Cryo-EM DAQ & Data Management

Yee-Ting Li, January 2019

Cryo-EM Training Workshop, SLAC

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Agenda

1. Remote Access: **FastX**
2. Experimental Metadata: **eLogbook**
3. Data Acquisition: **Pre-processing Pipeline**
4. Where is my data? **Data Transfer**

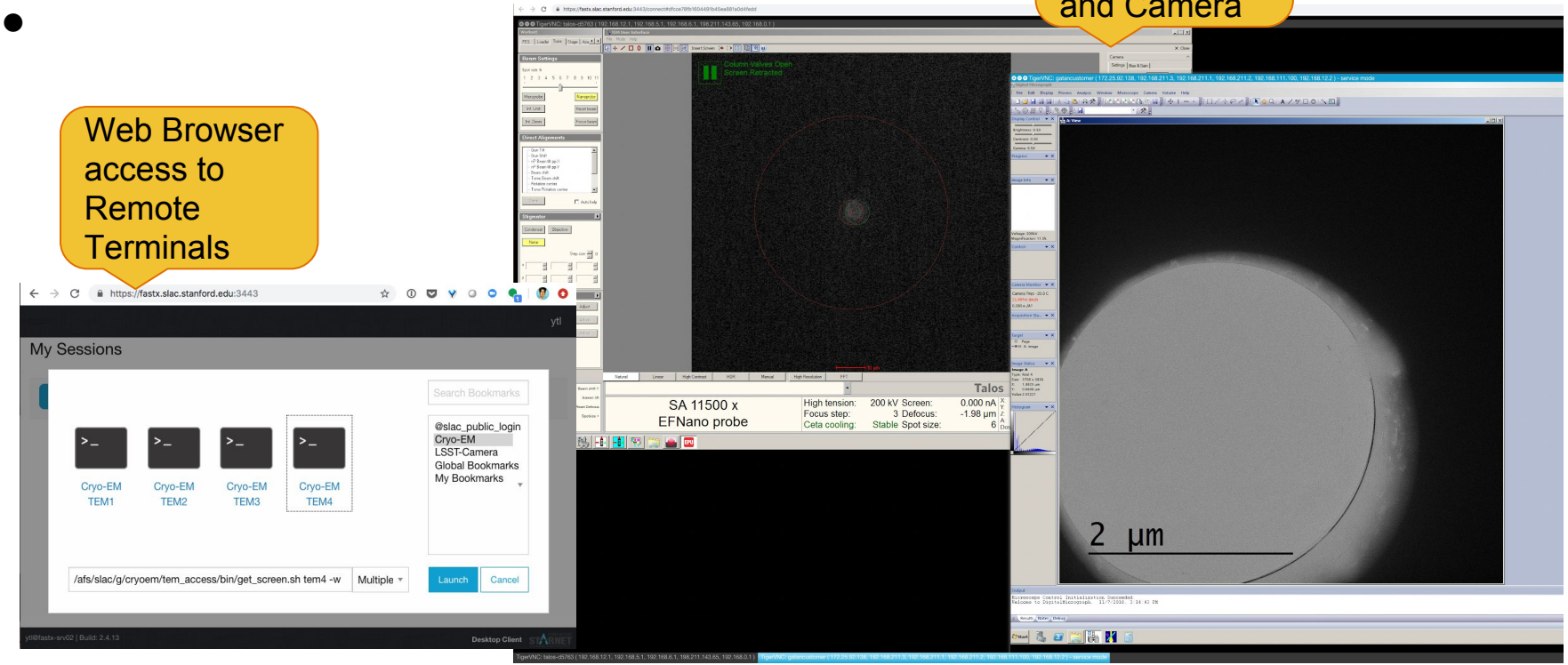
Question:

How do we provide the best experience for users of Stanford-SLAC's Cryo-EM Facility?

Remote Access

Remote control of Microscope and Camera

Web Browser access to Remote Terminals

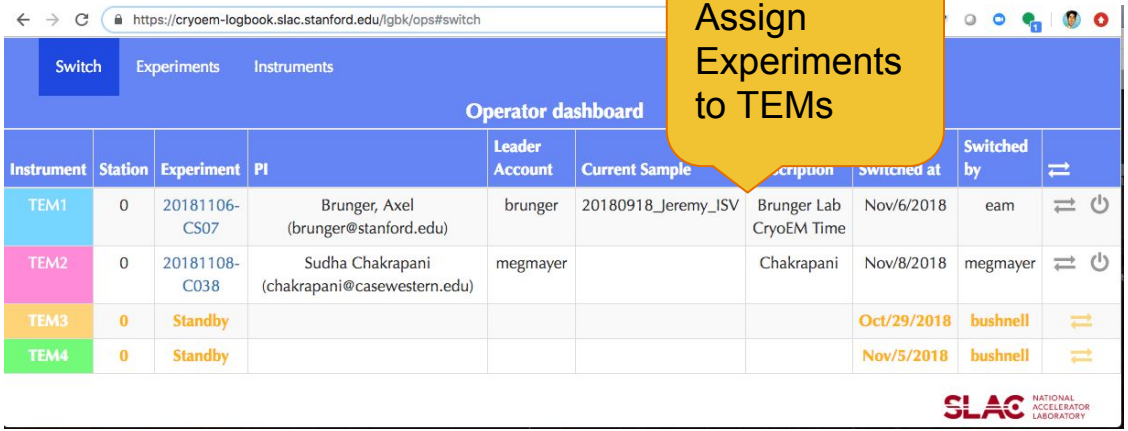


Enable Secure Remote Control of TEMs

Microscope Management and Experiment Tracking

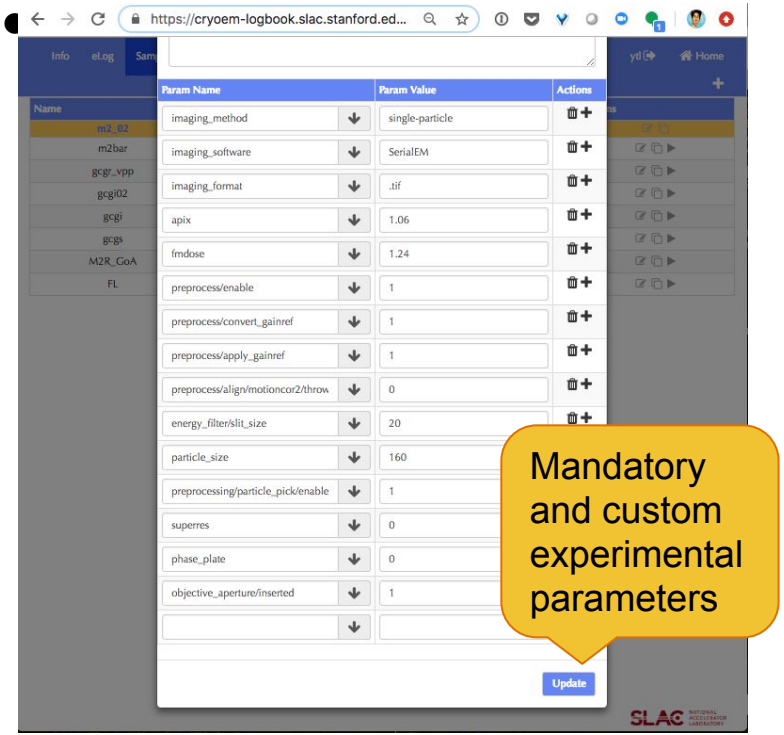
Problem: How to track usage and 500+ experiments for 4 microscopes each year

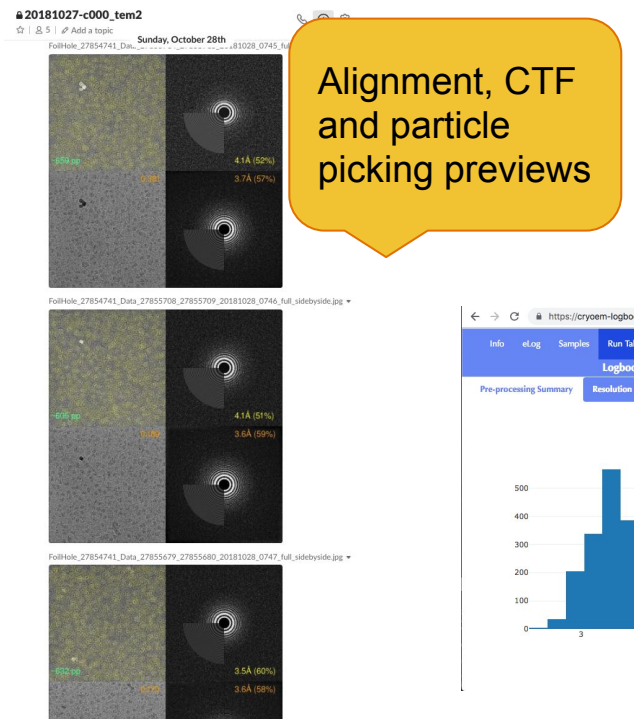
Solution: “elogbook” Web based logging and tracking tool



Make Data Management Easy and Consistent

Live Experimental Feedback for Users





Provide Data Logbook; Markup with Metadata; Provide Near Real-Time Feedback

0. Prerequisites

- Access to everything described here requires:
  - SLAC Unix Account
    - Request for yourself and all collaborators
    - Contact: **lisa@slac.stanford.edu**
  - An internet connection

Contact ytl@slac.stanford.edu for any compute/storage issues

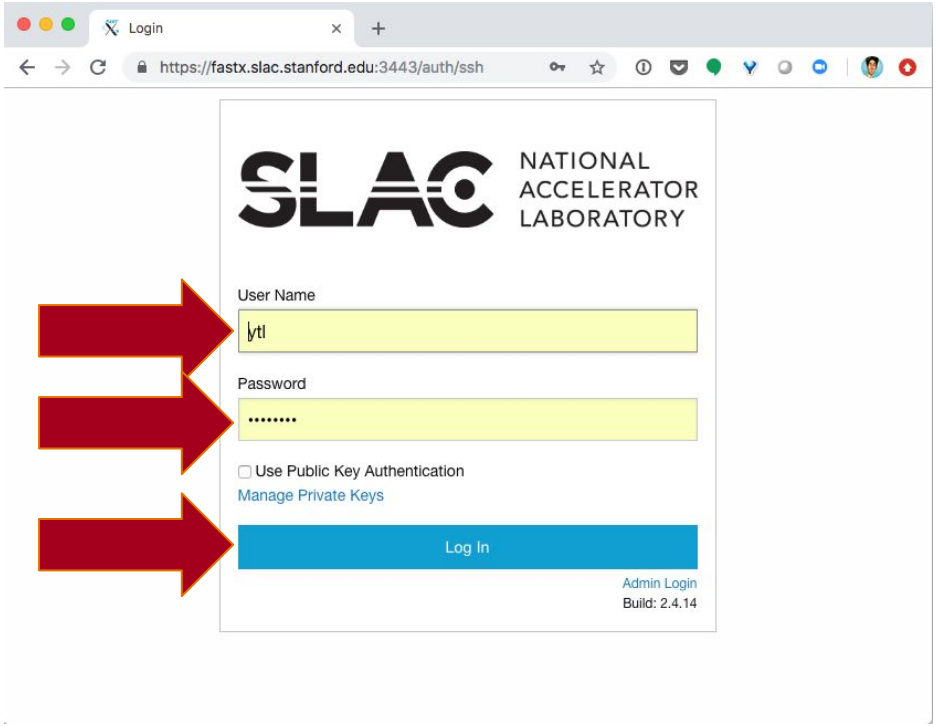
1. Remote Access

- Why?
  - Remotely control microscope to setup/change parameters, monitor live
- How?
  - Log onto website, or
  - Download ‘fastx client’
- Where?
  - https://fastx.slac.stanford.edu:3443

