

IT solutions to address operational challenges during COVID-19

Dr Neil A Smith

Diamond Light Source

18th June 2020

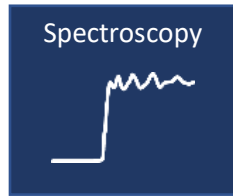
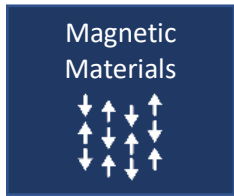
Contents

- Facility Overview
- Remote operation for users
- Computing Infrastructure
- Risks of remote operation
- Future plans

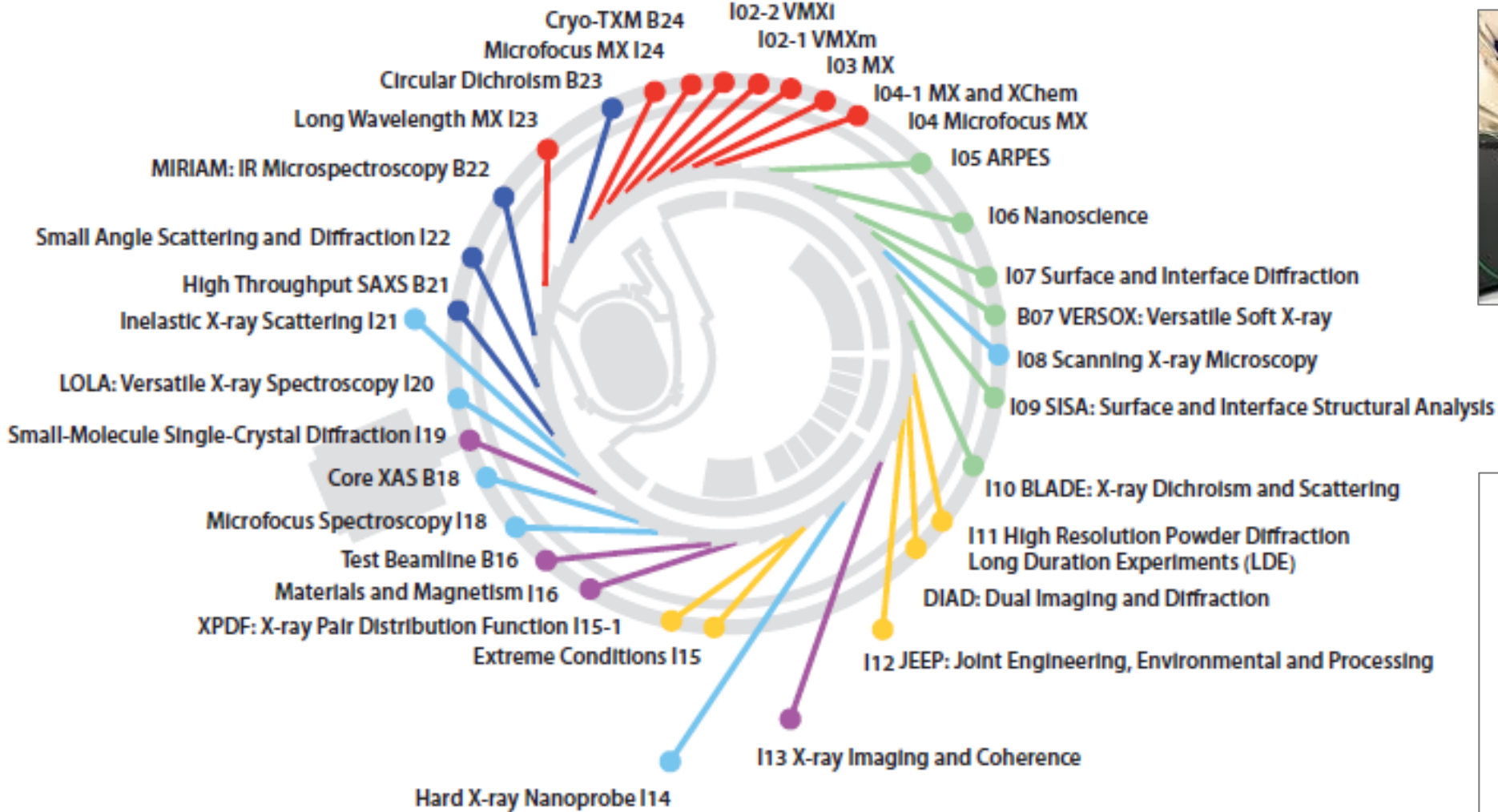


Diamond Light Source, Didcot, UK

Facility Overview



Synchrotron



Electron Bio-Imaging Centre (eBIC)

