

Overview of metadata and raw data cataloguing at Diamond.

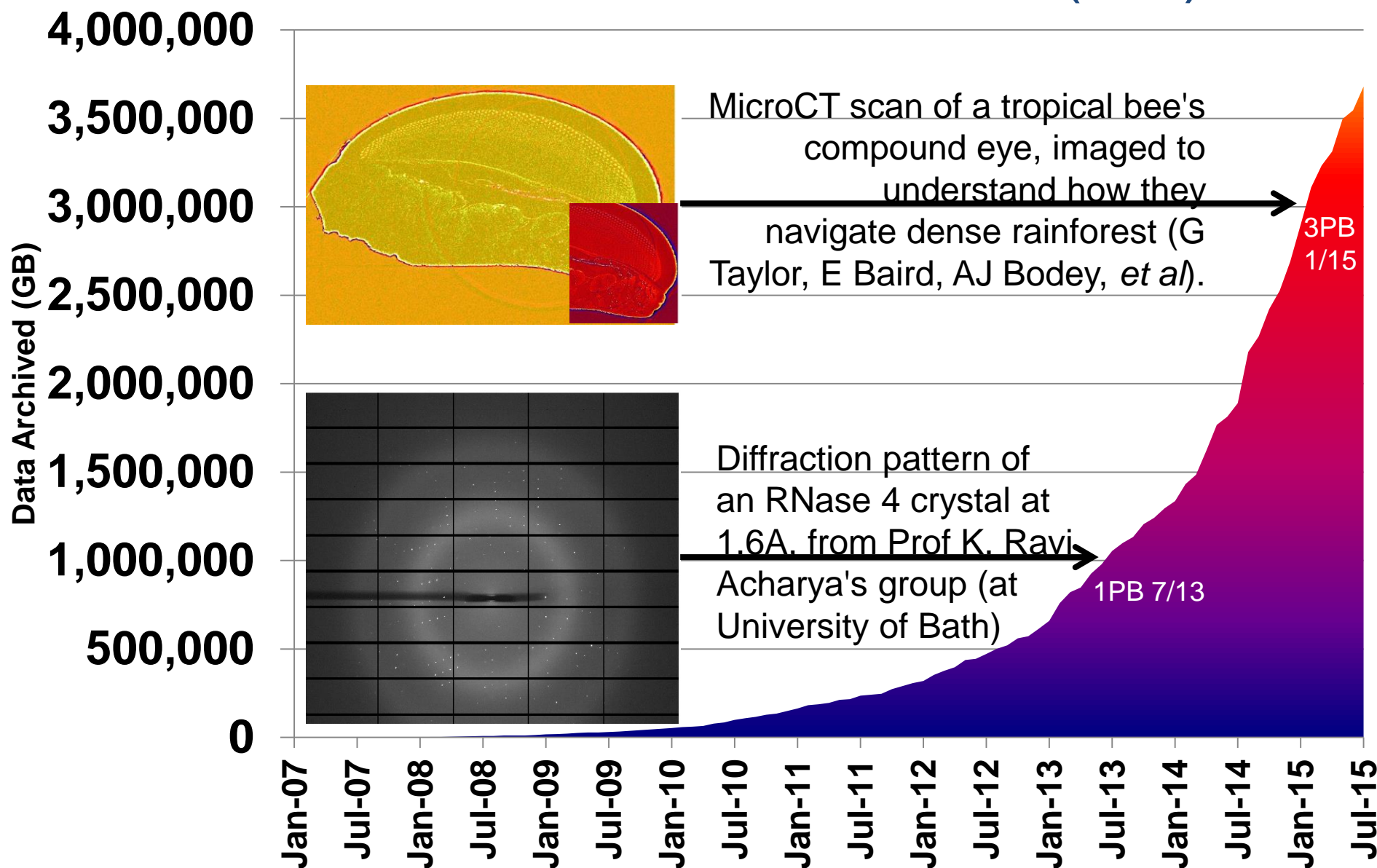
Pierre Aller

Alun Ashton

Further details from :
scientificsoftware@diamond.ac.uk

Diamond Data Archive

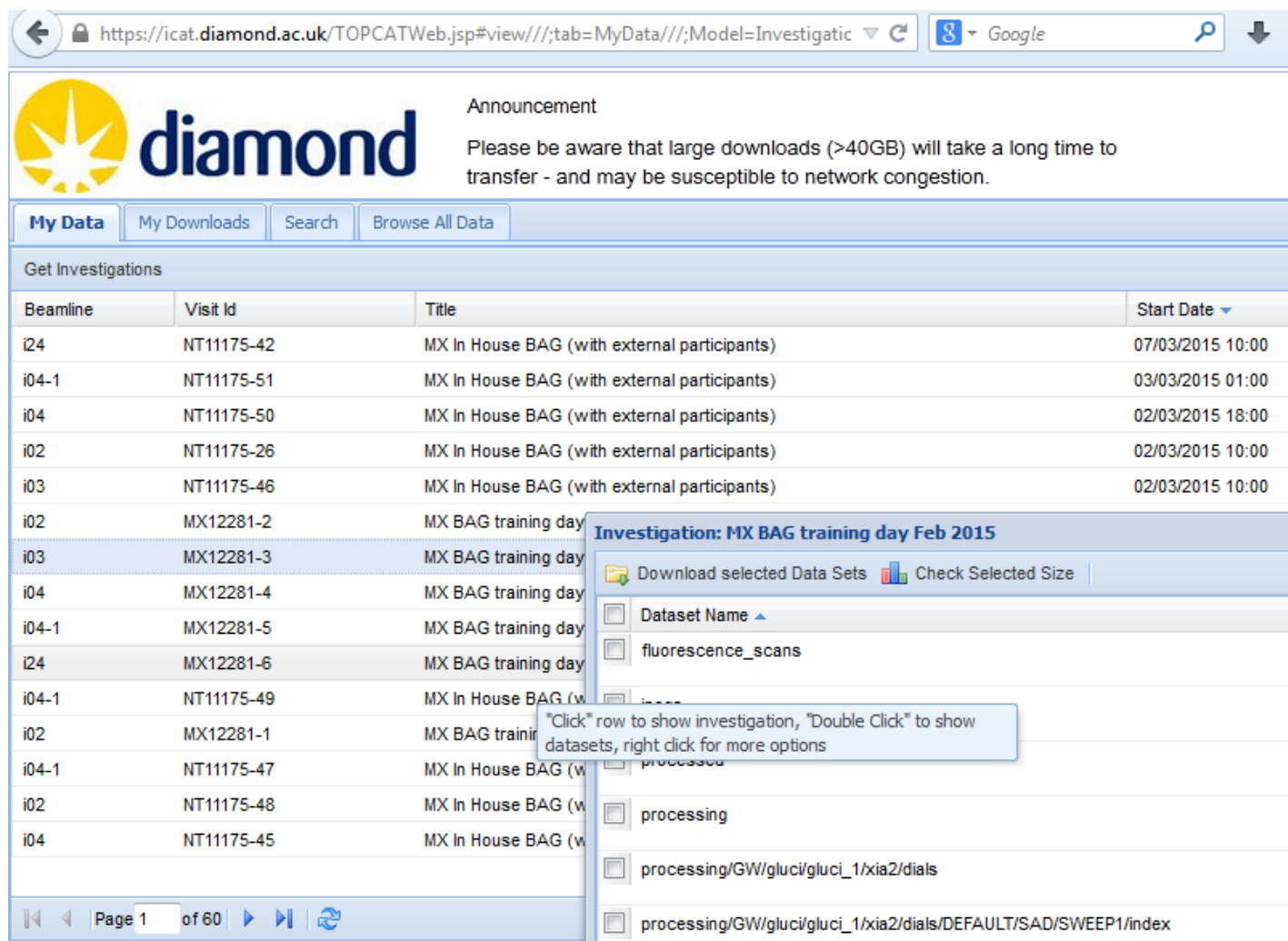
DLS Archived Data Volumes (GB)



Access to Diamond Data Archive

<http://icat.diamond.ac.uk>

Ongoing work to improve reliability, usability, scalability and downloading



The screenshot shows the Diamond Data Archive website. The header includes the Diamond logo and an announcement about large downloads. Below the header is a navigation bar with tabs: My Data, My Downloads, Search, and Browse All Data. The main content area displays a table of investigations. The table has columns for Beamline, Visit Id, Title, and Start Date. The table lists several investigations, including MX In House BAG (with external participants) and MX BAG training day. A sidebar on the right shows the details for the selected investigation, 'Investigation: MX BAG training day Feb 2015'. The sidebar includes a 'Download selected Data Sets' button, a 'Check Selected Size' button, and a list of dataset names with checkboxes. A tooltip is visible over the sidebar, stating: 'Click row to show investigation, Double Click to show datasets, right click for more options'. The footer shows 'Page 1 of 60'.