FastX

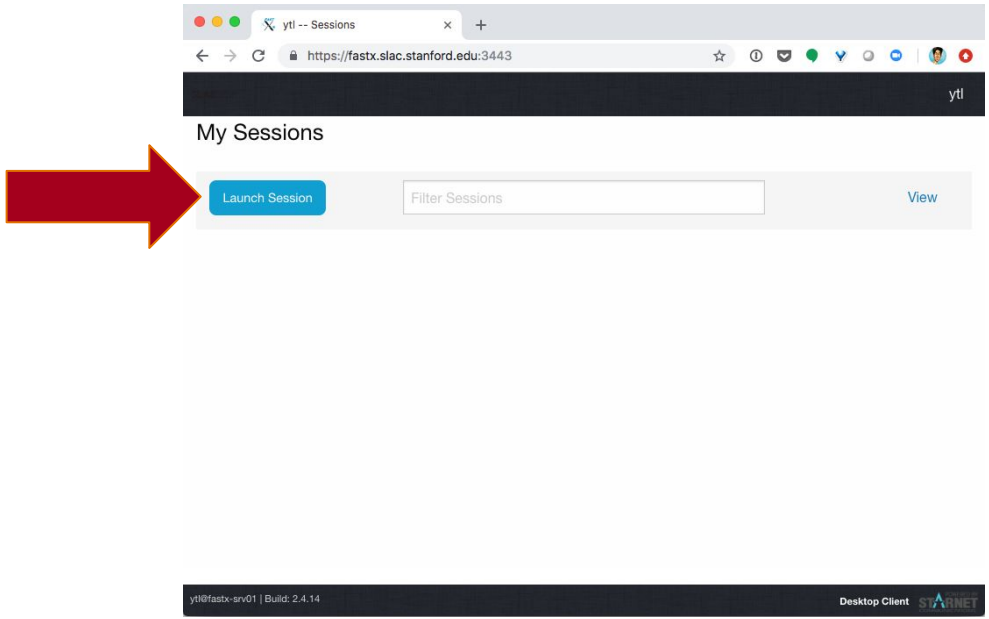
- Remote control software similar to NoMachine, TeamViewer etc.
- No software install required (just a standard browser)
- Client App allows microscope software windows to appear like normal windows on your Desktop

FastX - Logon

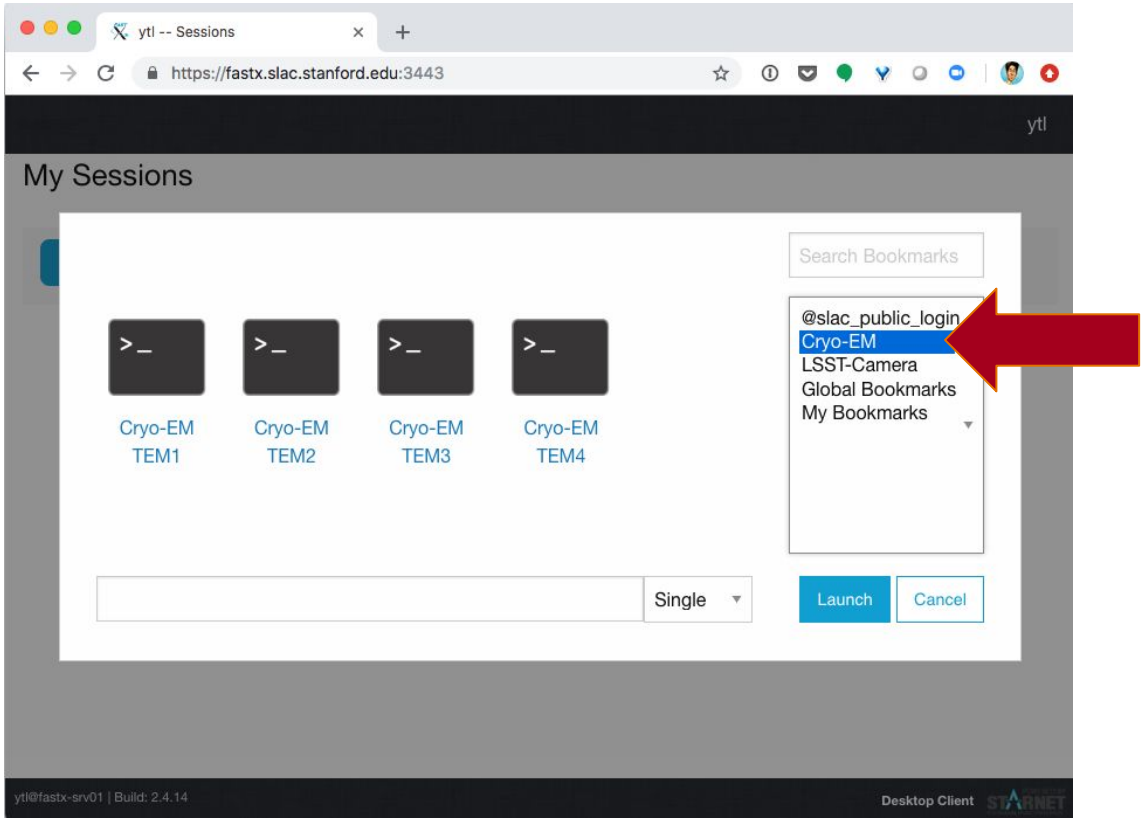


https://fastx.slac.stanford.edu:3443

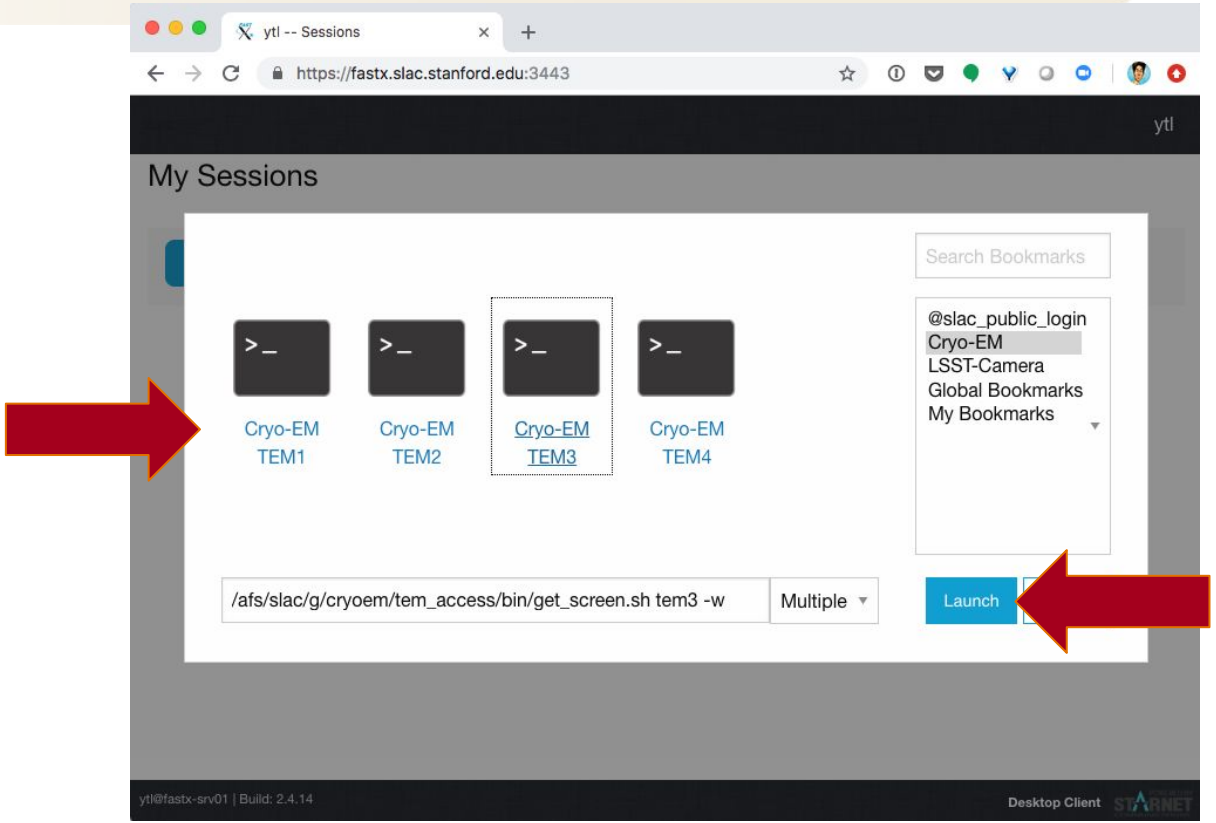
FastX - Click ‘Launch Session’

- 

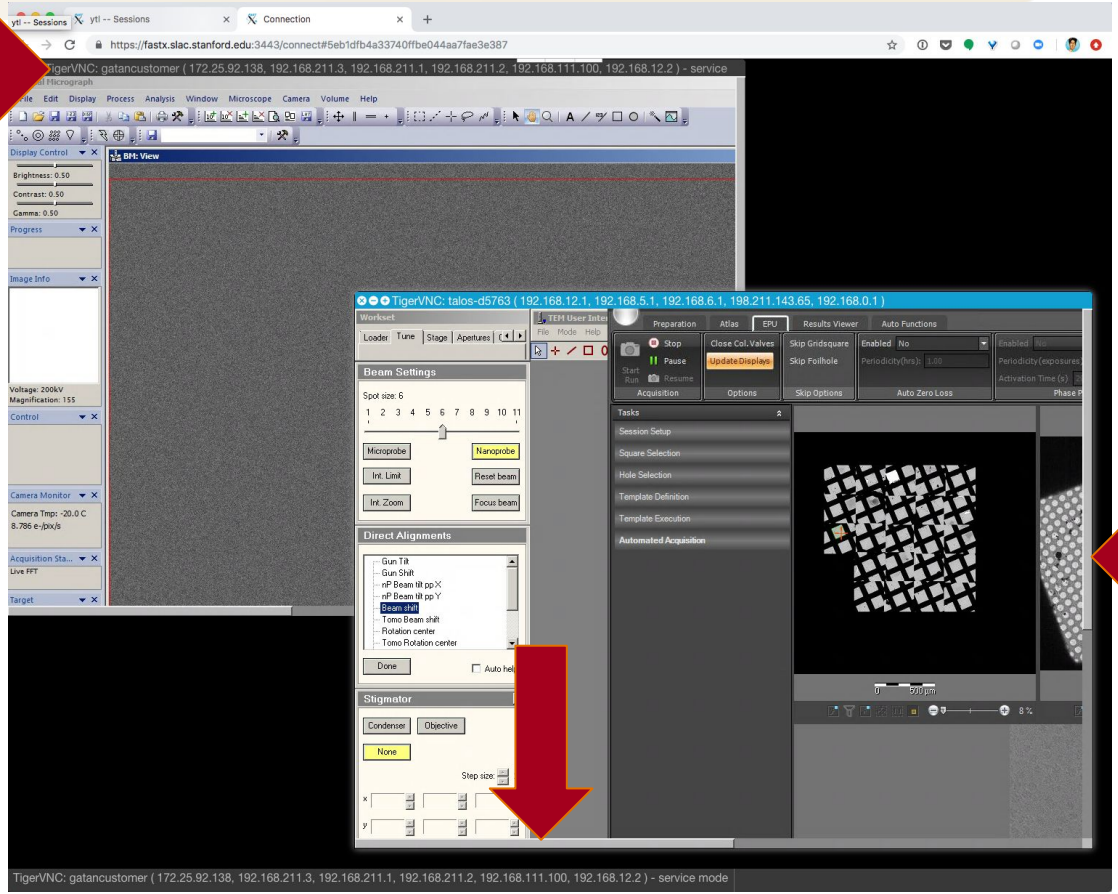
FastX - Select ‘Cryo-EM’