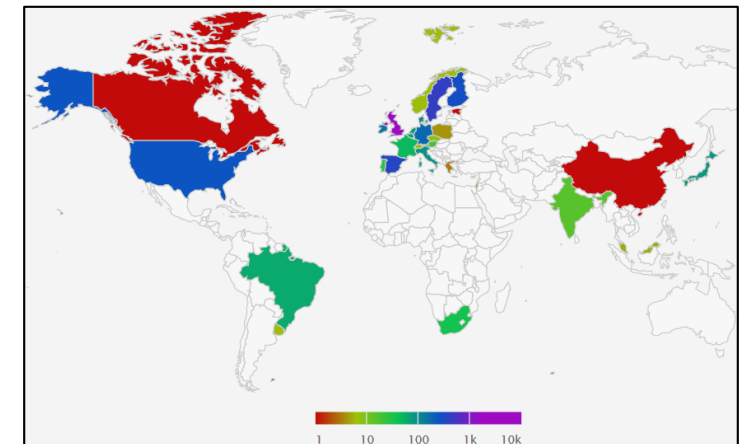
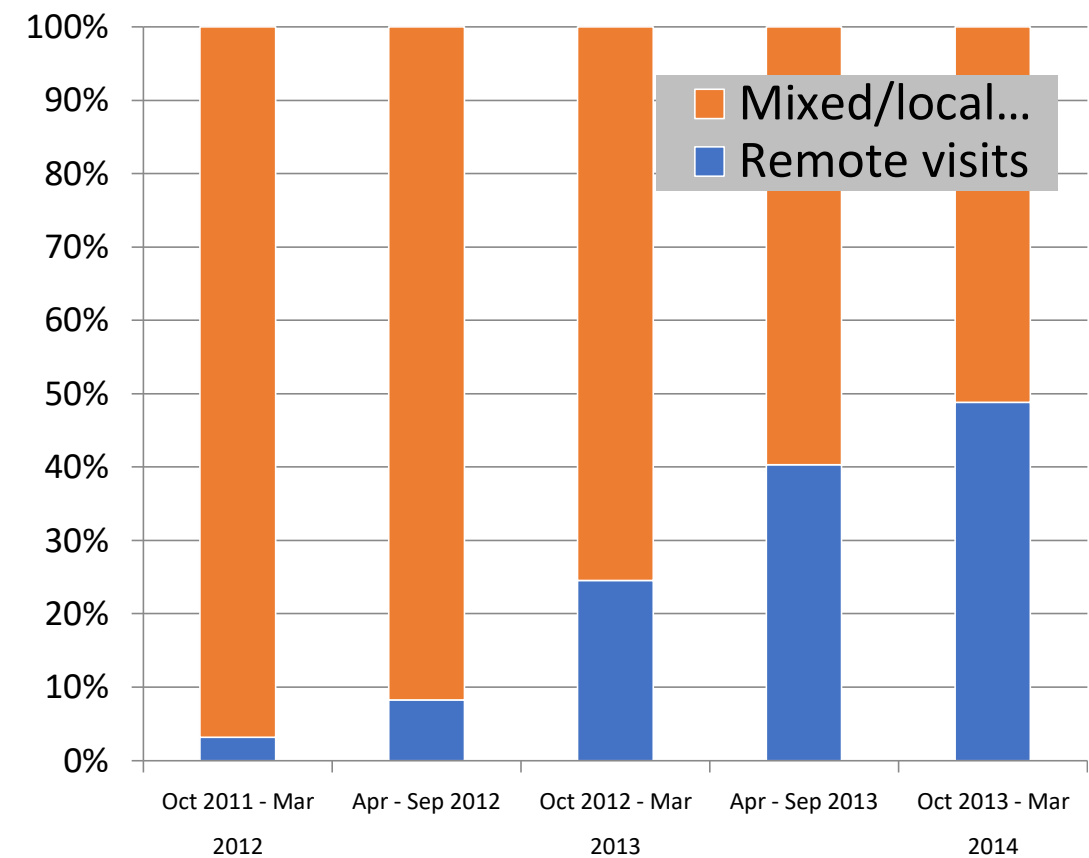


~5000 hours beamtime per year
 ~9000 visits/sessions per year
 30 Operational Beamlines
 6 Operational EMs
 Producing up to ~60TB per day
 Currently 25PB in archive, 6PB last year
 Univa Grid Engine based HPC
 (~6,500 CPU cores, ~240 GPUs)
 GPFS for storage

Remote operation for users

MX user access

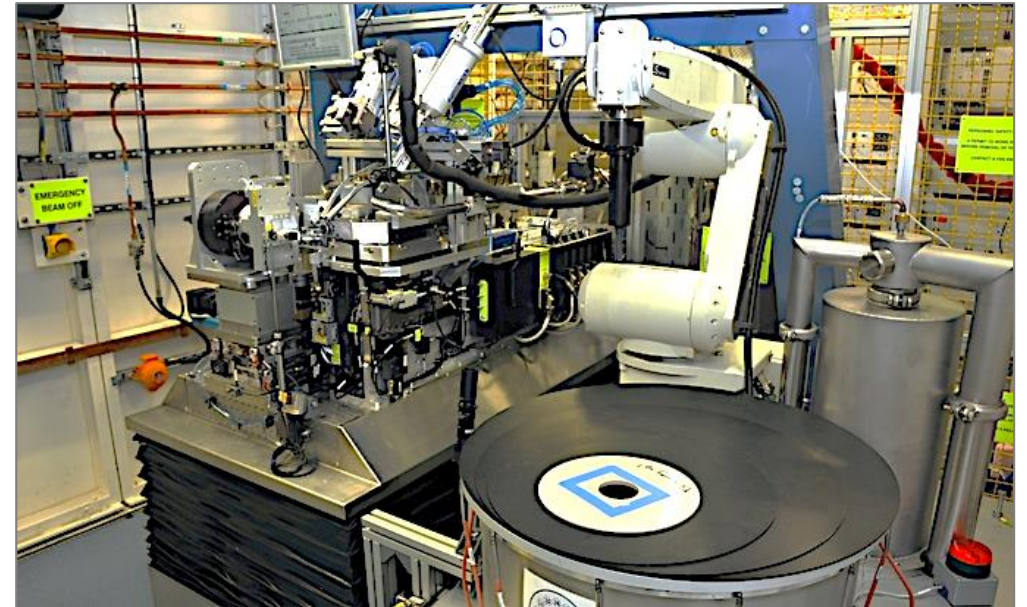
- Trend towards remote operation apparent before COVID
- Significant usage completely remote
 - 2015 ~50% Remote, 20% mixed, 30% on-site
 - 2018 70% visits to I03 and I04 fully remote
 - 2019 75-80% sessions on I03, I04, I04-1 and I24 remote or automated
- Unattended Data Collection (UDC)
- Some of heaviest remote users as close as Oxford and London



Dewar stats per country

MX access support

- Beamline staff load sample changer with pucks, perform beamline checks
- During a remote session users connect via NX (NoMachine)
- Local contact support available via phone
- Recently LCs have been using Zoom to help provide support during lockdown
- Results reviewed by users via Data Acquisition system (via NX) or SynchWeb/ISPyB



Beamline activity

- SynchWeb provides the main web user interface to ISPyB
<https://lspyb.diamond.ac.uk>
 - Designed to support remote viewing of active session – review results etc.
- Shows activities as they are generated/stored in ISPyB
 - Screenings, grid scans, full collections, robot actions
- Shows auto and downstream processing results
- Alerts user with about issues with processing (data collection warnings in future)
- Indications of image quality, radiation damage etc.