Announcement

Please be aware that large downloads (>40GB) will take a long time to transfer - and may be susceptible to network congestion.

My Data My Downloads Search Browse All Data

Get Investigations

| Beamline | Visit Id | Title | Start Date |
|----------|------------|--|------------------|
| i24 | NT11175-42 | MX In House BAG (with external participants) | 07/03/2015 10:00 |
| i04-1 | NT11175-51 | MX In House BAG (with external participants) | 03/03/2015 01:00 |
| i04 | NT11175-50 | MX In House BAG (with external participants) | 02/03/2015 18:00 |
| i02 | NT11175-26 | MX In House BAG (with external participants) | 02/03/2015 10:00 |
| i03 | NT11175-46 | MX In House BAG (with external participants) | 02/03/2015 10:00 |
| i02 | MX12281-2 | MX BAG training day | |
| i03 | MX12281-3 | MX BAG training day | |
| i04 | MX12281-4 | MX BAG training day | |
| i04-1 | MX12281-5 | MX BAG training day | |
| i24 | MX12281-6 | MX BAG training day | |
| i04-1 | NT11175-49 | MX In House BAG (w | |
| i02 | MX12281-1 | MX BAG trainin | |
| i04-1 | NT11175-47 | MX In House BAG (w | |
| i02 | NT11175-48 | MX In House BAG (w | |
| i04 | NT11175-45 | MX In House BAG (w | |

