RAPD Data Analysis at NE-CAT

MCE2021 Workshop

Frank Murphy – March 17, 2021

Agenda

NE-CAT

Automated data analysis @ NE-CAT

Current capabilities

Development of version 2

- Design & implementation
- ~~Problems~~ Opportunities

Acknowledgements

Contributors
Jon Schuermann*
David Neau
Kay Perry
Jim Withrow

NE-CAT Support Staff
Surajit Banerjee
Ali Kaya
Igor Kourinov
N. Sukumar

Support
NIGMS P30 GM124165

NE-CAT Beamlines

APS Sector 24

24-ID-C

Tunable 6.5-20keV

1.5X10^{13} photons/sec

25µmH X 10µmV

Eiger2 X 16M

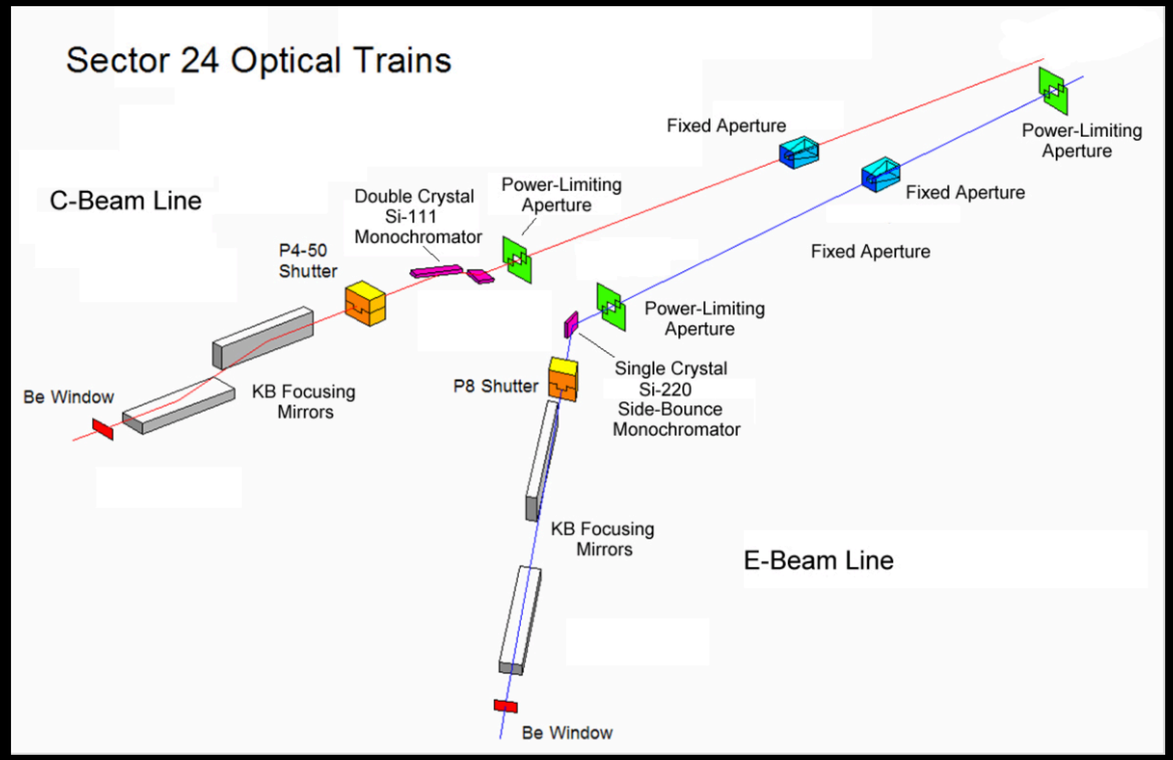
24-ID-E

Fixed 12,663eV

1.0X10^{13} photons/sec

35µmH X 5µmV

Eiger 16M



Our Goals For Automated Data Analysis

- Help users collect better data faster
- Enable collaboration with off-site users
- Project started in 2009