The screenshot displays the SynchWeb interface for a specific visit. At the top, there's a 'Data Collections' section with a progress bar. Below it, a navigation bar includes options like 'Assign Containers', 'Summary', 'Auto Processing', 'Visit Stats', 'Users', 'Devers', 'Sample Changer', 'Reprocessing', and 'Beamline Status'. The main content area is titled 'i03 Webcams & Beamline Status' and contains a 'Log' section with a list of data collection events. Below the log is a search bar and a set of tabs for different data types: 'Data Collections', 'Grid Scans', 'Full Collections', 'Auto Integrated', 'Processing Errors', 'Screenings', 'Edge Scans', 'MCA Spectra', 'Robot Actions', 'Sample Actions', and 'Favourites'. The 'Data Collections' tab is active, showing a table of processing results. The table has columns for 'Type', 'Resolution', 'Spacegroup', 'Min I/sig(I)', 'Rmeas Inner', 'Rmeas Outer', 'Completeness', and 'Cell'. Below the table, there are several sections: a 'Fast DIP' status indicator, a '0 Check(s) passed' message, a '4 alert(s)' warning, a 'Beam Centre' table, and a 'Downstream Processing' section with a 'Fast EP' status indicator. The 'Beam Centre' table shows X and Y coordinates for 'Start' and 'Refined' positions. The 'Downstream Processing' section shows a table of 'Shell' data with columns for 'Observations', 'Unique', 'Resolution', 'Rmeas', 'I/sig(I)', 'CC Half', 'Completeness', 'Multiplicity', 'Anom. Completeness', 'Anom. Multiplicity', and 'CC Anom'.

| Type | Resolution | Spacegroup | Min I/sig(I) | Rmeas Inner | Rmeas Outer | Completeness | Cell | Status |
|---------|--------------|------------|--------------|-------------|-------------|--------------|--------------------------------|---------------------------|
| fast_dp | 29.23 - 3.78 | P 4 2 2 | 19.2 | 0.045 | 34.439 | 85.6 | 86.78 86.78 182.87 90.00 90.00 | processing: success: full |

| Shell | Observations | Unique | Resolution | Rmeas | I/sig(I) | CC Half | Completeness | Multiplicity | Anom. Completeness | Anom. Multiplicity | CC Anom |
|------------|--------------|--------|---------------|--------|----------|---------|--------------|--------------|--------------------|--------------------|---------|
| outerShell | 99 | 88 | 3.78 - 3.87 | 34.439 | 0.0 | 0.7 | 10.0 | 1.1 | 1.5 | 1.0 | 0.0 |
| innerShell | 2460 | 99 | 16.89 - 29.23 | 0.045 | 75.1 | 1.0 | 81.8 | 24.7 | 90.0 | 18.6 | 0.3 |
| overall | 190796 | 6814 | 3.78 - 29.23 | 0.060 | 19.2 | 1.0 | 85.6 | 28.0 | 75.4 | 15.2 | -0.0 |

Beamline status and results

The screenshot shows the 'i03 Beamline Status' page. At the top, there are navigation tabs for 'Proposals', 'Projects', 'Unit Cell Search', 'Feedback', and 'Help'. A green notification bar states: 'Please note only valid approved samples can be added to containers. For other, general operating restrictions please see the Diamond website.' Below this, the 'i03 Beamline Status' section contains several green buttons: 'Ring Current 300.221', 'Refill 310.011', 'Hutch Locked', 'Port Shutter Open', 'Expt Shutter Open', 'Fast Shutter ***', 'Wavelength 0.9783', and 'Transmission 100'. The 'Webcams' section features two live video feeds from the 'I03 Sample Changer' dated 'Tue Jun 9 2020'. Below the webcams are sections for 'OAV', 'EPICS Screens', and 'GDA Log'. The 'GDA Log' displays a series of system messages, including 'DEBUG [object:server:32249@i03-control.diamond.ac.uk] ElgerDetectorWrapper - prepare_for_oscillation' and 'INFO [object:server:32249@i03-control.diamond.ac.uk] gda.device.attenuator.EpicsAttenuatorBase - Sending changefilter command'.

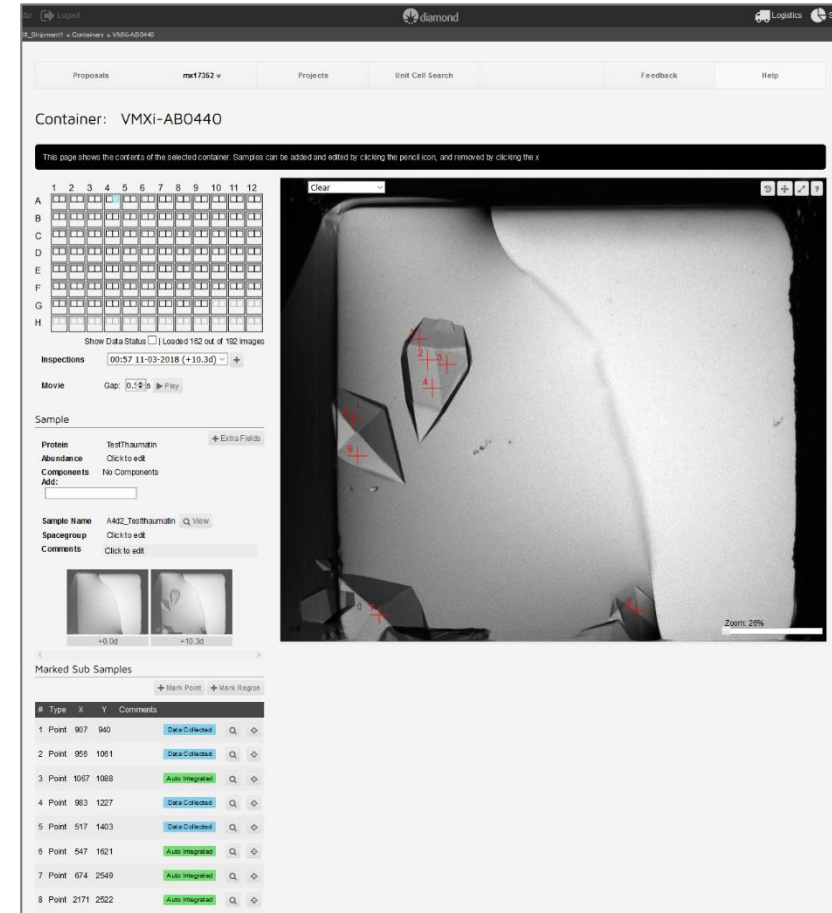
The screenshot shows the 'Image Viewer' interface for a specific experiment. The top navigation is similar to the status page, with 'Proposals' selected. A green notification bar is present. The main header displays the experiment ID: 'cm26459-6: 20200515/TestThaumatm/thaumatm16/thaumatm16_1_master.h5'. Below this, a black bar contains instructions: 'This page is a full scale diffraction image viewer. Mousewheel zooms in and out, drag click to pan around the image. Press > to go to the next image and < the previous.' The viewer shows a diffraction pattern with a central spot and surrounding rings. A resolution of '1.41Å' is indicated. The interface includes control elements for 'Exposure 0.012s', 'Transmission 100.00%', 'Resolution 1.34Å', 'Wavelength 0.9795Å', and 'Oscillation 0.10°'. There are also sliders for 'Zoom 117%', 'Contrast 0', and 'Brightness 0'. The diffraction image is overlaid with a yellow grid and a circular pattern.

