", "Resolution High", and "Resolution Low". It also has a "Run Data Processing On Selected Sweep" button.
- Dimple:** A dialog box for Dimple input parameters, including fields for "PDB file" and "MTZ file".

The desktop background shows several icons: "release.sh", "ftpanon on diamftpserv01", "i03 connect", and "i16-4775.nxs". The taskbar at the bottom shows multiple instances of the terminal and CCP4 windows.

ISPyB- MX LIMS interface (also in SynchWeb)

The screenshot displays the ISPyB- MX LIMS interface. The main window shows a table of data collections with columns for Name, Date, #Images, Protein Acronym, Crystal Type, Sample Name, Completeness [%], and I/Sigma. Below this, the Autoprocesing Results section shows a tree view of data collections and their resolutions. A Rerun Wizard dialog is open, prompting the user to assign ranges of images to be extracted from each data collection. The dialog includes a table for Name and Sweep(s), and two plots: a scatter plot of Number of Spots vs. Image Number and a resolution plot for a specific image.

Name	Date	#Images	Protein Acronym	Crystal Type	Sample Name	Completeness [%]	I/Sigma
cd44_3_1	2014-08-27 10:49:05	140	-	-	-	-	-
th_8_1	2014-08-27 10:30:59	540	-	-	-	99.4	13.4
cd44_3_1	2014-08-26 11:59:04	140	-	-	-	-	-
th_8_1	2014-08-18 15:19:00	540	-	-	-	99.4	13.4
th_8_1	2014-08-18 15:17:53	540	-	-	-	-	-
th_8_1	2014-08-08 14:54:18	540	-	-	-	99.4	13.4
th_8_1	2014-08-08 14:22:21	540	-	-	-	99.4	13.4
th_8_1	2014-08-08 14:07:47	540	-	-	-	99.4	13.4
th_8_1	2014-08-05 11:28:10	540	-	-	-	99.4	13.4

Name	Resolution (Å)	Type
th_8_1		
xia2 (3dai)	57.78 - 1.39	overall
xia2 (3da)	20.81 - 1.39	overall
xia2 (2da)	22.95 - 1.37	overall
fast_dp	28.89 - 1.33	overall

Name	Sweep(s)
th_8_1	208:510
th_8_1	all

th_8_1
Omega (°)