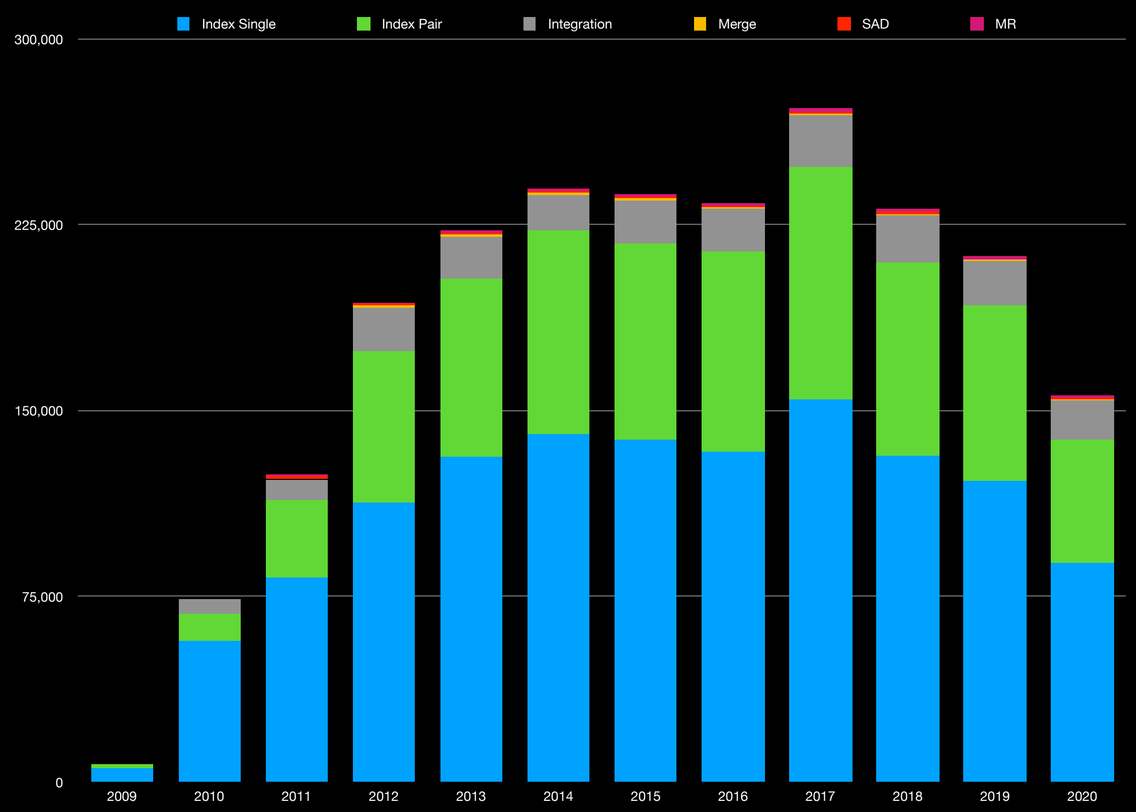
Current Status – RAPDv1

RAPD Capabilities

- Indexing & Strategy
- Integration, Scaling & Analysis
- Merging
- SAD
- MR

RAPD Capabilities

- Indexing & Strategy – DISTL, Labelit, BEST, Mosflm
- Integration, Scaling & Analysis – XDS, Aimless, Pointless, Phaser, Xtriage
- Merging – Aimless, Pointless
- SAD – SHELX, Phenix
- MR – Phaser



RAPD Index

rapd.nec.aps.anl.gov

Data | Structure

Snaps | Runs

NE38_7_PAIR_0_1+2.img
NE38_7_PAIR_0_000002.img
NE38_7_PAIR_0_000001.img
NE38_6_PAIR_0_1+2.img
NE38_6_PAIR_0_000002.img
NE38_6_PAIR_0_000001.img
NE38_5_PAIR_0_1+2.img
NE38_5_PAIR_0_000002.img
NE38_5_PAIR_0_000001.img
NE38_4_PAIR_0_1+2.img
NE38_4_PAIR_0_000002.img
NE38_4_PAIR_0_000001.img
NE38_3_PAIR_0_1+2.img
NE38_3_PAIR_0_000002.img
NE38_3_PAIR_0_000001.img
NE38_2_0_000002.img
NE38_2_0_000001.img
NE38_1_PAIR_0_1+2.img
NE38_1_PAIR_0_000002.img
NE38_1_PAIR_0_000001.img