Diffraction image viewer: Provides images extracted from nexus files

Webcams only accessible for users during session

VMXi; in-situ beamline

- SynchWeb is the UI for data collection
- Formulatrix crystal imager takes images of plates over a predefined schedule
- User marks locations of crystals within plate locations
- Requests for data collection stored within ISPyB
- Process monitors the queue and instructs Data Acquisition system (GDA) to load the plate, performs the data collection then store the plate back in the imager
- Asynchronous process => capture user intent in ISPyB, use GDA service to perform collection
- Users notified via e-mail that results are available
- **NB: Limited interaction of user with data acquisition system**

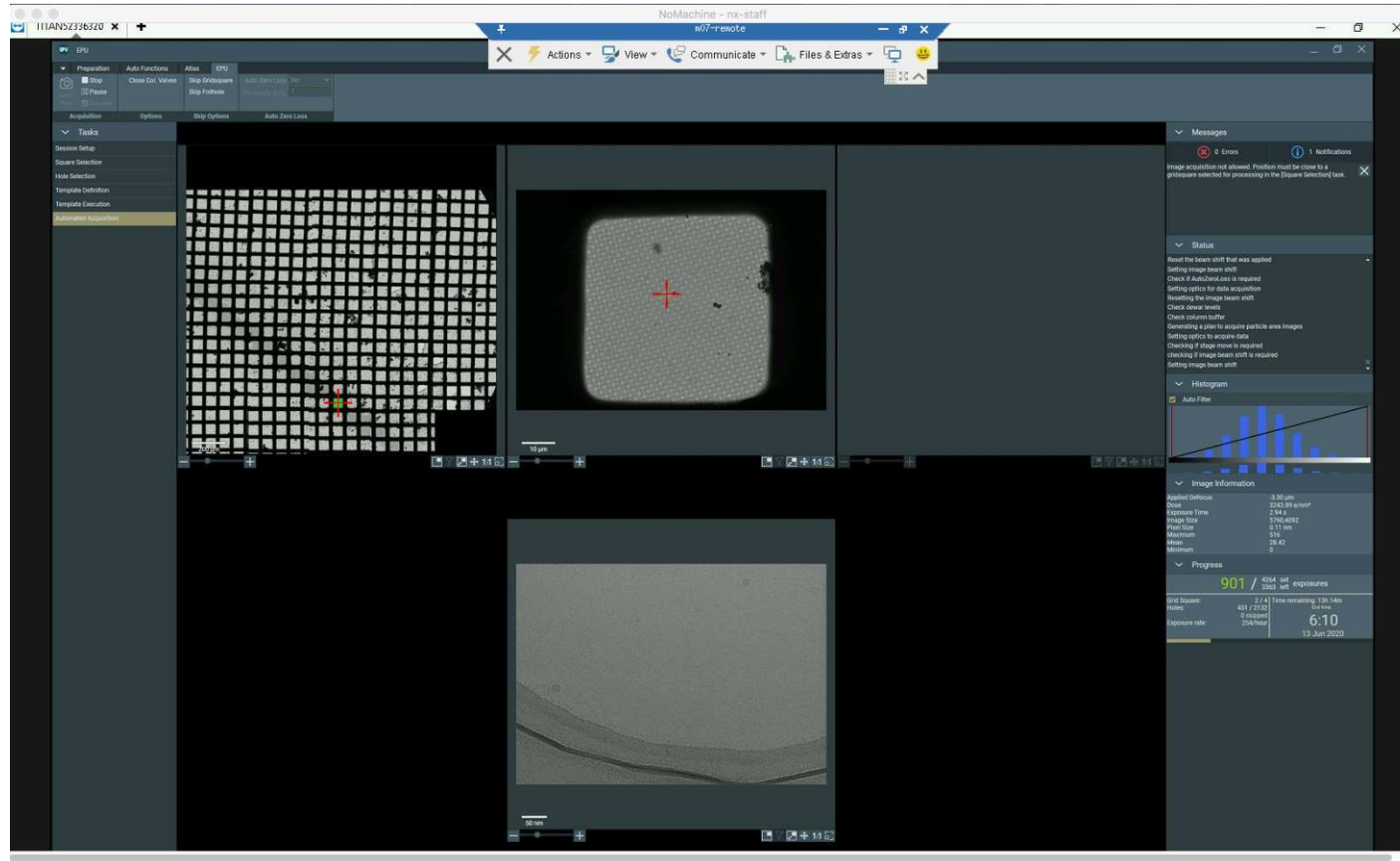


The screenshot displays the SynchWeb interface for the VMXi-AB0440 container. The interface includes a navigation bar with tabs for Proposals, Projects, Unit Cell Search, Feedback, and Help. Below the navigation bar, the container name "VMXi-AB0440" is displayed. A grid of plate locations (A-H, 1-12) is shown, with a large image of the plate showing crystal locations marked with red crosses. The interface also includes a search bar, a "Show Data Status" button, and a "Marked Sub Samples" table.