Number of Spots

Image Number

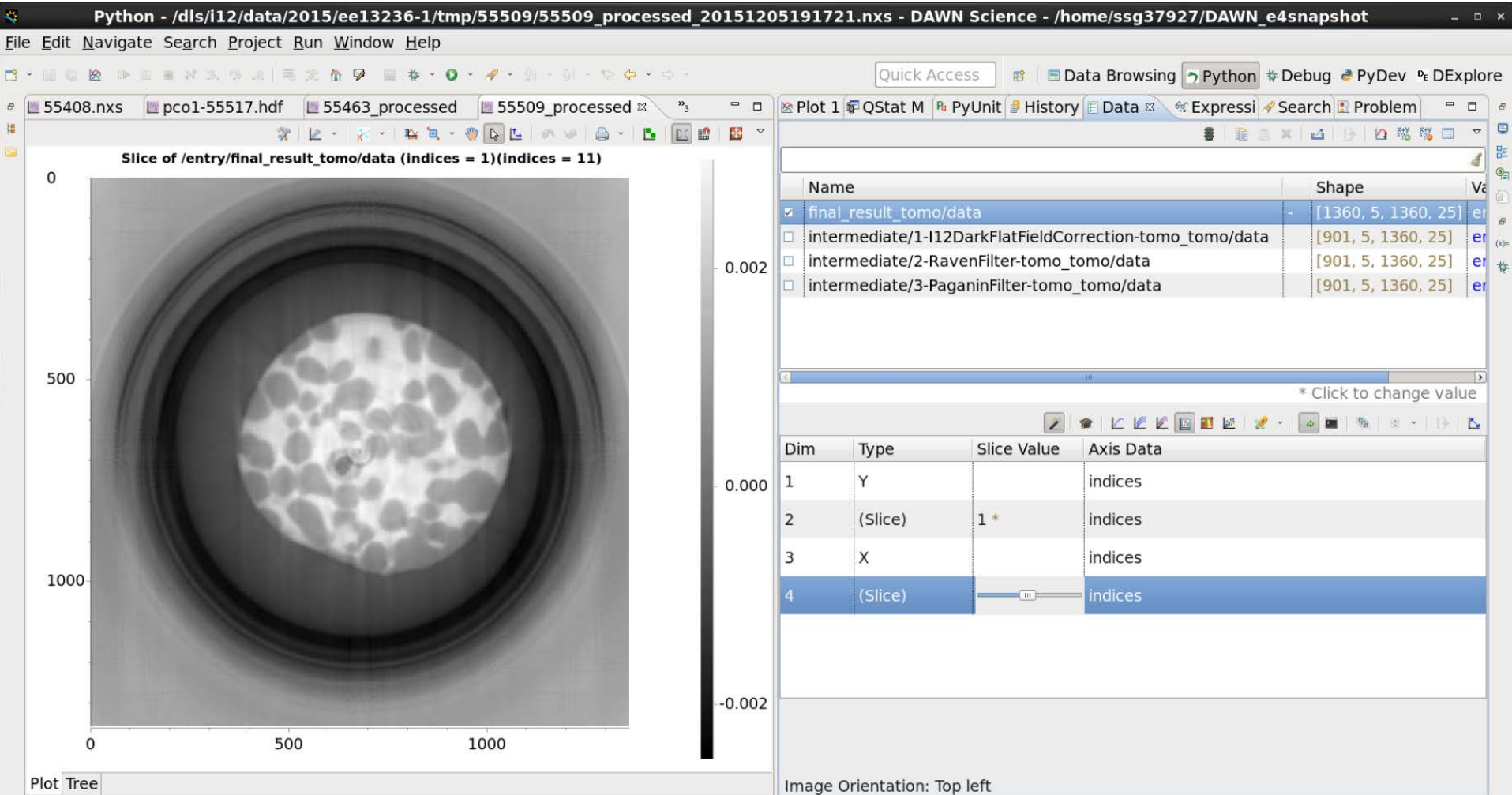
Legend: + Found Spots (th_8_1), x Good Bragg Candidates (th_8_1), • Resolution

th_8_1_0123.cbf

Resolution

Buttons: < Back, Next >, Finish, Cancel

Reconstructed data (in collaboration with STFC and ISIS)



The framework: Savu

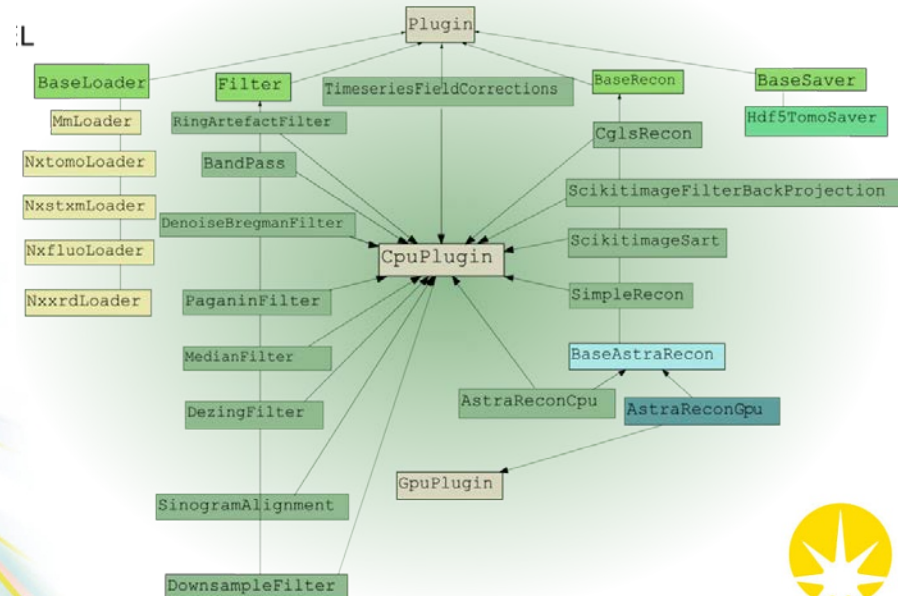
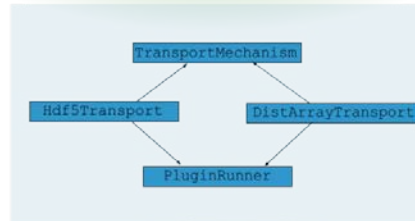
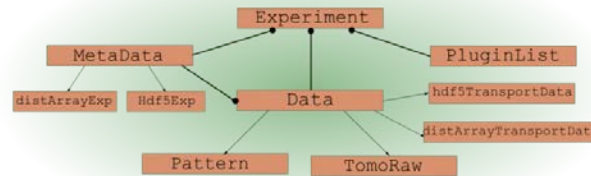
Data Layer