FastX - Select assigned TEM and ‘Launch’



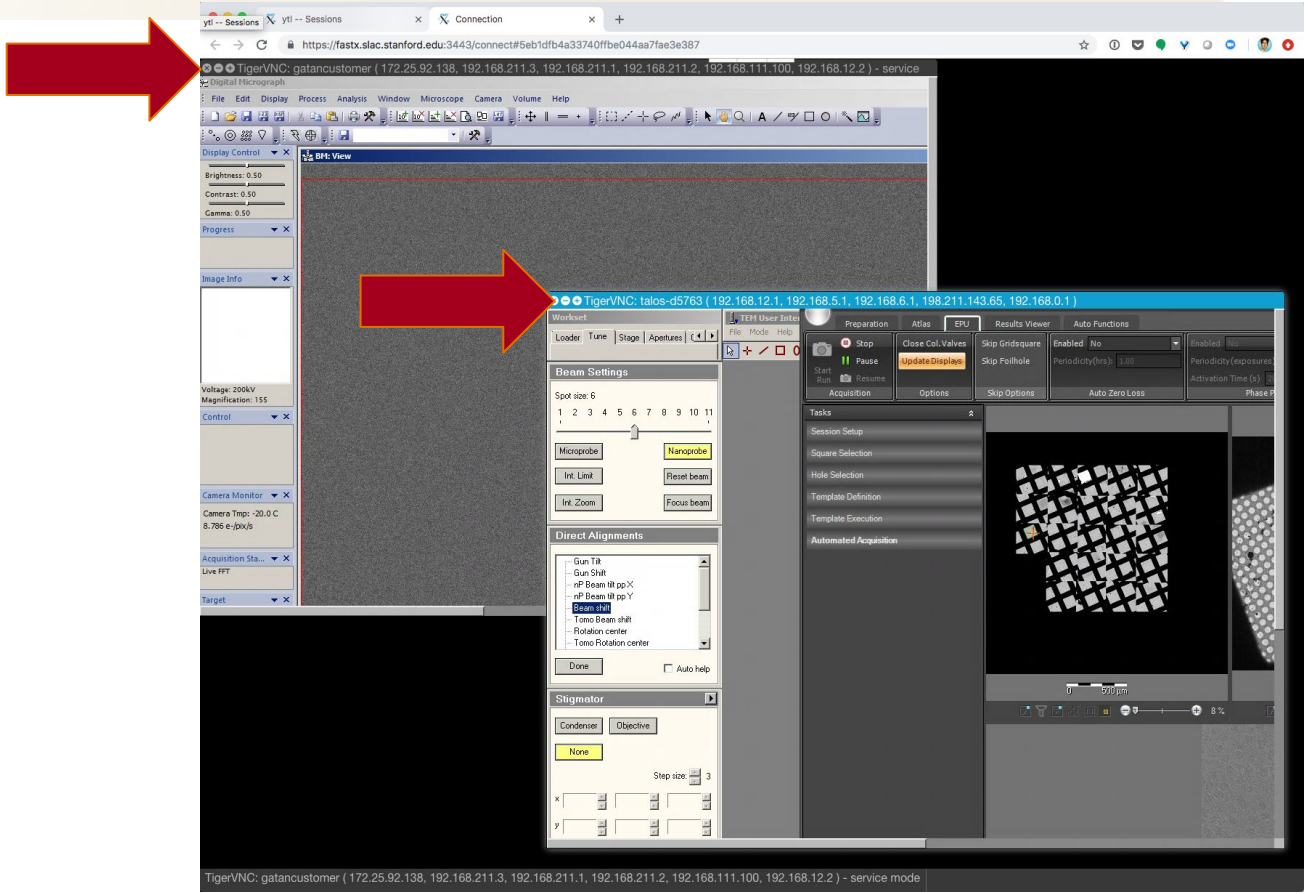
FastX - Two Windows (K2/3 & EPU/SerialEM)



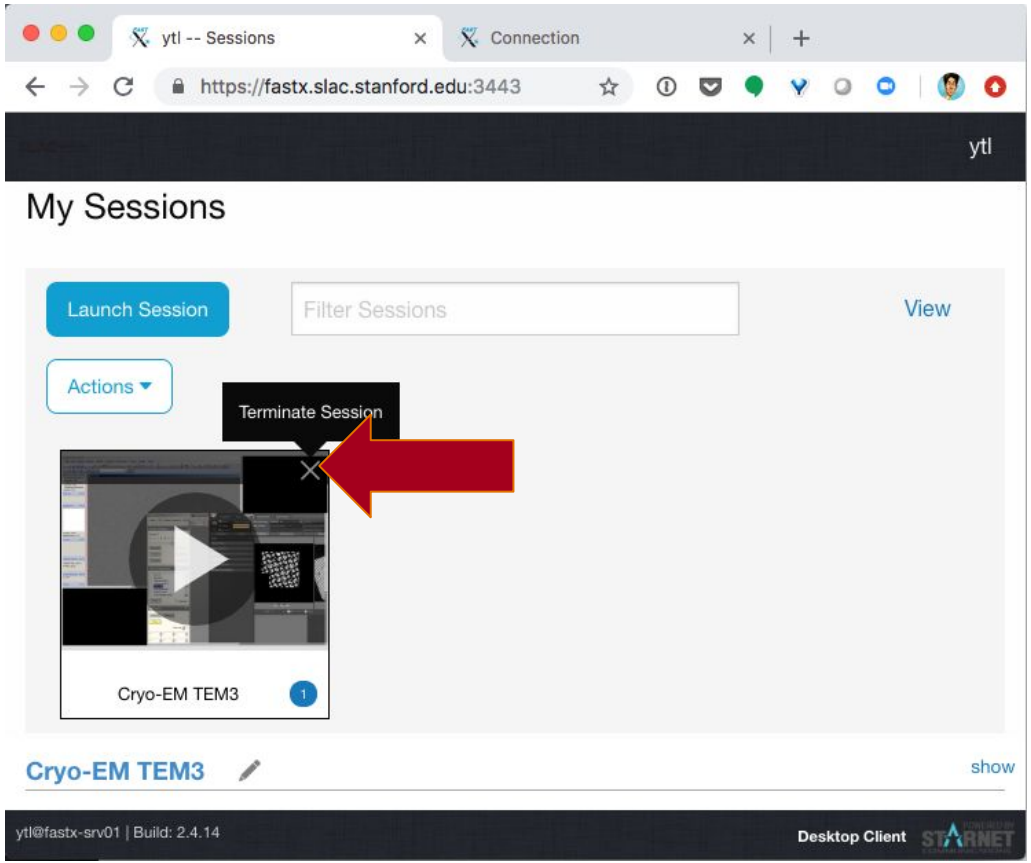
FastX - Exiting

- Session persists if you close your browser window
  - i.e. you will still be logged on
- Need to explicitly terminate session...

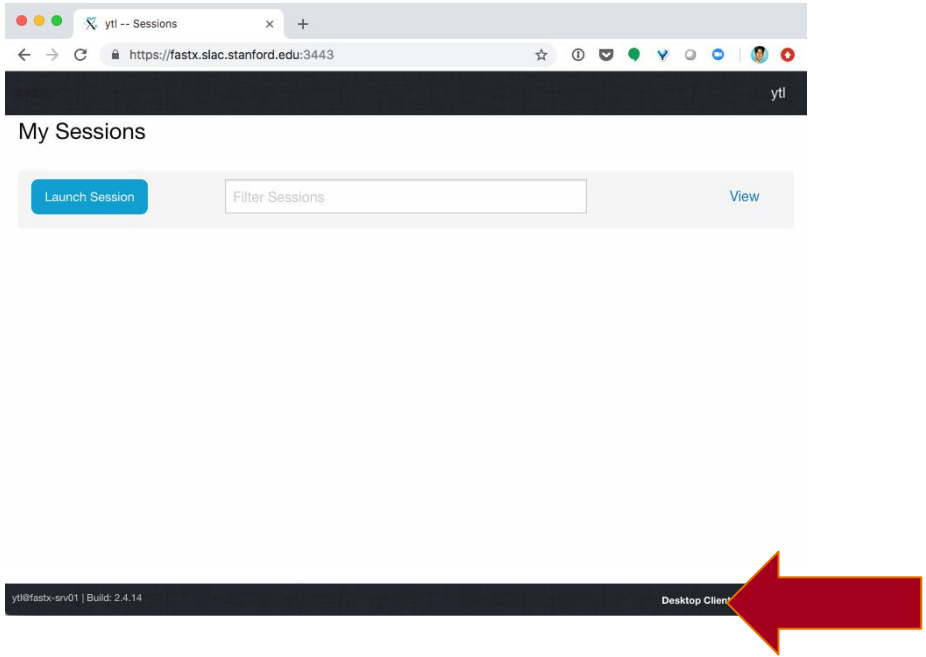
FastX - Exiting (option 1)

- 

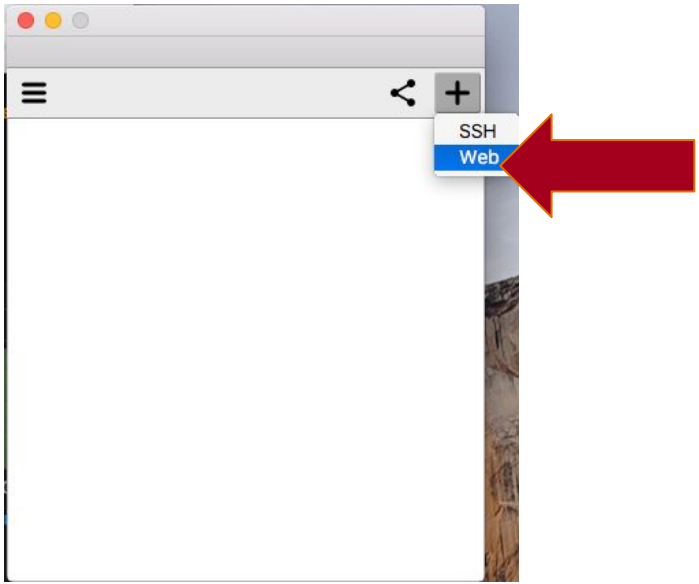
FastX - Exiting (option 2)