Summary | Detail | Plots | Snaps | Runs | Samples

Labelit autoindexing summary for:

Image: /gpfs1/users/cornell/chem7880_C_5704/images/haoyue/snaps/NE38_7_PAIR_0_000001.cbf

Image: /gpfs1/users/cornell/chem7880_C_5704/images/haoyue/snaps/NE38_7_PAIR_0_000002.cbf

Space Group	a	b	c	α	β	γ	mosaicity	resolution
P4	57.83	57.83	150.13	90.00	90.00	90.00	0.100000	1.25

Detector distance

Resolution limit at edge	Detector distance
1.2	150
1.3	180.0
1.4	200.0

Data collection strategy from BEST

N	Omega Start	Omega End	Rot Range	N of Images	Delta Omega	Exposure time	% Transmission
1	82.00	140.0	58.0	290	0.20	0.3	1.5

ANOMALOUS data collection strategy from BEST

N	Omega Start	Omega End	Rot Range	N of Images	Delta Omega	Exposure time	% Transmission
1	78.00	170.0	92.0	368	0.25	0.2	1.5

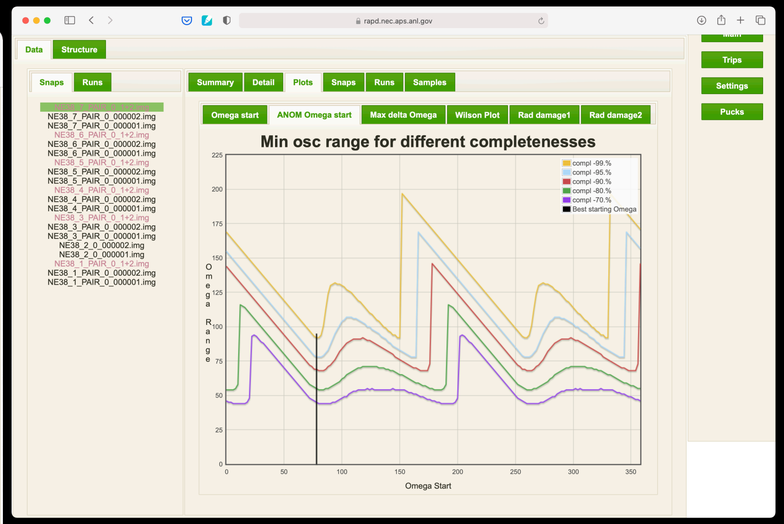
Controls

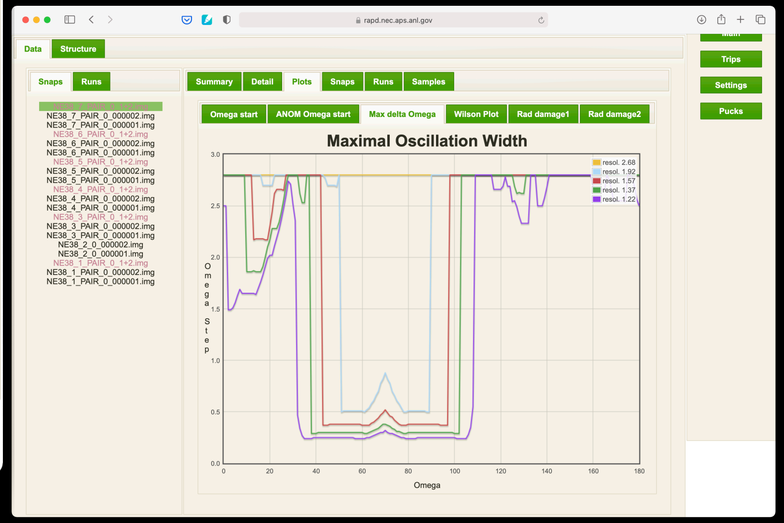
Main

Trips

Settings

Pucks





RAPD Integrate

rapd.nec.aps.anl.gov

Data | Structure

Snaps | Runs

NE38_7_1
NE38_1_1
NE38_5_1
NE38_4_1
NE38_3_1
NE38_2_1

Summary | Detail | Plots | Analysis | Snaps | Runs | Samples

Processing Results for NE38_1_1

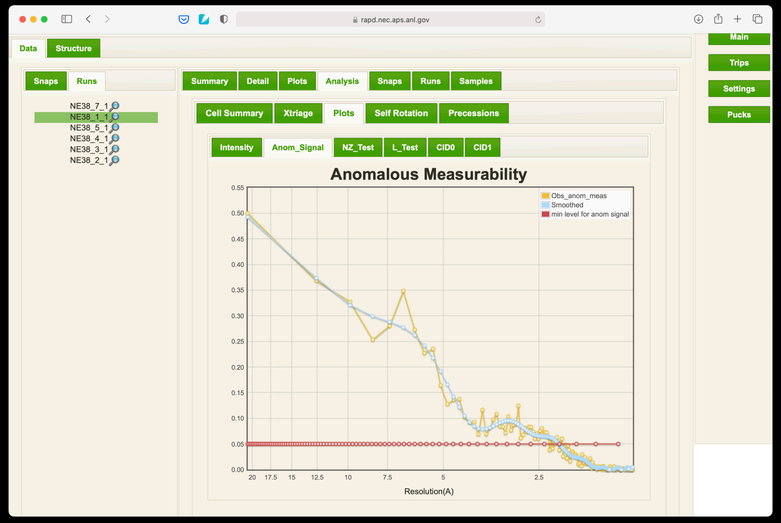
Images 1-1800

Spacegroup: P 41 21 2
Unit Cell: 57.85 57.85 150.15 90.00 90.00 90.00