Control Layer



Plugin Layer

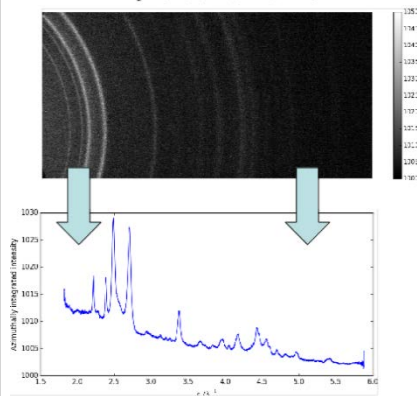


Multi-modal data processing

No other existing pipeline can perform the whole process.

XRD reduction:

Data Shape: (52,7,101,2083,4150)



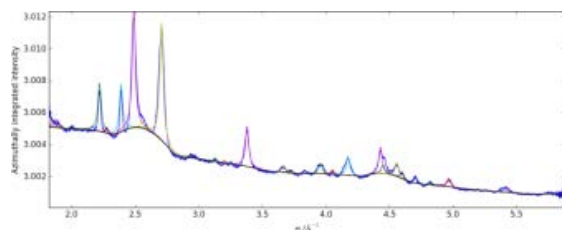
Data Shape: (52,7,101,4643)

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

Background removal:

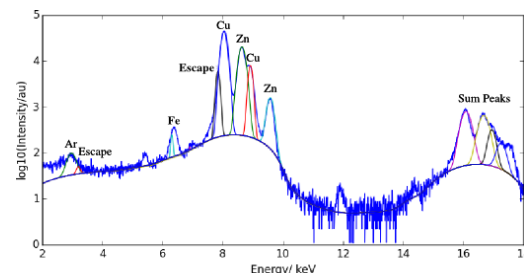
- 1.5 hours single threaded.
- 90s using Savu.

XRD peak fitting:



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

XRF Peak fitting:

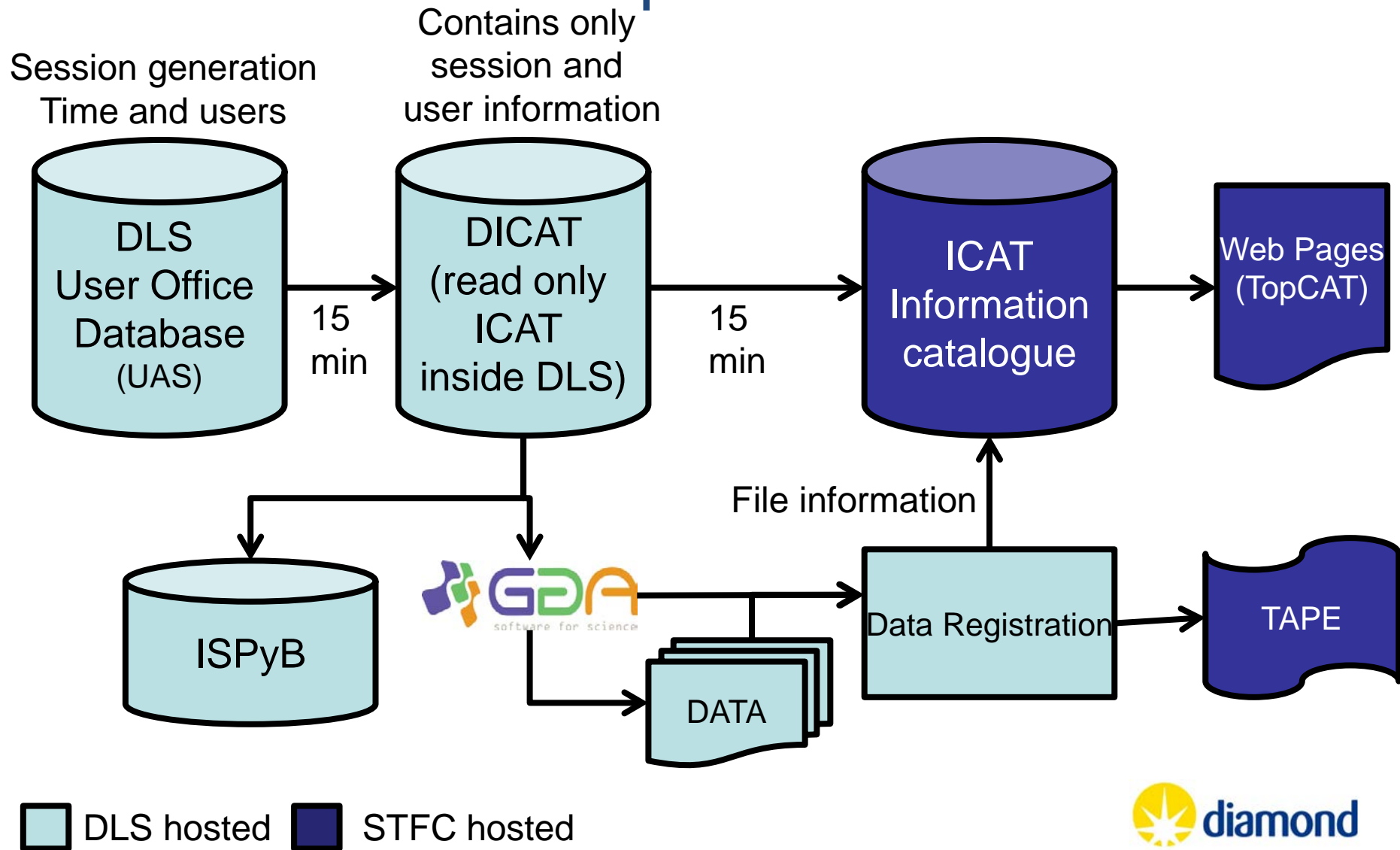


- 2 hours single threaded.
- 5 minutes using Savu.

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.

Top / DIAMOND / CM12150 / I03 2015 Commissioning / Dataset

Name	Size	Create Time	Modified Time
processing	⚙	2015-06-02 16:26:08 +0100	2015-06-02 16:26:08 +0100
.ispyb	⚙	2015-06-02 16:26:09 +0100	2015-06-02 16:26:09 +0100
2015-10-06/fake105537	⚙	2015-10-06 11:05:00 +0100	2015-10-06 11:05:00 +0100
2015-10-06/fake114906	⚙	2015-10-06 11:57:59 +0100	2015-10-06 11:57:59 +0100
processed	⚙	2016-02-03 05:00:34 +0000	2016-02-03 05:00:34 +0000
jpegs	⚙	2016-02-03 05:02:30 +0000	2016-02-03 05:02:30 +0000
20150716/setup/gw/2A	⚙	2016-02-03 05:02:34 +0000	2016-02-03 05:02:34 +0000
2015-07-10/fake083041	⚙	2016-02-03 05:02:39 +0000	2016-02-03 05:02:39 +0000
gw/20150710/kdp/19keV/2	⚙	2016-02-03 05:02:45 +0000	2016-02-03 05:02:45 +0000
domjoe/AmyR	⚙	2016-02-03 05:02:46 +0000	2016-02-03 05:02:46 +0000

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.

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