Investigation: MX BAG training day Feb 2015

Download selected Data Sets Check Selected Size

Dataset Name

fluorescence_scans

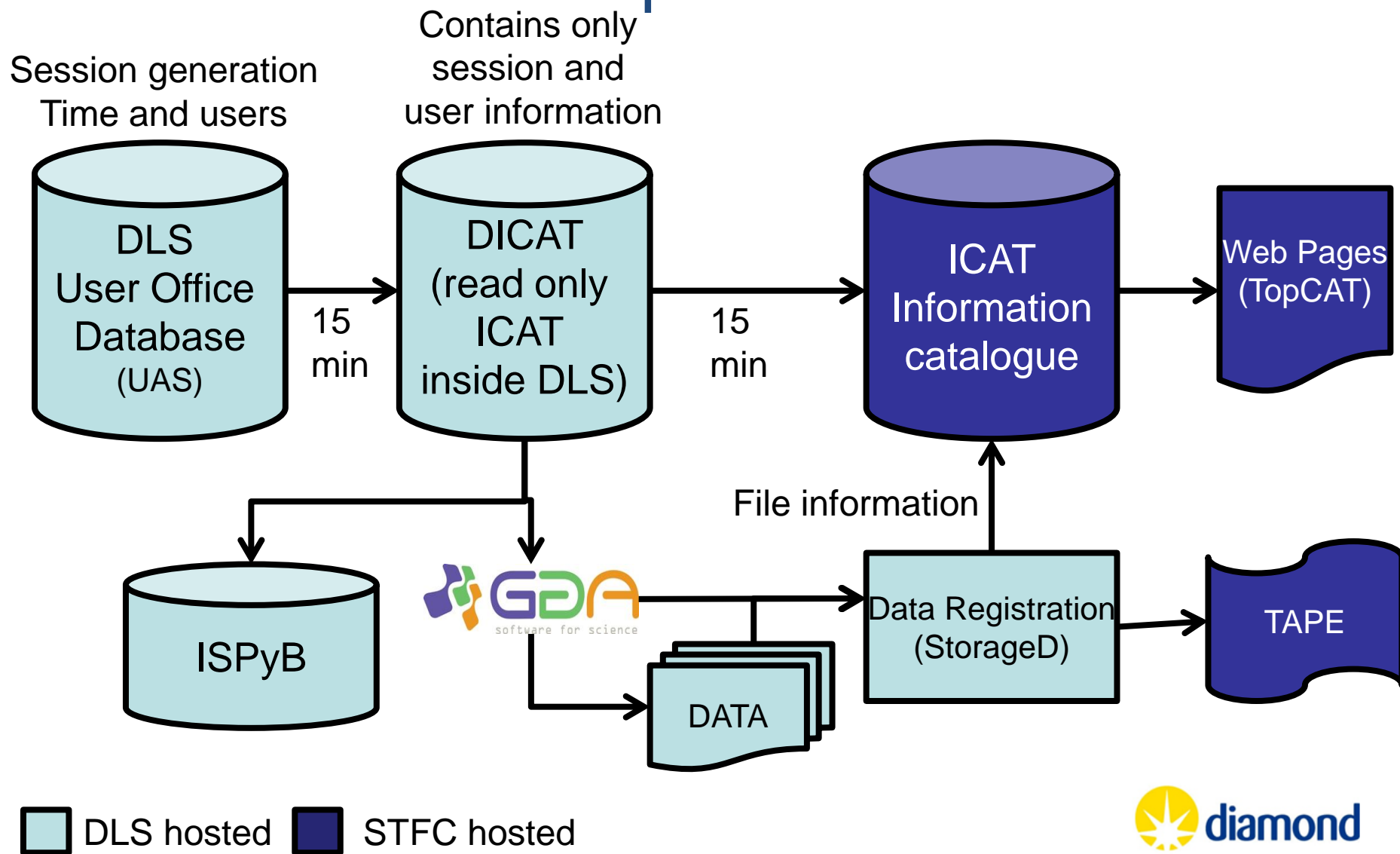
processing

processing/GW/gluci/gluci_1/xia2/dials

processing/GW/gluci/gluci_1/xia2/dials/DEFAULT/SAD/SWEEP1/index

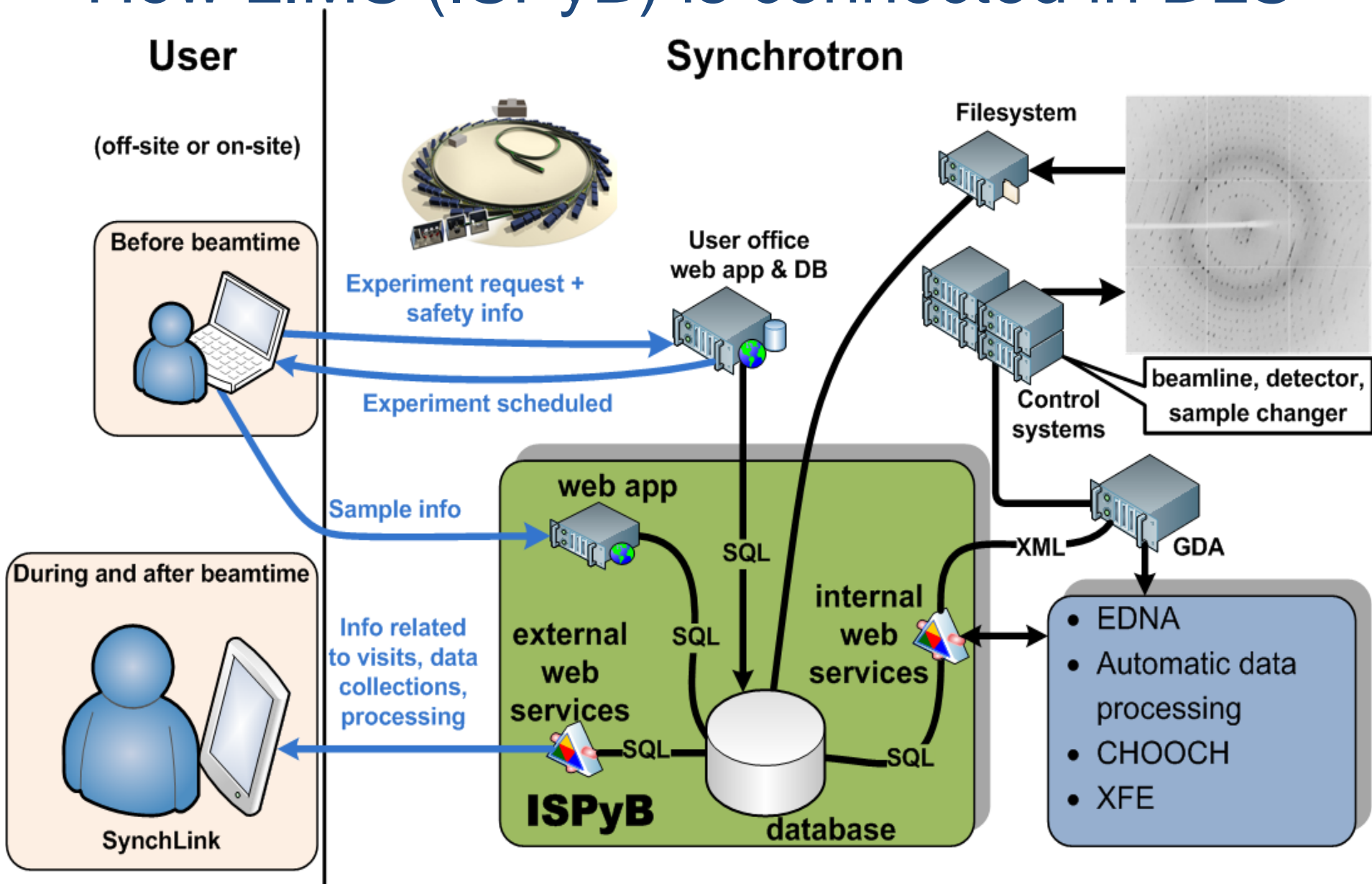
Page 1 of 60

Data is captured from every stage of the process



LIMS – Experiment Information Management.

How LIMS (ISPyB) is connected in DLS



Interfaces

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

Sample: Thaum 13 Ω Start: 78.0° 250 1.2

Ω Osc: 0.10°

No. Images: 1400

Wavelength: 0.9795 Å

Transmission: 5.99%

Type: Data Collection

Comment: (300,-230 subWedge:1Aperture

Auto Processing

Fast DP **DIALS** XIA2 3c

Space Group **A**

P 41 21 2 57.91

Shell **Observat**

outerShell 5113

innerShell 5113

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

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 Experime

Crystal

Actor Attributes

Spot Summary

Experimental Details

test_M6S1_1

Images Collected 2,400

Wavelength 2 Å

1 2 3 4

ther_09_remo

ther_09_infl

ther_09_peak

ther_08_remo

Experiment param

Images collected

Wavelength

Omega start

Omega end

Rotation per image

Exposure time

Beamsize X

Beamsize Y

Transmission

ther_09_remo

Experiment det...

Diff

Experiments (i24) Exp

Autoprocessing re

1 - fast_dp results

2 - xia2 results

3 - xia2 results

4 - xia2 results

Comments

(-31164,-176,98)

Associated images

Crystal snapshots

Diffraction images

Experiment details

3 - xia2 results

xia2 -min_images 3 -3da -xparallel -1...

4 - xia2 results

xia2 -min_images 3 -3dail -xparallel -1...

Comments

(-31164,-176,98)