SIGMAR (Mosaicity): 0.09329°
Asymptotic limit of I/sigma (ISa) = 23.81

	Overall	Inner Shell	Outer Shell
High resolution limit	1.27	6.94	1.27
Low resolution limit	150.15	150.15	1.29
Completeness	99.9	100.0	98.6
Multiplicity	25.7	18.2	24.0
I/sigma	30.2	81.7	3.5
CC(1/2)	0.999	0.999	0.875
Rmerge	0.102	0.081	1.055
Rmerge (anomalous)	0.100	0.079	1.037
Rmeas	0.104	0.083	1.078
Rmeas (anomalous)	0.104	0.082	1.080
Rpim	0.020	0.018	0.215
Rpim (anomalous)	0.028	0.023	0.299
Anomalous completeness	99.8	100.0	96.8
Anomalous multiplicity	13.7	12.1	12.7
Anomalous correlation	-0.131	0.047	-0.089
Anomalous slope	1.060	--	--
Total observations	1763371	9851	78787

Controls

Main
Trips
Settings
Pucks



RAPD SAD

rapd.nec.aps.anl.gov

Data | Structure

SAD | MR | MAD

NE38_7_1
NE38_1_1

Summary | Plots | Autobuild

SHELXC Results

SAD

Resolution	5.73	3.53	2.66	2.17	1.86	1.64	1.47	1.34	1.23	1.14	1.07
N(data)	906	2616	4392	6311	8030	9896	12142	13801	16625	18748	19463
Chi^2	1.19	1.02	1.04	1.18	1.35	1.53	1.54	1.51	1.45	1.35	0.80
<I/sig>	162.3	182.0	154.5	135.7	107.4	67.3	44.3	26.5	16.7	11.5	5.2
% Complete	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.1
<d"/sig>	3.60	2.37	2.13	2.07	1.72	1.32	1.04	0.85	0.77	0.73	0.69
CC(1/2)	95.7	93.4	79.4	70.1	52.1	31.7	19.4	10.4	5.1	2.5	-3.0