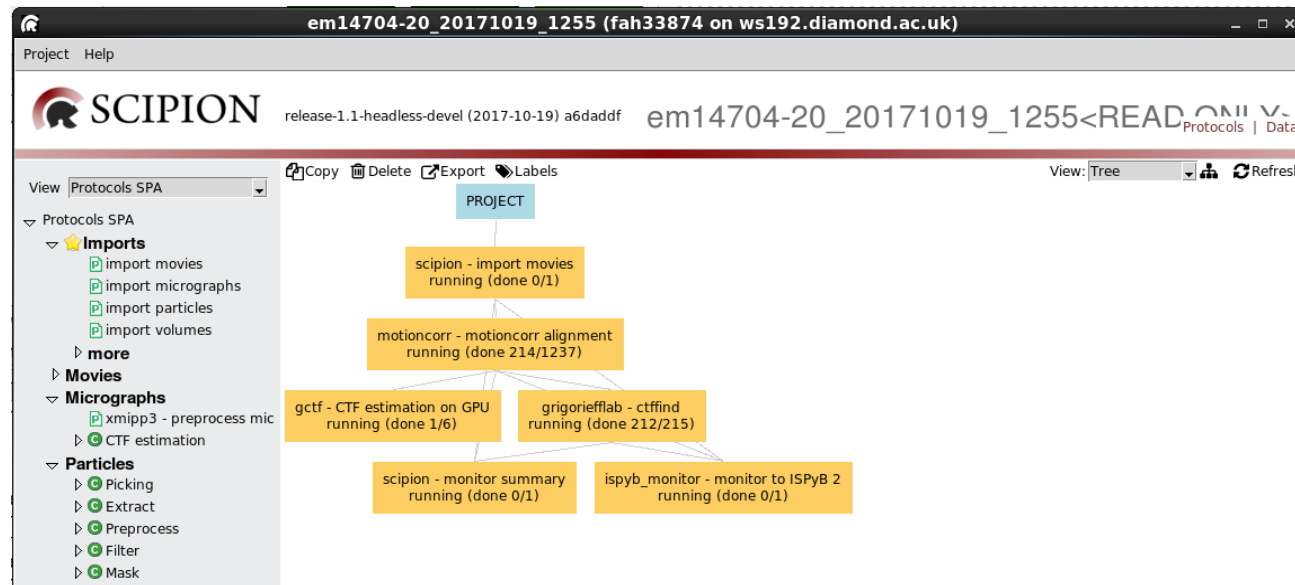
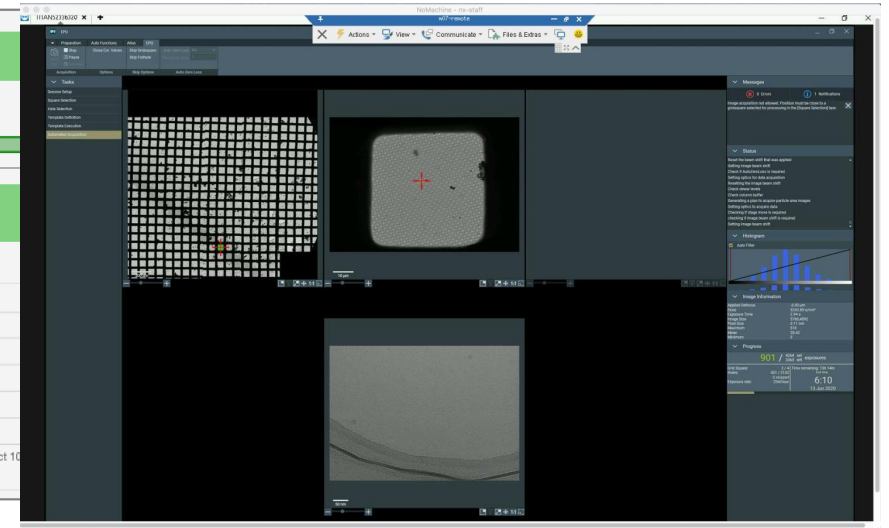
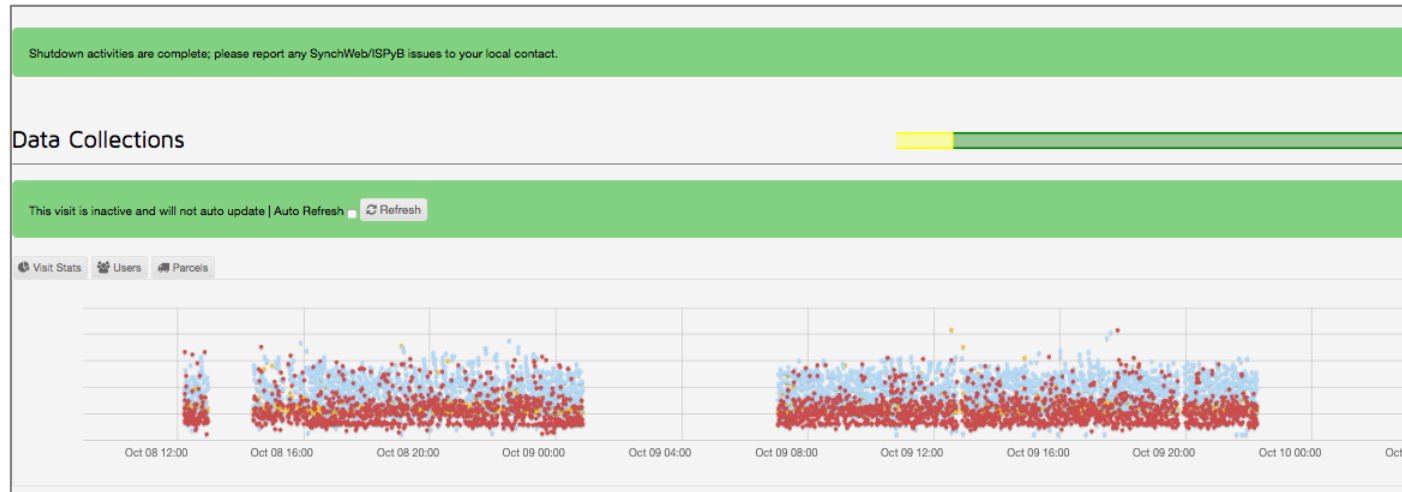
| Type | X | Y | Comments |
|---------|------|------|-----------------|
| 1 Point | 907 | 940 | Data Collected |
| 2 Point | 956 | 1061 | Data Collected |
| 3 Point | 1057 | 1088 | Auto Integrated |
| 4 Point | 983 | 1227 | Data Collected |
| 5 Point | 517 | 1403 | Data Collected |
| 6 Point | 547 | 1521 | Auto Integrated |
| 7 Point | 674 | 2549 | Auto Integrated |
| 8 Point | 2171 | 2522 | Auto Integrated |