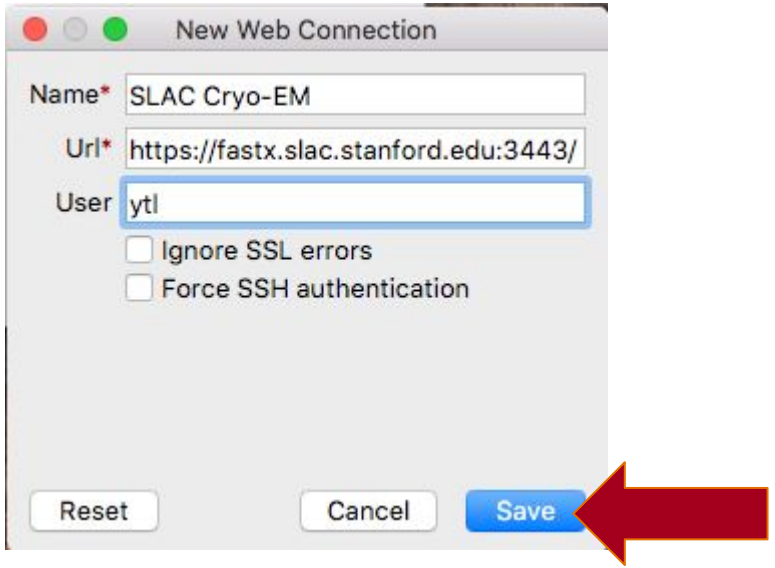


FastX - Download Native Client

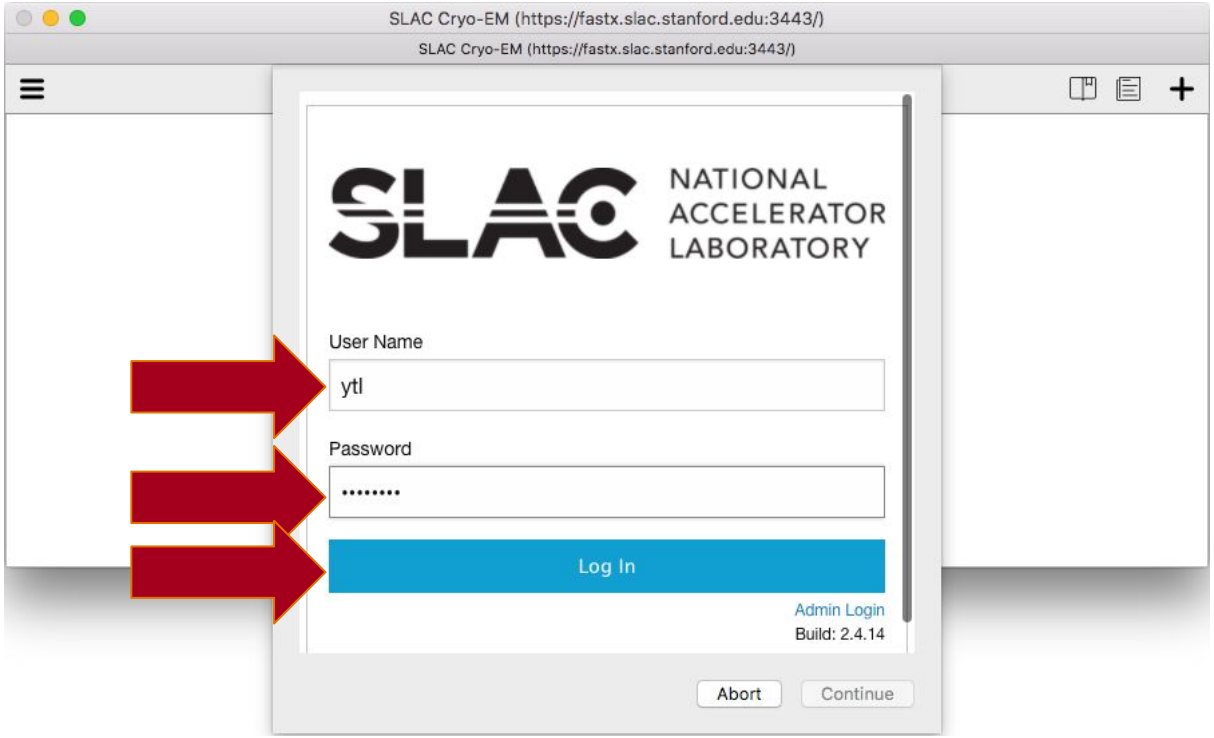


FastX - Configuring Native Client (1)

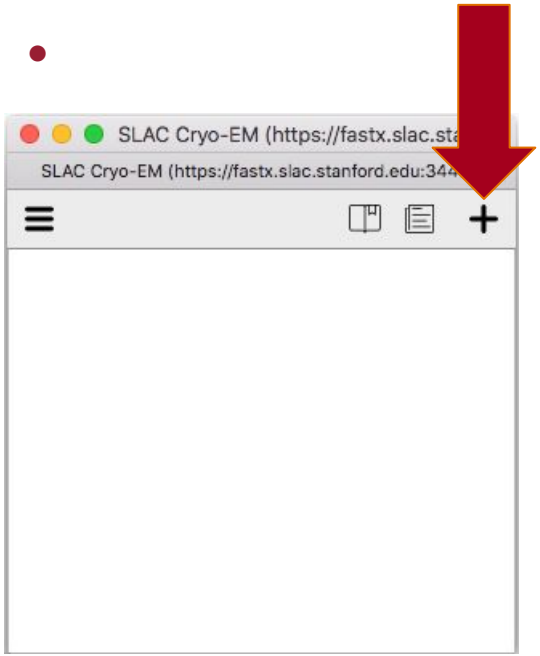
- 

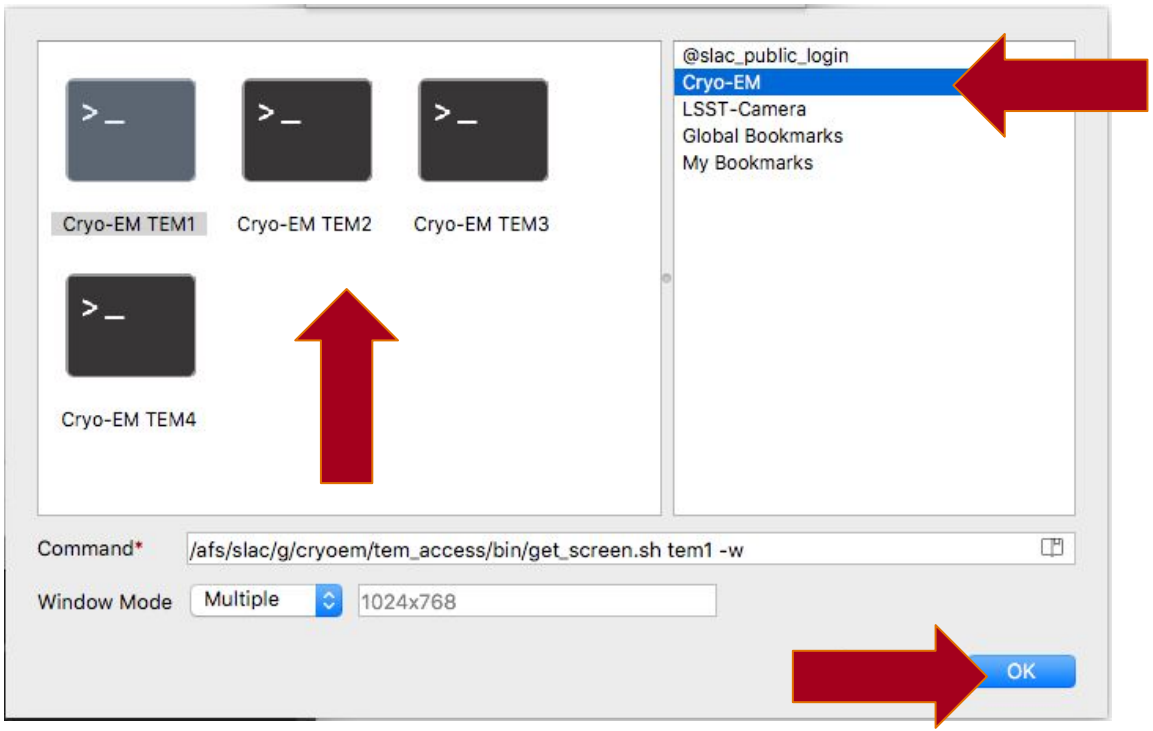


FastX - Configuring Native Client (2)

- 

FastX - Configuring Native Client (3)





1. Experimental Metadata

- Why?
  - Keep information about the experimental setup and image quality
- How?
  - Use an Electronic Notebook to store, and view information about experiment
- Where?
  - https://cryoem-logbook.slac.stanford.edu

Cryo-EM elogbook is the portal to your experiment at SLAC

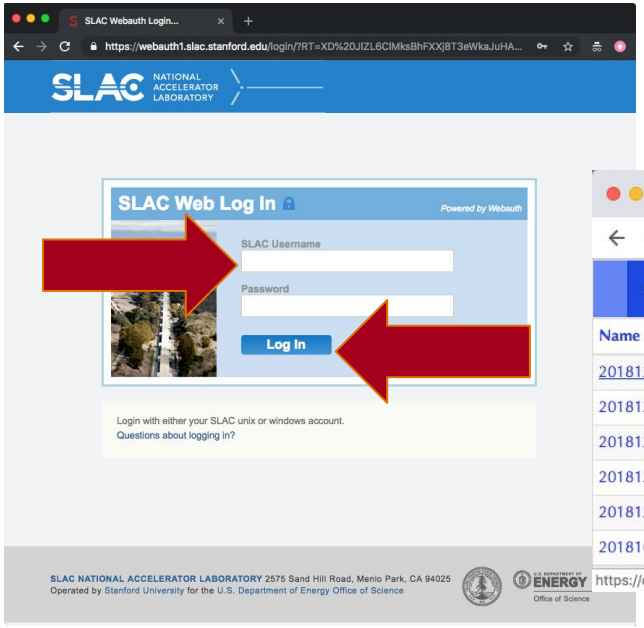
ELogbook

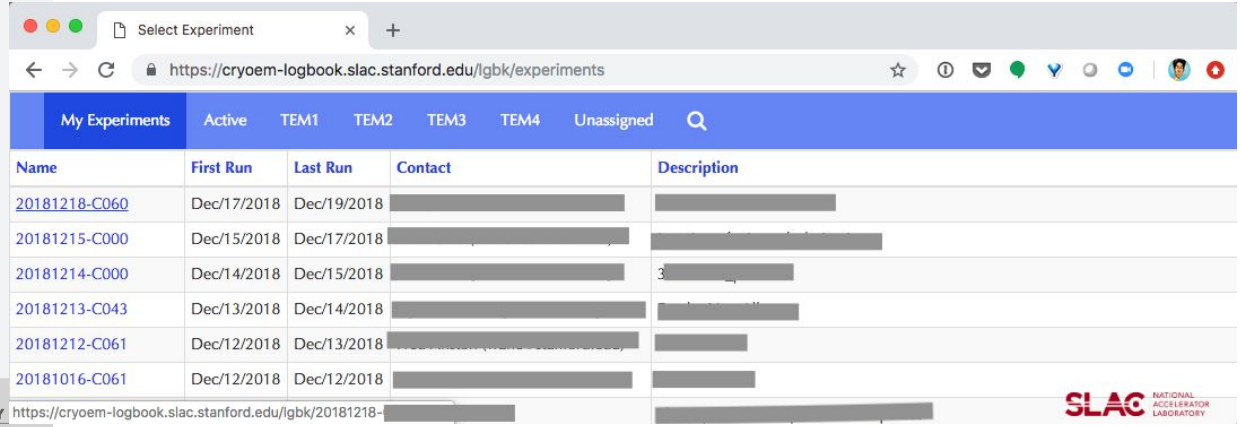
- Provide portal for all Cryo-EM experimental data
  - Assign samples
  - Annotate new metadata
  - Preview results
  - Set permissions for data on disk
  - Summary of image preprocessing
  - Initiate pre-processing pipeline

When should I use it?