SHELXD Results

Space Group	Max CCall	Max CCweak	Max CFOM	Max PATFOM
P41212 (P43212)	**29.41**	**20.27**	**49.68**	**9.61**
P4122 (P4322)	22.39	14.72	37.11	3.61
P4212	9.13	5.02	14.14	2.94
P422	8.55	4.53	13.08	3.1
P42212	8.55	4.71	13.27	2.79
P4222	10.61	6.00	16.6	2.63

SHELXE Results

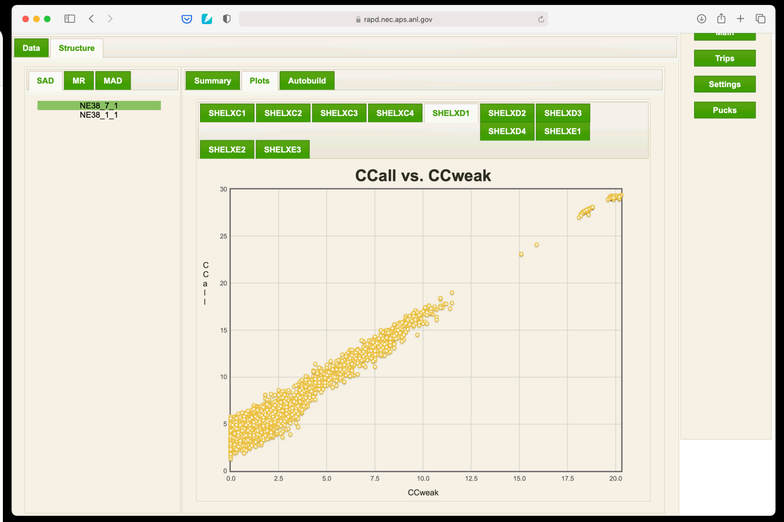
SHELXE may have found a solution in P41212 with a solvent fraction of 0.55

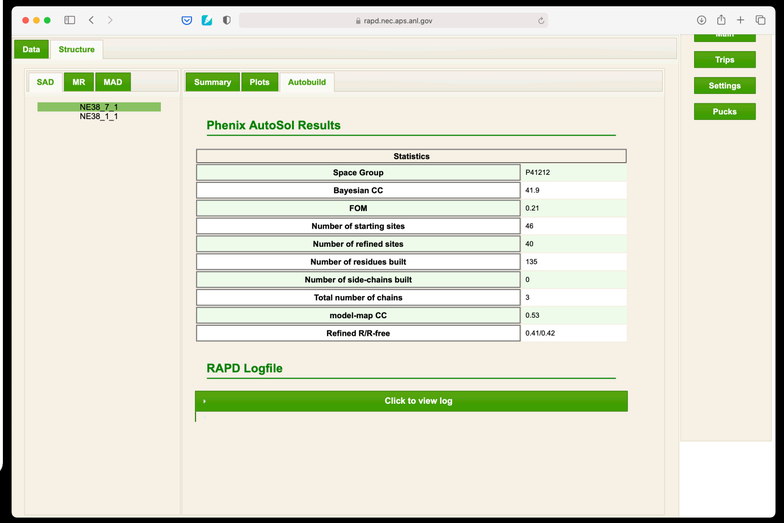
Site	x	y	z	occ*Z	density
1	0.5121	0.0412	0.0813	34.0000	70.78
2	0.4828	-0.2186	-0.0078	31.8478	63.24
3	0.2879	-0.0853	-0.0604	22.9296	43.89
4	0.5232	0.1897	0.0867	19.9104	38.48
5	0.5288	-0.0034	0.0945	17.0748	29.60
6	0.4847	0.0388	0.1893	14.4772	30.83
7	0.8224	-0.0473	0.1008	11.0296	18.88
8	0.1737	0.0207	-0.0578	8.6598	15.94