eBIC: Remote connection and beamline control for staff

- Access to microscope support PC with FedID and NoMachine (NX)
- Full Access to the Microscope and Detector PC's using TeamViewer (because Windows control PCs)
- Full access to microscope controls - virtual hand panels + microscope alignment



eBIC: Remote connection for user



- ISPyB monitoring of the visit, alternatively the Scipion project can be viewed using NoMachine (User instructions on eBIC webpage)
- Access to microscope control via TeamViewer monitored by staff
- Work ongoing to support Relion processing and expand EM view in SynchWeb

eBIC: ISPyB – MX Style Sample registration

Shipment Contents

Select a dewar by clicking on the row in the table below. Dewar details are then shown below. Click the + icon to add a container to the selected dewar

+ Add Dewar

| Name | Barcode | Facility Code | First Experiment | Tracking # to | Tracking # from | Status | Location | Containers | |
|-------------|----------------------|---------------|------------------|---------------|-----------------|--------|----------|--------------------------|---|
| DLS-01-0001 | cm4950-0003872 | Click to edit | Click to edit | Click to edit | Click to edit | opened | | 0 | + |
| Dewar1 | cm4950-1-i03-0003870 | Click to edit | cm4950-1 | Click to edit | Click to edit | opened | | Click to add a container | + |
| rt | cm4950-0003871 | Click to edit | Click to edit | Click to edit | Click to edit | opened | | 0 | + |
| DLS-01-0001 | cm4950-0003873 | Click to edit | Click to edit | Click to edit | Click to edit | opened | | 0 | + |
| DLS-01-0001 | cm4950-0003874 | DLS-01-0001 | Click to edit | Click to edit | Click to edit | opened | | 0 | + |

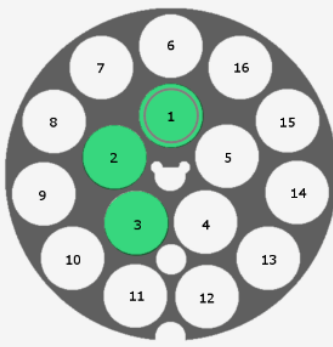
Add Container

This page allows you to add containers to the selected dewar and shipment. If the protein you want to use is not listed type in a new name and press tab. This will create a new protein

Shipment:

Dewar: Dewar1

Puck Name:



| Location | Protein Acronym | Sample Name | Spacegroup | Barcode | Comment |
|----------|----------------------------------|-----------------------------------|-------------------------------|-------------------------------|-------------------------------|
| 1 | <input type="text" value="lys"/> | <input type="text" value="lys1"/> | <input type="text" value=""/> | <input type="text" value=""/> | <input type="text" value=""/> |
| 2 | <input type="text" value="lys"/> | <input type="text" value="lys2"/> | <input type="text" value=""/> | <input type="text" value=""/> | <input type="text" value=""/> |
| 3 | <input type="text" value="lys"/> | <input type="text" value="lys3"/> | <input type="text" value=""/> | <input type="text" value=""/> | <input type="text" value=""/> |



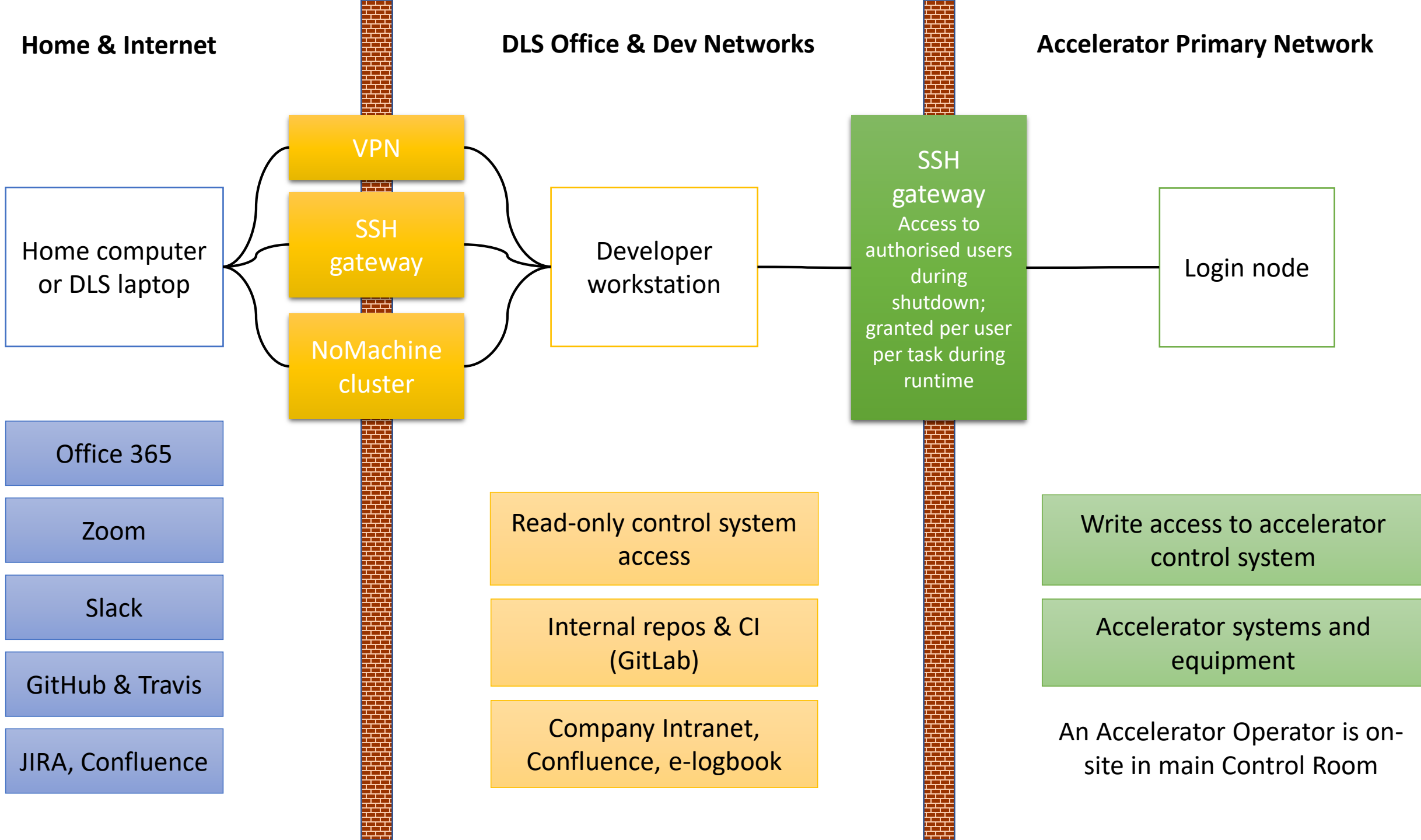
- Diamond labels used to identify dewars in ISPyB – Unique barcode
- eBIC getting transport canes and pucks which will be distributed to our BAG's
- Need to modify SynchWeb/ISPyB MX sample pages to facilitate this style of puck