- Upon new data for a new sample
- To register a change in sample during an experiment
- To allow collaborators access to data (on disk)
- To allow collaborators access to Remote Access to TEM
- To take notes/annotate the experiments (text, pictures etc)
- Obtain statistics about data quality
- To setup pre-processing

Accessing the eLogbook



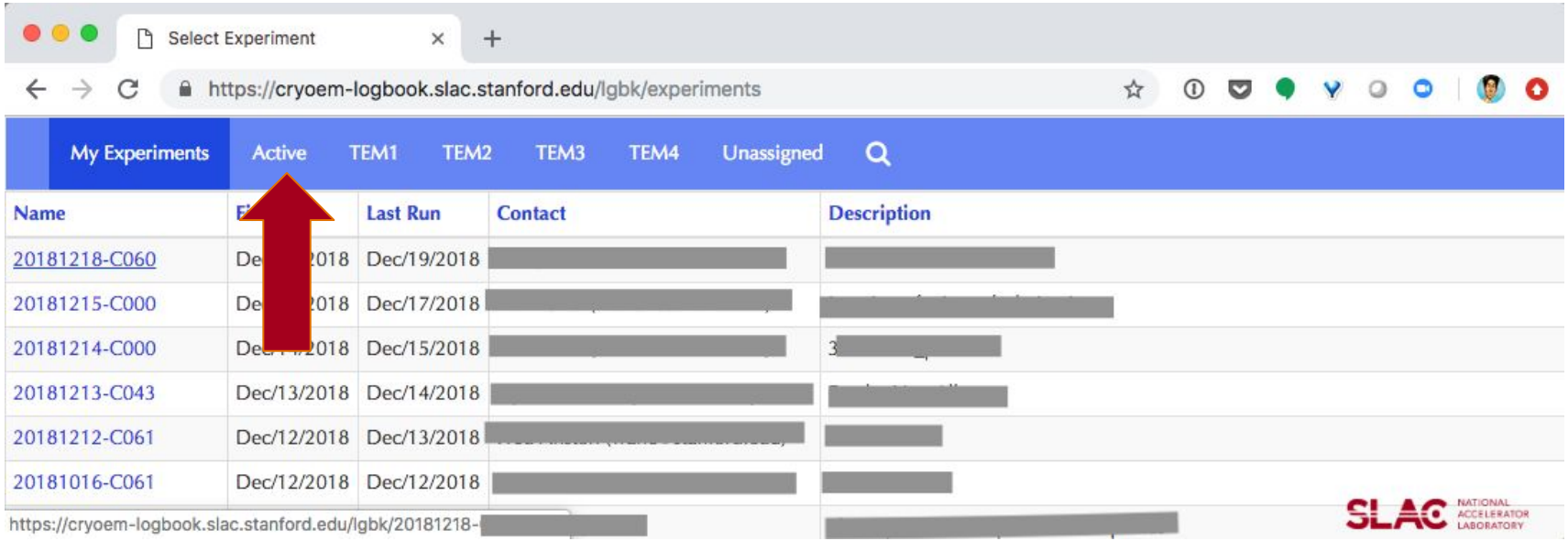


https://cryoem-logbook.slac.stanford.edu - use SLAC account

Experiment and Samples

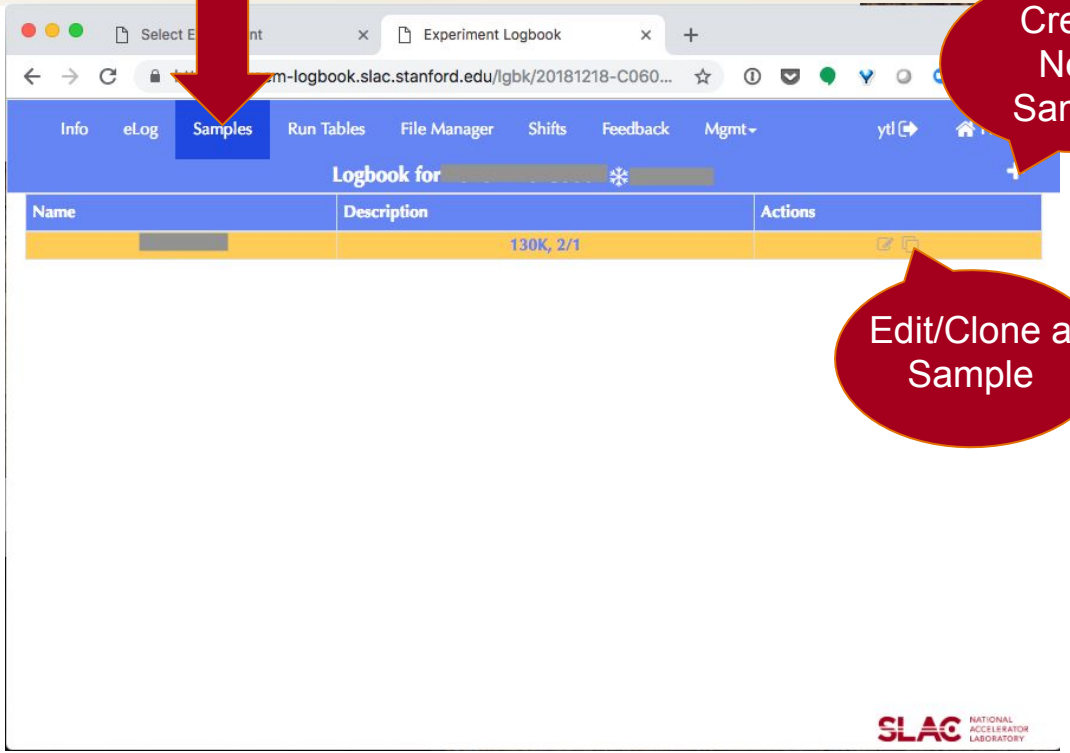
- What is a ‘Sample’?
  - A set of data (movies) collected under identical specimen and experimental (microscope) parameters.
- When should I create a new ‘Sample’?
  - When you either change specimen or EPU/SerialEM settings
- Data for each new Sample is copied to new directory (see later)

Is my experiment active?



If your experiment is not Active, then you will not be able to Remote Access!

Create a new Sample

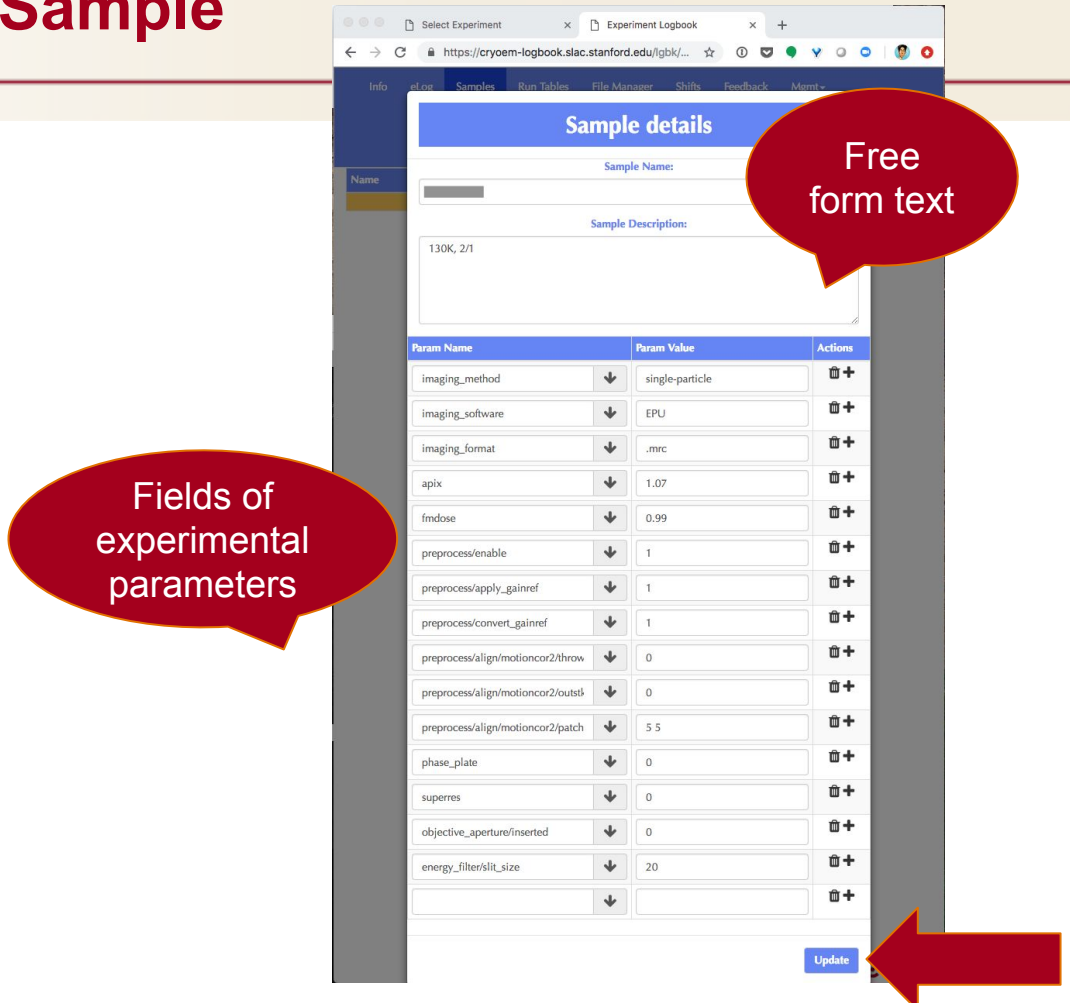


Yellow highlighted sample is Active

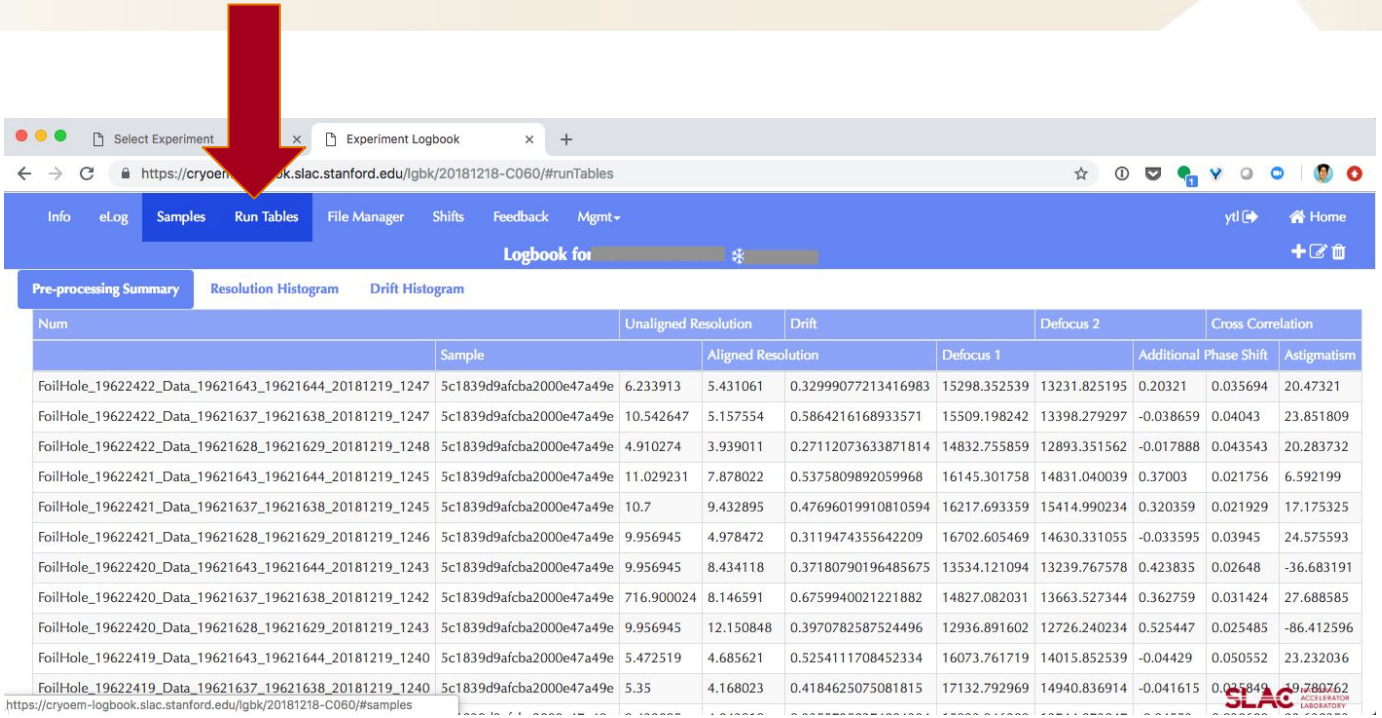
Sample Parameters

- Keep track of experimental parameters for each sample as key-value pair
- Keys
  - Selectable from a list (pre-defined)
  - Free form keys (user definable)
- Some keys are mandatory
- Most important keys:
  - fmdose, apix, superres, phase_plate