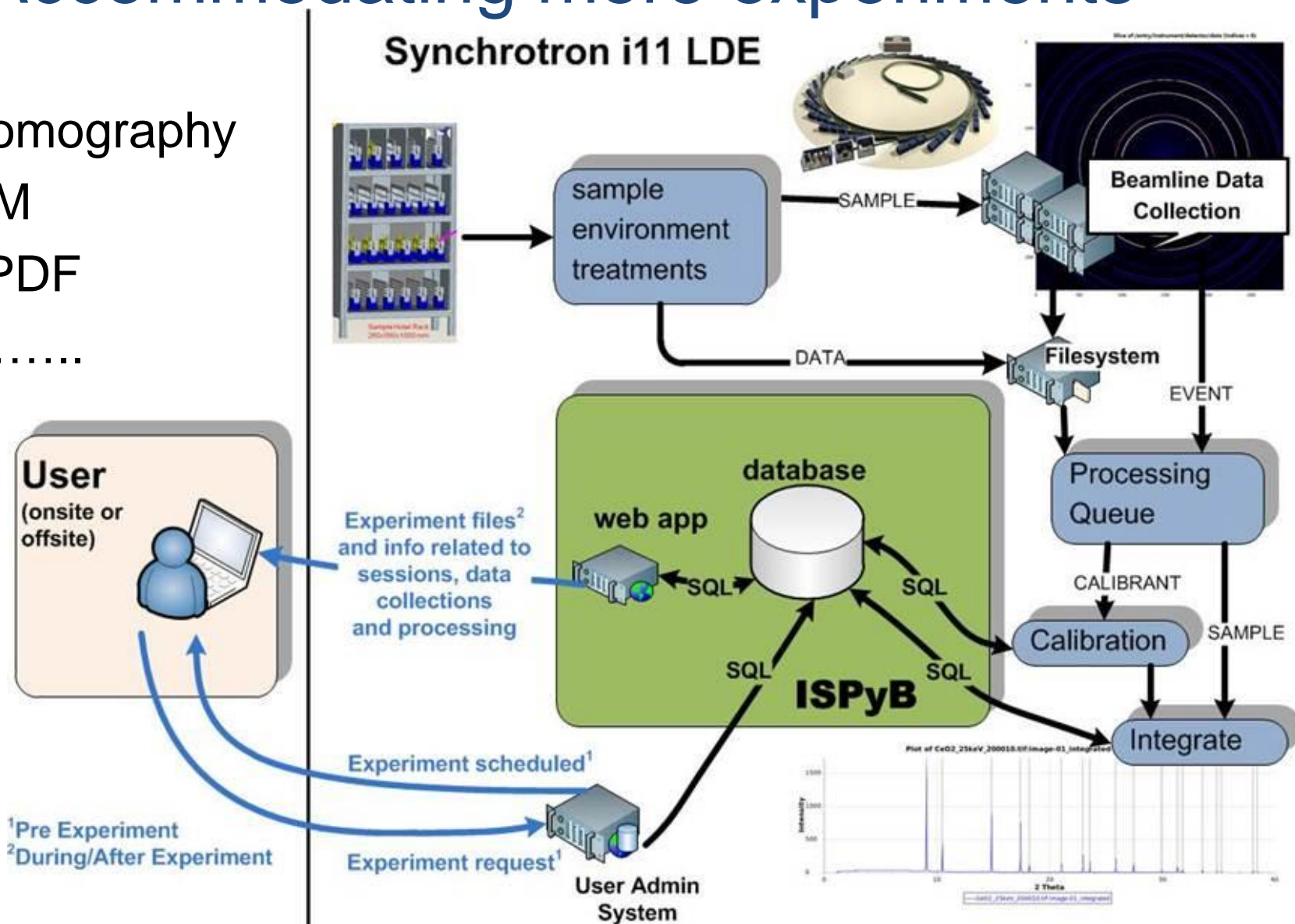
Associated images

Crystal snapshots

Diffraction images

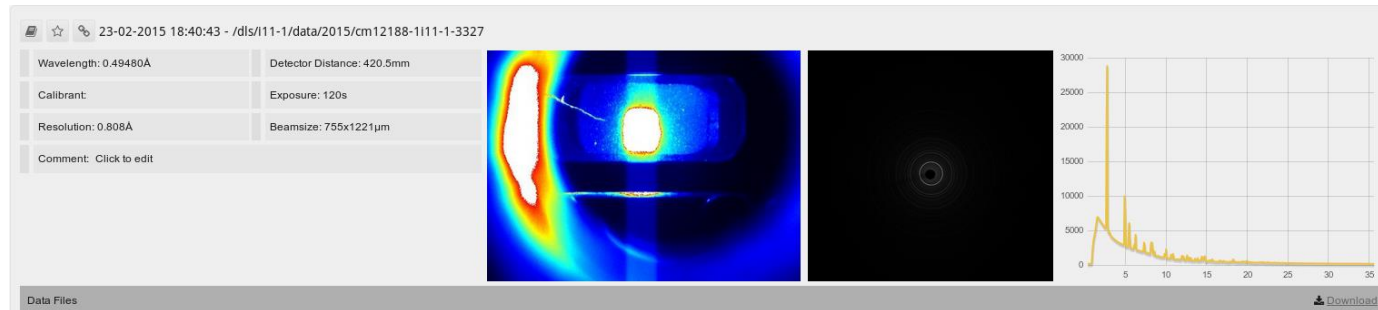
Accommodating more experiments

- Tomography
- EM
- xPDF
-



...With interfaces...