Current Infrastructure

Home & Internet

DLS Office & Dev Networks

Accelerator Primary Network



VPN

SSH
gateway

NoMachine
cluster

Home computer
or DLS laptop

Office 365

Zoom

Slack

GitHub & Travis

JIRA, Confluence

Developer
workstation

Read-only control system
access

Internal repos & CI
(GitLab)

Company Intranet,
Confluence, e-logbook

SSH

gateway
Access to
authorised users
during
shutdown;
granted per user
per task during
runtime

Login node

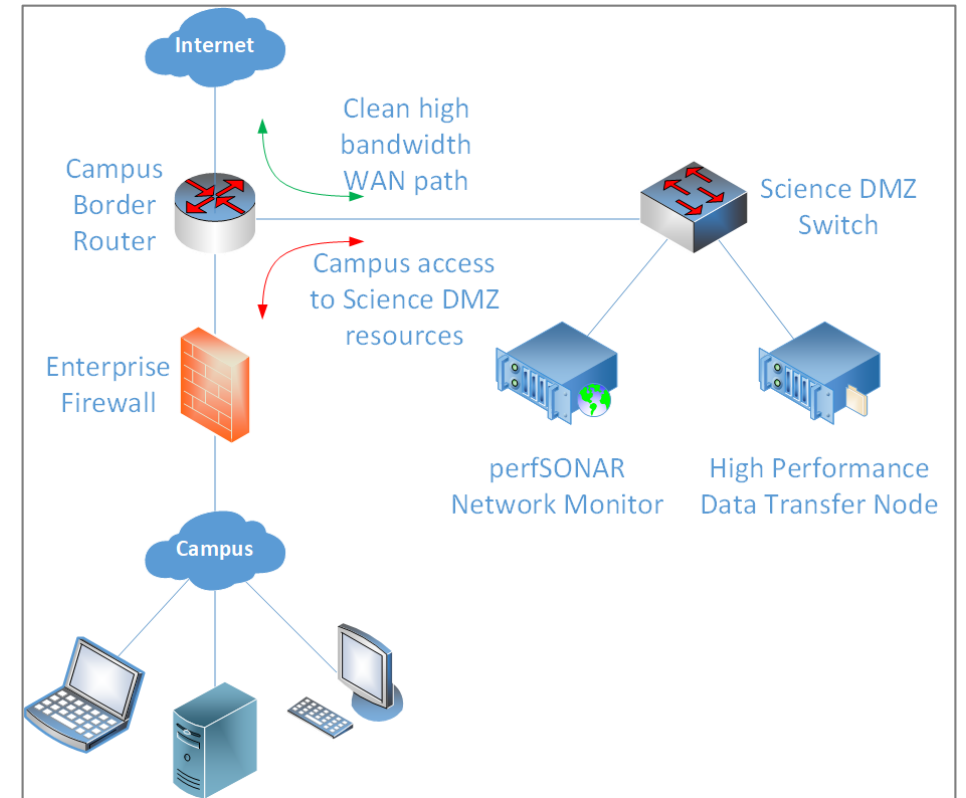
Write access to accelerator
control system

Accelerator systems and
equipment

An Accelerator Operator is on-
site in main Control Room

Data Retrieval

- Data kept on disk for at least 40 days
- Within the 40 day window, users can download data from DLS workstation or via Globus
 - STFP for data sets < 20Gb
 - Globus for data sets > 20Gb
- After 40 days data is archived via iCAT
 - Data sets can be downloaded via TopCAT
 - Data can be restaged to DLS filesystem, or in future other providers e.g. STFC HPC such as Scientific Computing Application Resource for Facilities (SCARF)
- Note: The data that the user can access on the filesystem underneath Globus, or the data that is presented to them in the Topcat interface, is only that that the user is entitled to see.