Edit a Sample

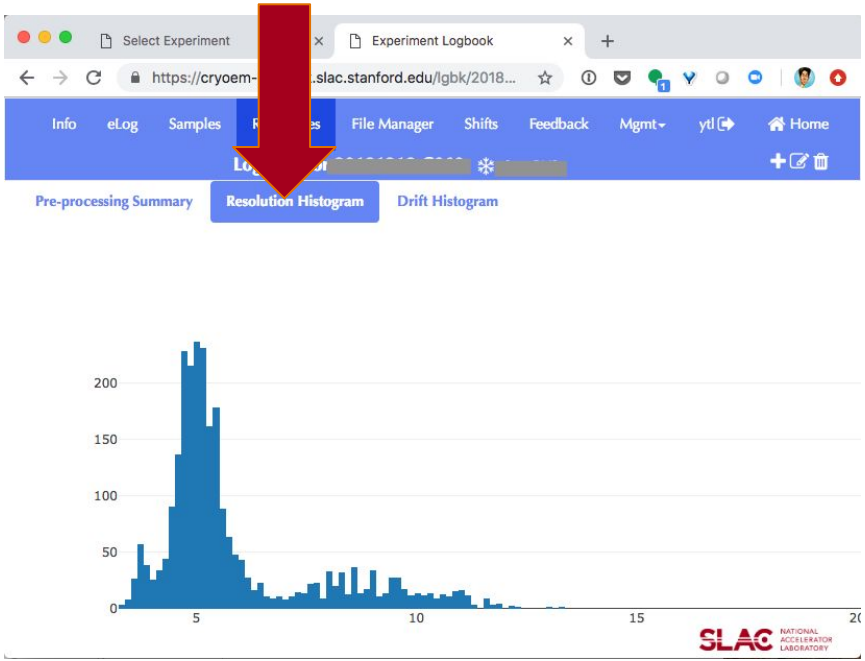


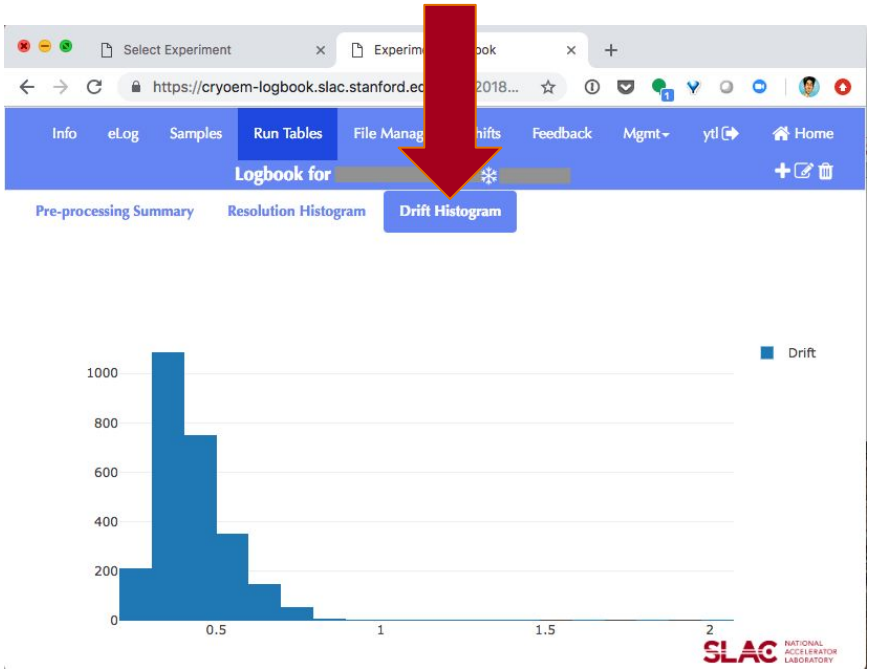
Run Tables



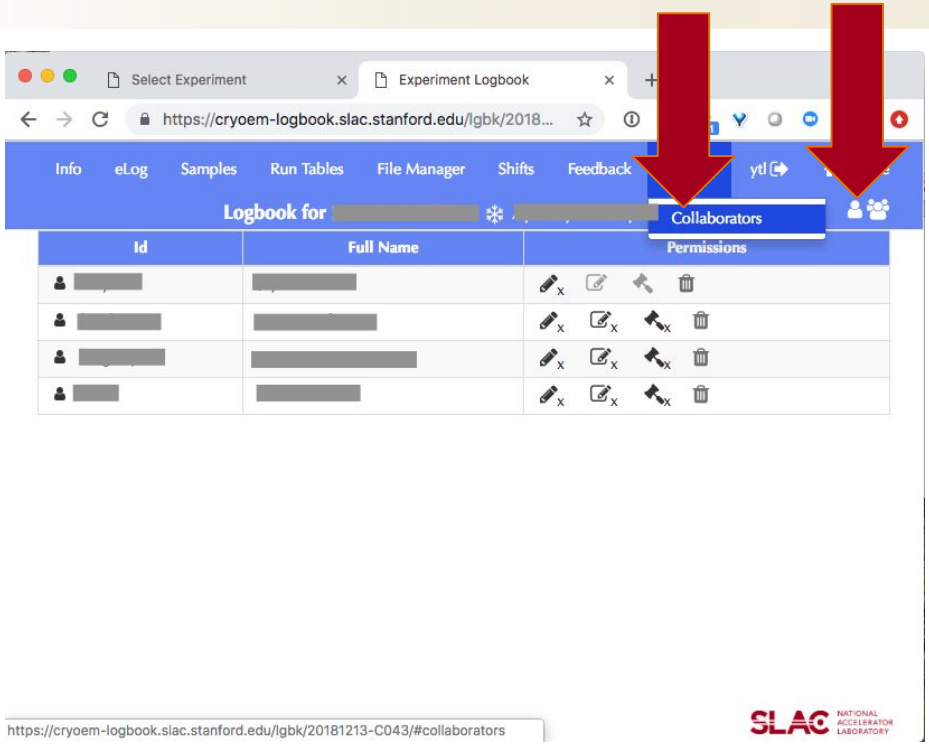
Num		Unaligned Resolution	Drift	Defocus 2		Cross Correlation	
	Sample	Aligned Resolution		Defocus 1		Additional Phase Shift	Astigmatism
FoilHole_19622422_Data_19621643_19621644_20181219_1247	5c1839d9afcba2000e47a49e	6.233913	5.431061	0.32999077213416983	15298.352539	13231.825195	0.20321	0.035694	20.47321
FoilHole_19622422_Data_19621637_19621638_20181219_1247	5c1839d9afcba2000e47a49e	10.542647	5.157554	0.5864216168933571	15509.198242	13398.279297	-0.038659	0.04043	23.851809
FoilHole_19622422_Data_19621628_19621629_20181219_1248	5c1839d9afcba2000e47a49e	4.910274	3.939011	0.27112073633871814	14832.755859	12893.351562	-0.017888	0.043543	20.283732
FoilHole_19622421_Data_19621643_19621644_20181219_1245	5c1839d9afcba2000e47a49e	11.029231	7.878022	0.5375809892059968	16145.301758	14831.040039	0.37003	0.021756	6.592199
FoilHole_19622421_Data_19621637_19621638_20181219_1245	5c1839d9afcba2000e47a49e	10.7	9.432895	0.47696019910810594	16217.693359	15414.990234	0.320359	0.021929	17.175325
FoilHole_19622421_Data_19621628_19621629_20181219_1246	5c1839d9afcba2000e47a49e	9.956945	4.978472	0.3119474355642209	16702.605469	14630.331055	-0.033595	0.03945	24.575593
FoilHole_19622420_Data_19621643_19621644_20181219_1243	5c1839d9afcba2000e47a49e	9.956945	8.434118	0.37180790196485675	13534.121094	13239.767578	0.423835	0.02648	-36.683191
FoilHole_19622420_Data_19621637_19621638_20181219_1242	5c1839d9afcba2000e47a49e	716.900024	8.146591	0.6759940021221882	14827.082031	13663.527344	0.362759	0.031424	27.688585
FoilHole_19622420_Data_19621628_19621629_20181219_1243	5c1839d9afcba2000e47a49e	9.956945	12.150848	0.3970782587524496	12936.891602	12726.240234	0.525447	0.025485	-86.412596
FoilHole_19622419_Data_19621643_19621644_20181219_1240	5c1839d9afcba2000e47a49e	5.472519	4.685621	0.5254111708452334	16073.761719	14015.852539	-0.04429	0.050552	23.232036
FoilHole_19622419_Data_19621637_19621638_20181219_1240	5c1839d9afcba2000e47a49e	5.35	4.168023	0.4184625075081815	17132.792969	14940.836914	-0.041615	0.035849	19.780762

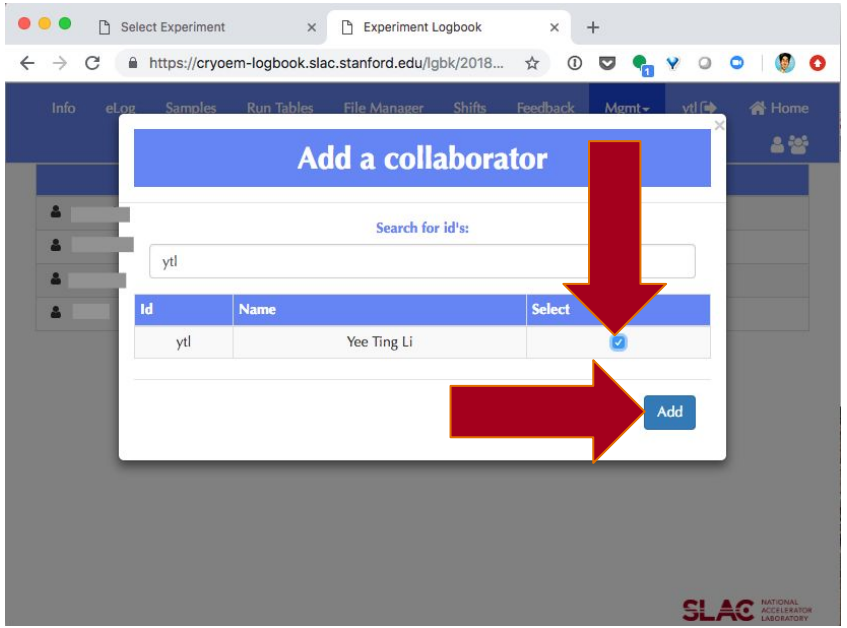
Run Tables


Select Experiment
Experiment Logbook
Info
eLog
Samples
File Manager
Shifts
Feedback
Mgmt
ytl
Home
Pre-processing Summary
Resolution Histogram
Drift Histogram
200
150
100
50
0
5
10
15
SLAC NATIONAL ACCELERATOR LABORATORY


Select Experiment
Info
eLog
Samples
Run Tables
File Manag
Feedback
Mgmt
ytl
Home
Logbook for
Pre-processing Summary
Resolution Histogram
Drift Histogram
Drift
1000
800
600
400
200
0
0.5
1
1.5
2
SLAC NATIONAL ACCELERATOR LABORATORY