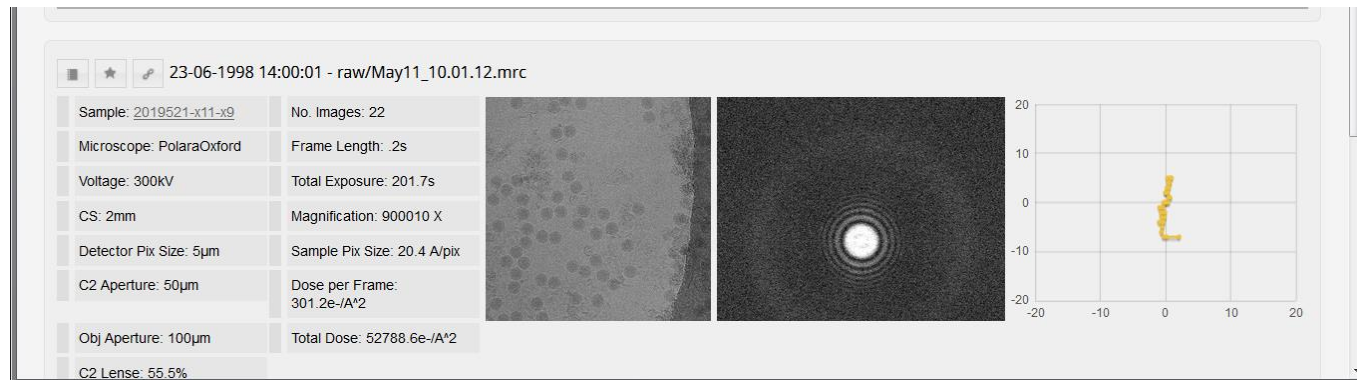
- Powder Diffraction



- Tomography



- EM



The RAW and processed data files

NeXus for RAW data

- Diamond Light Source is still committed to establishing NeXus as its primary file format for its facilities.
 - Key drivers are both scientific and technical.
- A combination of the Data Acquisition Software (GDA) and controls software EPICS available on all photon beamlines facilitates data collection in NeXus

HDF5 improvements for NeXus

- Diamond with other partners are funding improvement in HDF5 for
 - SWMR (Single Writer Multiple Reader)
 - Virtual Dataset, ability to link together multiple data sources/files.
- Diamond have contributed to h5py and Java data bindings to allow use of SWMR.

NeXus for experiments

- Diamond is working with communities to establish standards for all discipline areas e.g.
 - nxMX (investigating bridging imgCIF/CBF)
 - Andreas Förster: EIGER HDF5 data and NeXus format
 - Herbert J. Bernstein: Common diffraction image metadata specification in imgCIF, HDF5 and NeXus
 - Using nxTomo, nxARPES including data processing
 - Used for Non Crystalline Diffraction and soon for Pair Distribution Function experiments
 - nxSTXM, nxFLUO also being investigated

Future integration.

- Continue move towards full NeXus defined RAW and processed (including provenance) data
- Record more experiment and experiment types in a LIMS
- Harvest more information from raw data files and from LIMS to suitably enrich long term data archive
- Better links into external databases and repositories such as wwPDB, CCDC and other facility catalogues.

Summary

- The LIMS systems at Diamond record all aspects of the experiment but not the raw data. This will eventually cover all experiments.
- The 'files' Diamond experiments produce should all eventually adhere to a descriptive standard
- The Diamond data archive should provide a link between the data and reasonable experiment parameters/results

Thanks

- NIAC: NeXus International Advisory Committee
- HDF5 Group
- ISPyB Collaborators
- STFC and ICAT Collaborators
- Diamond computing, software and science groups