“Science DMZ” network architecture

Risks of remote operation

Diamond Remote Working / Access Risks

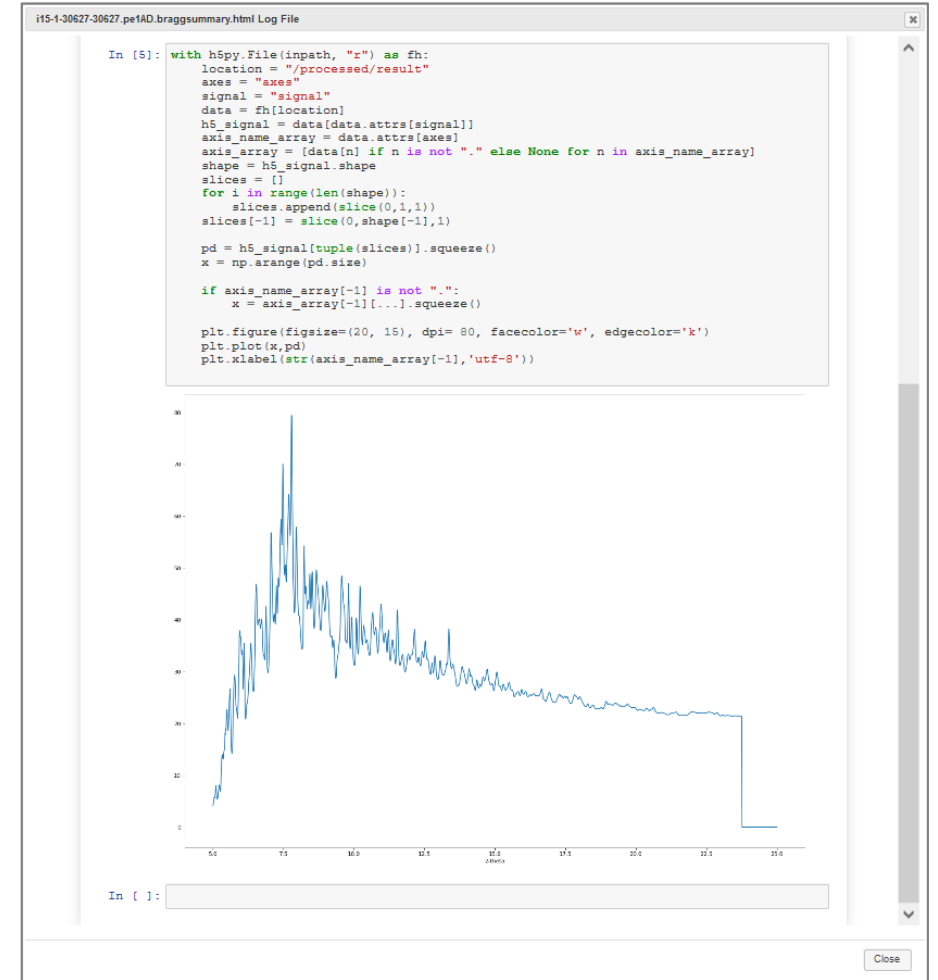
| | Risk | Mitigation |
|------|---|--|
| i | New operational information security risks resulting from home working (in particular use of computers that do not have latest software versions and anti-virus software) | <ul style="list-style-type: none"> Publish '10 Cyber Security Tips' - on Diamond COVID-19 site Offer dedicated support contact; with quick response guaranteed Link to NCSC advice on home working Cyber Security talk given in 'Learning at Home/Work Week' series Expedited delivery of more secure replacement Virtual Private Network |
| ii | New Covid-19 related threats, in particular related to Diamond's R&D activities; notably phishing emails | <ul style="list-style-type: none"> Collect intelligence from: security experts, various fora, and in particular: National Cyber Security Centre, Cyber Security Information Sharing Partnership; Jisc, CNR Network Reporting, Oxford University CISO <ul style="list-style-type: none"> Inform staff; created Cyber Alert page on COVID-19 site Mitigate risks; make changes to firewall |
| iii | Use of Video Conferencing platforms insecurely leading to unintentional release of data (video conferences have increased in number almost 100 times since the start of lockdown) | <ul style="list-style-type: none"> Publish '10 Microsoft Teams Tips', 10 Video Conferencing Tips (Zoom), 10 Video Conferencing Etiquette and Effectiveness Tips - on COVID-19 Internal site Appraise Teams configuration settings; adjust to improve security Develop full Teams User-Guide - on COVID-19 site Change Zoom defaults (including requiring passwords for all video conferences) |
| iv | IG and IT risks in Diamond risk register require to take into account COVID-19 | <ul style="list-style-type: none"> Re-appraised security risks, updated risk scores as required; reported to Information Governance Committee and Risk Coordination Committee – mitigations agreed |
| v | Office 365 not configured sufficiently securely across Diamond | <ul style="list-style-type: none"> Microsoft Secure Score for Diamond improved (including requiring Multi Factor Authentication for administrators) |
| vi | Display Screen Equipment (DSE) not set correctly in home environment | <ul style="list-style-type: none"> Home working guidance provided - on COVID-19 site Staff able to borrow screens and chairs |
| vii | Staff not sufficiently informed about GDPR risks | <ul style="list-style-type: none"> New cyber security awareness training software system purchased from <i>Cybsafe</i>, about to be rolled out |
| viii | Threat to Diamond following attack to European academic community supercomputers (BBC report) | <ul style="list-style-type: none"> Incident Response Team mobilized, threat communicated to users, steps taken to protect Diamond High Performance Computing Services from related attack |
| vix | Authentication security not sufficient for remote users | <ul style="list-style-type: none"> Implement project to enable Multi Factor Authentication for remote access for all systems |



Future plans

Increased emphasis on remote operation and mail-in for samples

- Project in early stages to allow shipment/mail-in through for Soft Condensed Matter (SCM) group
- Expand use of mail-in for Crystallography group
 - i19 beamlines already using MX capability,
 - i15-1 XPDF work on going – new robot plus JupyterHub for analysis
- Main challenge is extending ISPyB schema to support non-MX sample descriptions



Jupyter Notebook stored as processing attachment in ISPyB

Cloud Roadmap

Cloud (off – premise) computing is seen as a key requirement / enabler for faster, broader, better post-processing of experimental data at Diamond

- Completed pilot project to investigate AWS and Azure for cloud bursting.
- Recruited dedicated cloud engineer in Scientific Computing

On premise computing infrastructure critical for data acquisition and near real time processing of current sessions

- Provisioned private hub for docker images in Diamond to ensure security on local clusters. More recent work on adopting Singularity as container of choice.
- Newest ‘Hamilton’ cluster now supports Docker and Singularity images. Acting as a testbed and starting point for supporting users with applications that run in a container.
- Provisioned on-premise JupyterHub and Kubernetes cluster
- Training staff to use podman to develop and deploy images

Summary

- In general existing remote capability has worked well for instruments that are still running
 - VPN upgrades in progress
 - Already well provisioned with NX capability but increased memory usage per server to cope with increased demand
- Impact on data analysis software engineers manageable via ssh/NX to workstations or via VPN
- More impact on controls software engineers due to limited access to lab based hardware
 - Also callouts / trouble shooting more difficult off-site
- Investigating other remote desktop technologies and VDI solutions
- Continue to investigate cloud bursting/re-hosting data on cloud services for off-site data analysis