Adding Collaborators





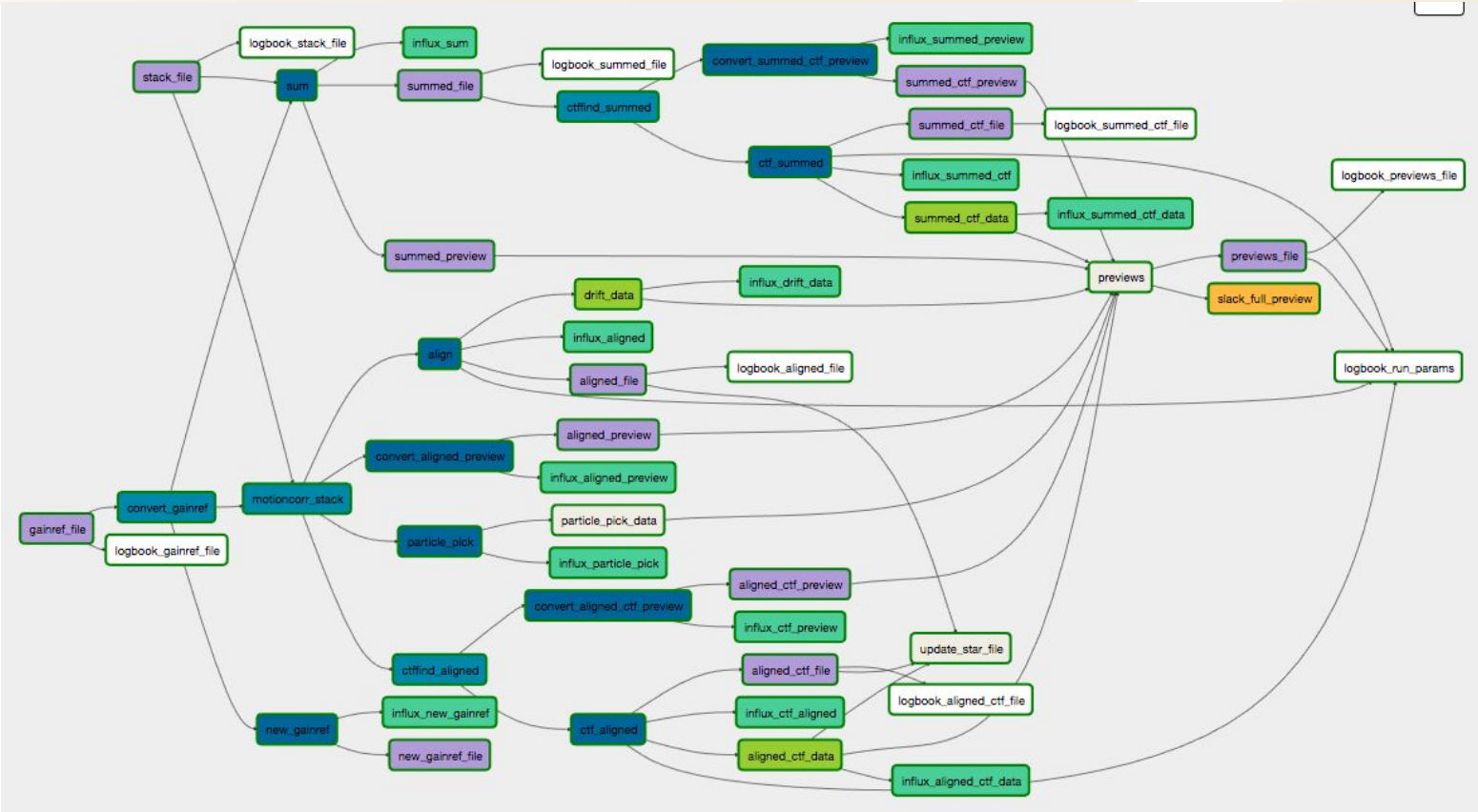
Added Collaborators will have Remote Access & Data Access to Experiment

1. Pre-Processing

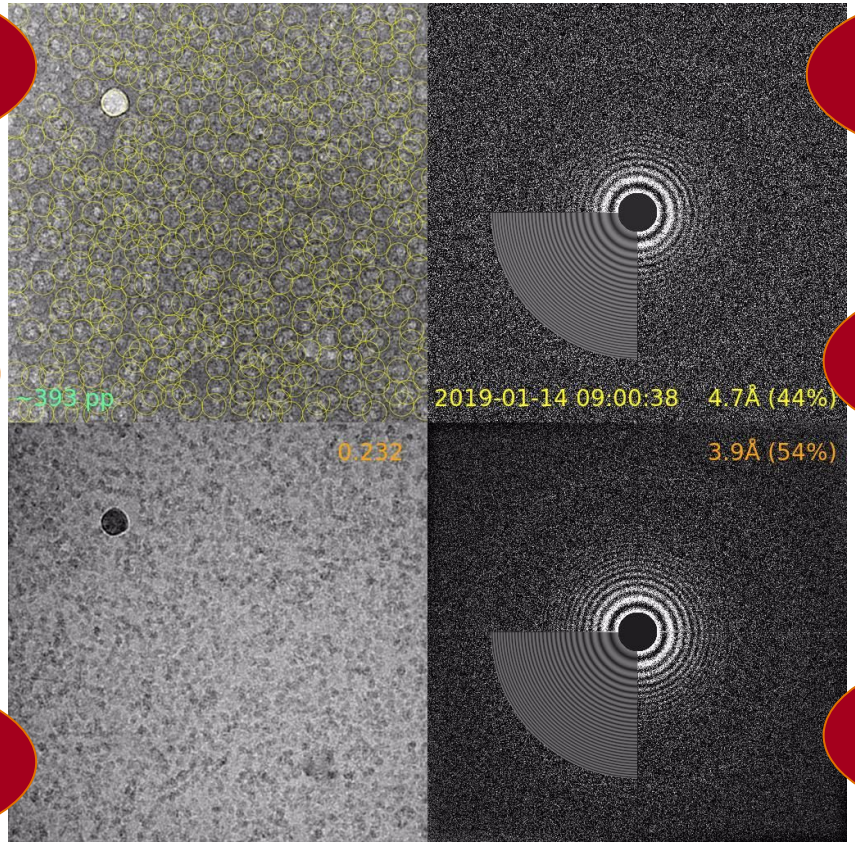
- Why?
  - Common tasks required prior to 2D classification; we provide aligned images and CTF ‘out-of-the-box’
- How?
  - As we stream the movies from the microscopes, we also do pre-processing and logging at the same time.
- Where?
  - Pre-processed data in the `./aligned` directory.

Aligned Images are previewed via SLACK

All Movies are put into Pipeline



Preview Generated



Particle Picked

CTF Before Alignment

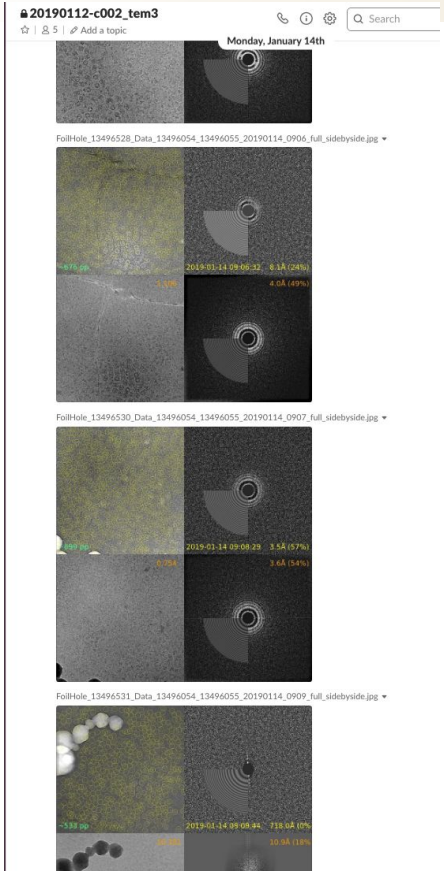
Num. Picked particles / RMS Drift

Est. Resolution / % of Nyquist

Aligned Image

CTF After Alignment

Live Stream of Previews



- Sign up at
  - **slac-cryoem.slack.com**
- Use home institution email for signup
- Operator may need to invite you to experiment

3. Where is my data?

SLAC

- Why?
  - Once your experiment is over, you want access to the raw and pre-processed data.
- How?
  - All data is kept at SLAC
  - Current plan is that data will be retained for ~5 years
- Where?
  - `/gpfs/slac/cryo/fs1/exp/`

All experimental data kept on Long Term Storage

Directory Layout

- Parent Directory: **/gpfs/slac/cryo/fs1/exp/**
- Each experiment kept under:
  - **./<YYYYMM>/<YYYYMMDD>-<PROP-ID>_<TEM#>/**
- Each experiment folder has n samples:

```
# pwd
/gpfs/slac/cryo/fs1/exp/201901/20190111-CS30_TEM3
# ls -lah
...
drwxrwsr-x+  7 ytl 4.0K Jan 11 19:37 5c391d7c79c214000e33cbdd
lrwxrwxrwx   1 ytl   24 Jan 11 17:34 vista -> 5c391d7c79c214000e33cbdd
...
```

Directory Layout (2)

- Each sample folder has:

```
$ ls -lah
total 256K
drwxrwsr-x+ 7 ytl 4.0K Jan 11 19:37 .
drwxrwsr-x+ 3 ytl 4.0K Jan 11 17:34 ..
drwxrwsr-x+ 3 ytl 4.0K Jan 11 17:41 aligned
-rw-rw-r--+ 1 ytl 139K Jan 12 07:25 images.star
drwxrwsr-x+ 3 ytl 4.0K Jan 11 17:42 particles
drwxrwsr-x+ 2 ytl  48K Jan 12 07:25 previews
drwxrwxrwx+ 6 ytl 4.0K Jan 12 12:42 raw
drwxrwsr-x+ 3 ytl 4.0K Jan 11 17:41 summed
```

Raw Data: ./raw/

- Duplication of files kept on the microscopes
  - Ensure EPU/SerialEM is configured to store all data to **X:** drive
- Do NOT move files around on the microscope computers
  - It will cause duplication of data as its streamed onto disk

Aligned Data: `./aligned/`

- Pre-processed data kept here
- Aligned image:
  - `./motioncor2/<version>/`
    - `<filename>_aligned.mrc`
    - `<filename>_aligned_DW.mrc`
    -
- CTFs
  - `./motioncor2/<version>/ctffind4/<version>/`
    - `<filename>_aligned_ctf.txt`
    - `<filename>_aligned_ctf.mrc`

How do I copy my data?

- Can use standard UNIX tools
  - scp, rsync
    - `dtn01.slac.stanford.edu`
- Globus Online
  - Send ytl@slac.stanford.edu an email to enable your SLAC account
    - https://www.globus.org/
    - `slac#cryoem`

RSYNC

- Copy entire experiment:

3+ TB

```
rsync -avz \
  dtn01.slac.stanford.edu:/gpfs/slac/cryo/fs1/exp/201901/20190107-CS01_TEM4 \
  .
```

- Copy just aligned images:

~500 GB

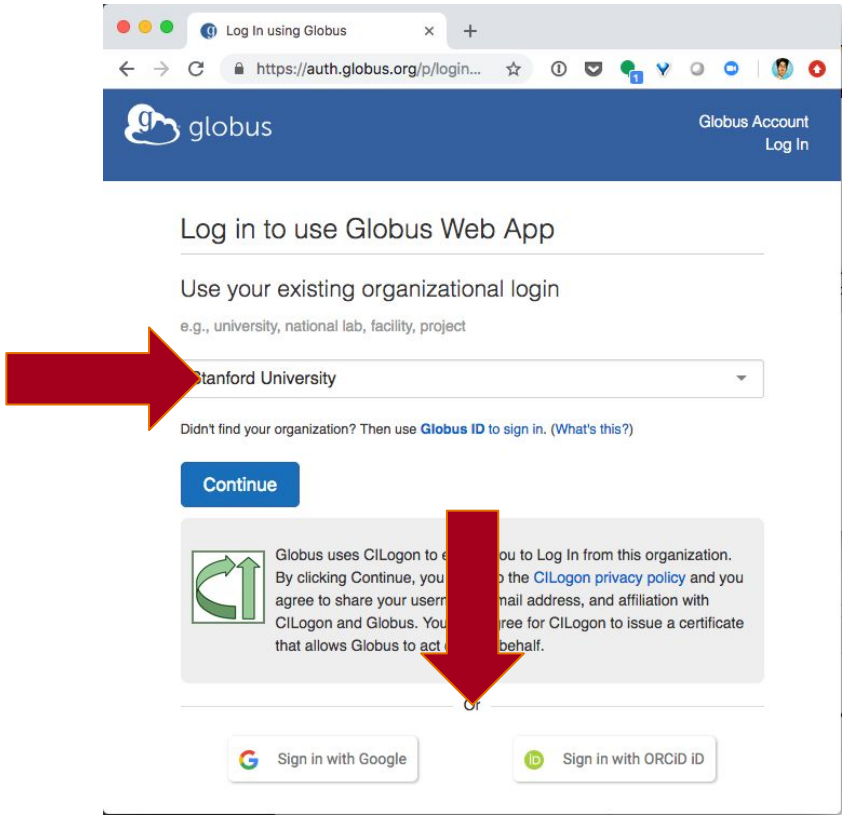
```
rsync -avz \
  dtn01.slac.stanford.edu:/gpfs/slac/cryo/fs1/exp/201901/20190107-CS01_TEM4/aligned
\
  .
```

- Copy just raw images:

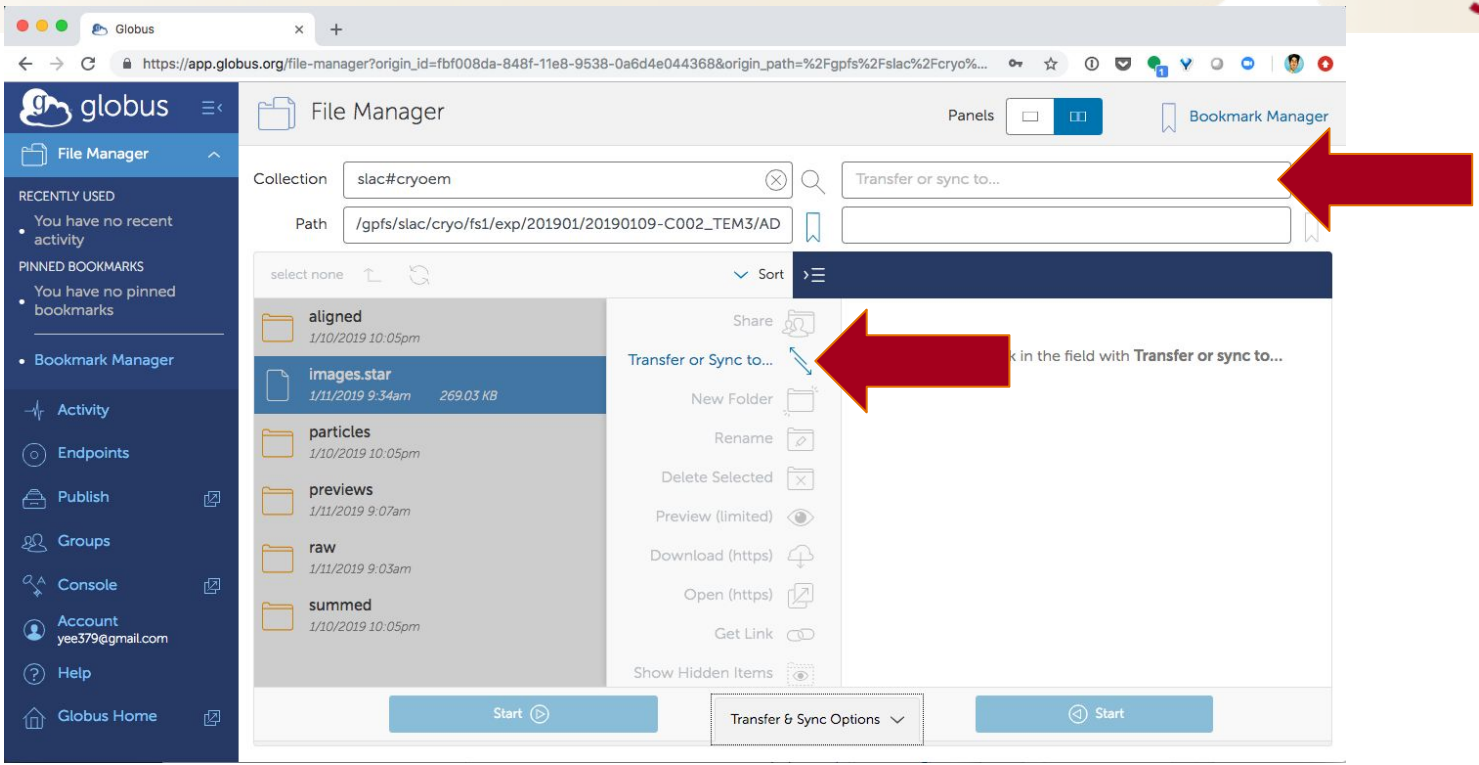
2+ TB

```
rsync -avz \
  dtn01.slac.stanford.edu/gpfs/slac/cryo/fs1/exp/201901/20190107-CS01_TEM4/raw \
  .
```

Globus Online

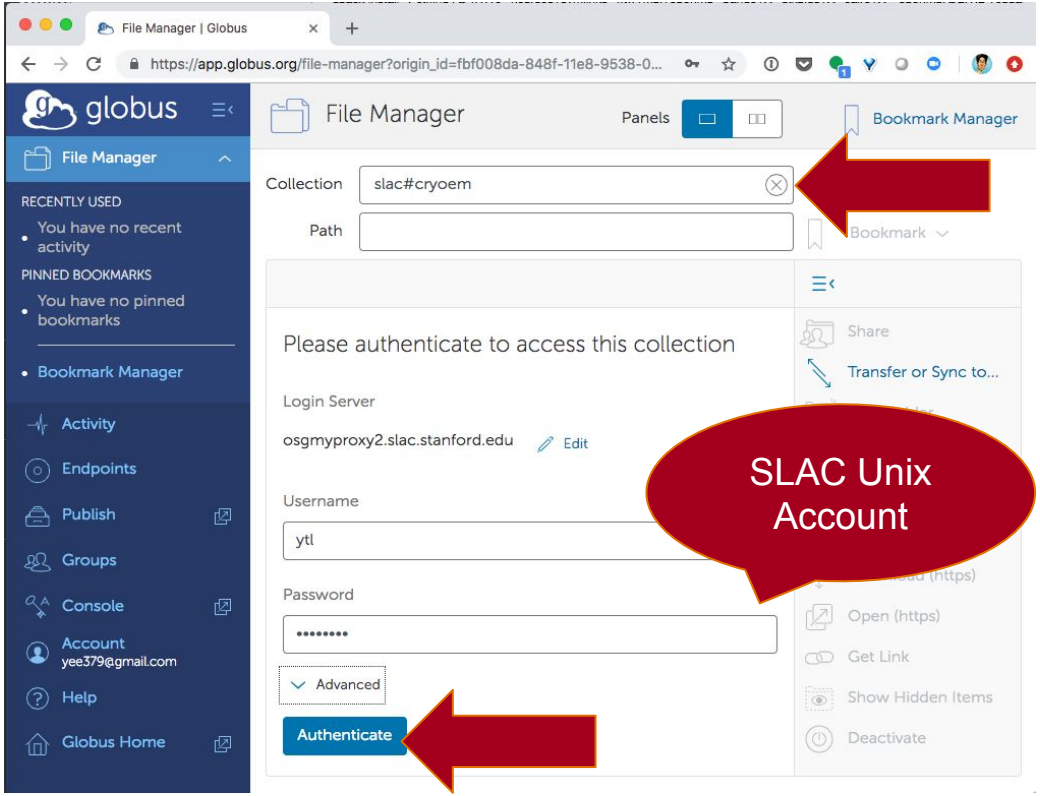


Globus Online



Download and Install ‘Globus Personal Connect’ to transfer to your Desktop

Globus Online



Other transfer tools

- CyberDuck
- WinSCP
-

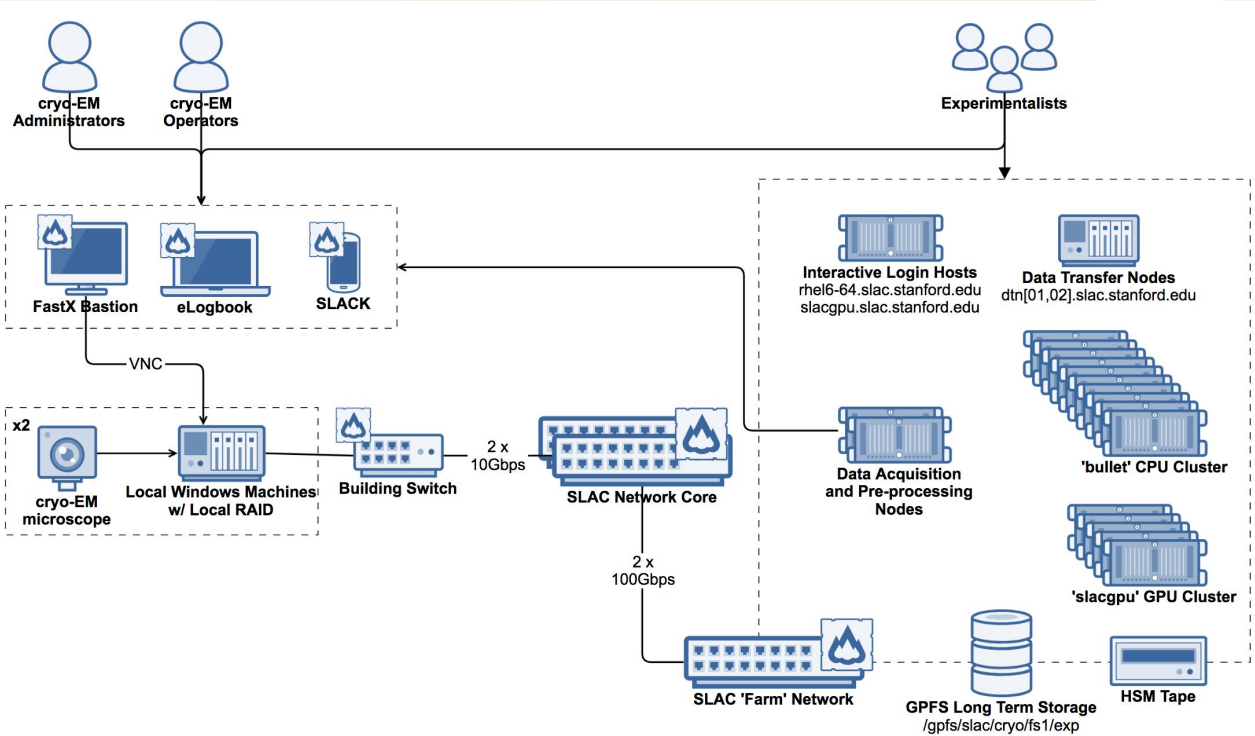
Summary

- Remote Access to TEM
  - **https://fastx.slac.stanford.edu:3443**
- Samples in eLogbook
  - **Activate new samples** for each new microscope config
- Pre-processing Pipeline and near-real-time previews
  - **slac-cryoem.slack.com**
- Data Transfer
  - **slac#cryoem** on Globus
  - **dtn01.slac.stanford.edu** for rsync etc.

Questions?

-

DAQ and Data Management Overview



Provide ‘Fast Feedback’ to Users