SHELXE FOM	0.842

Controls

Main | Trips | Settings | Pucks

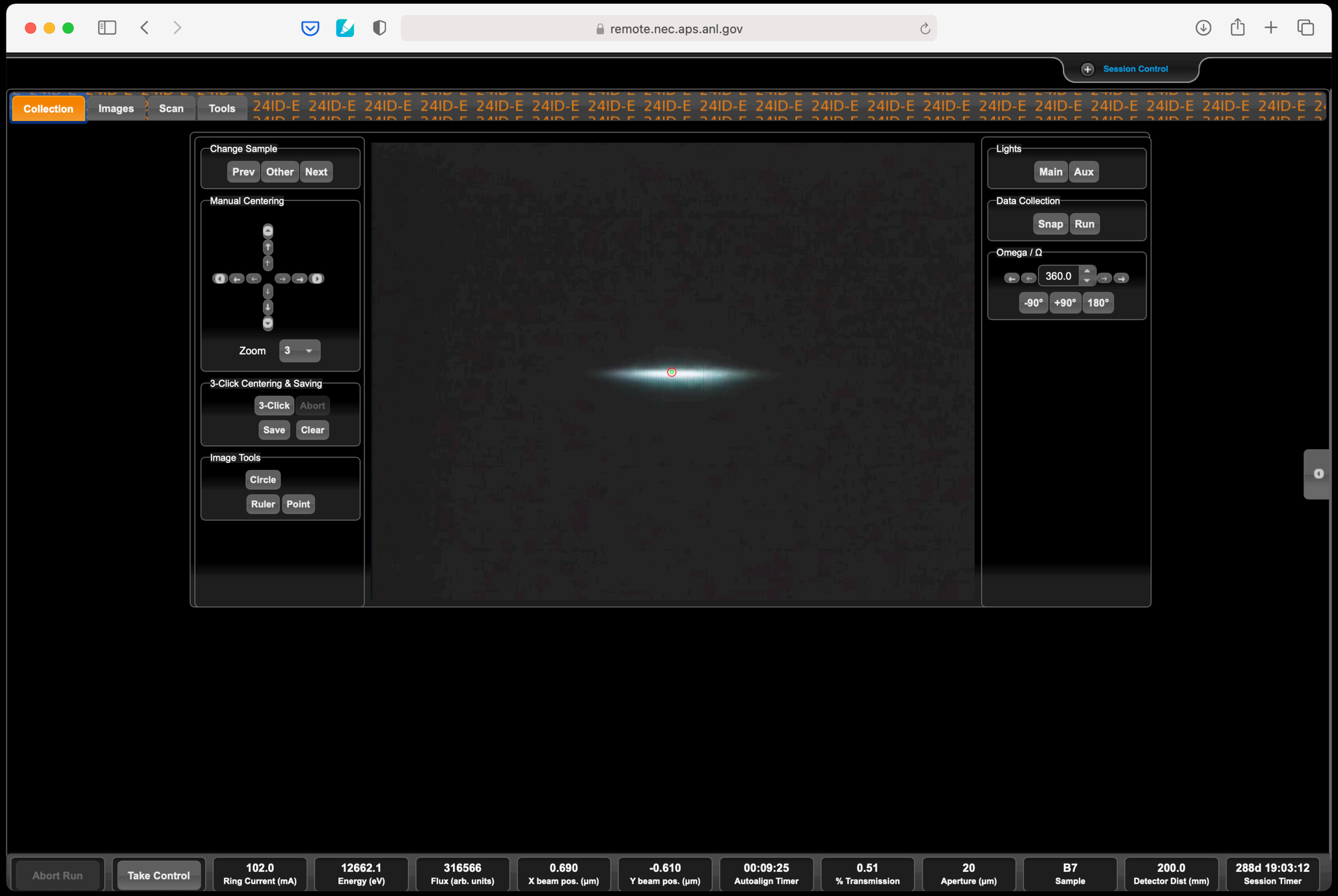




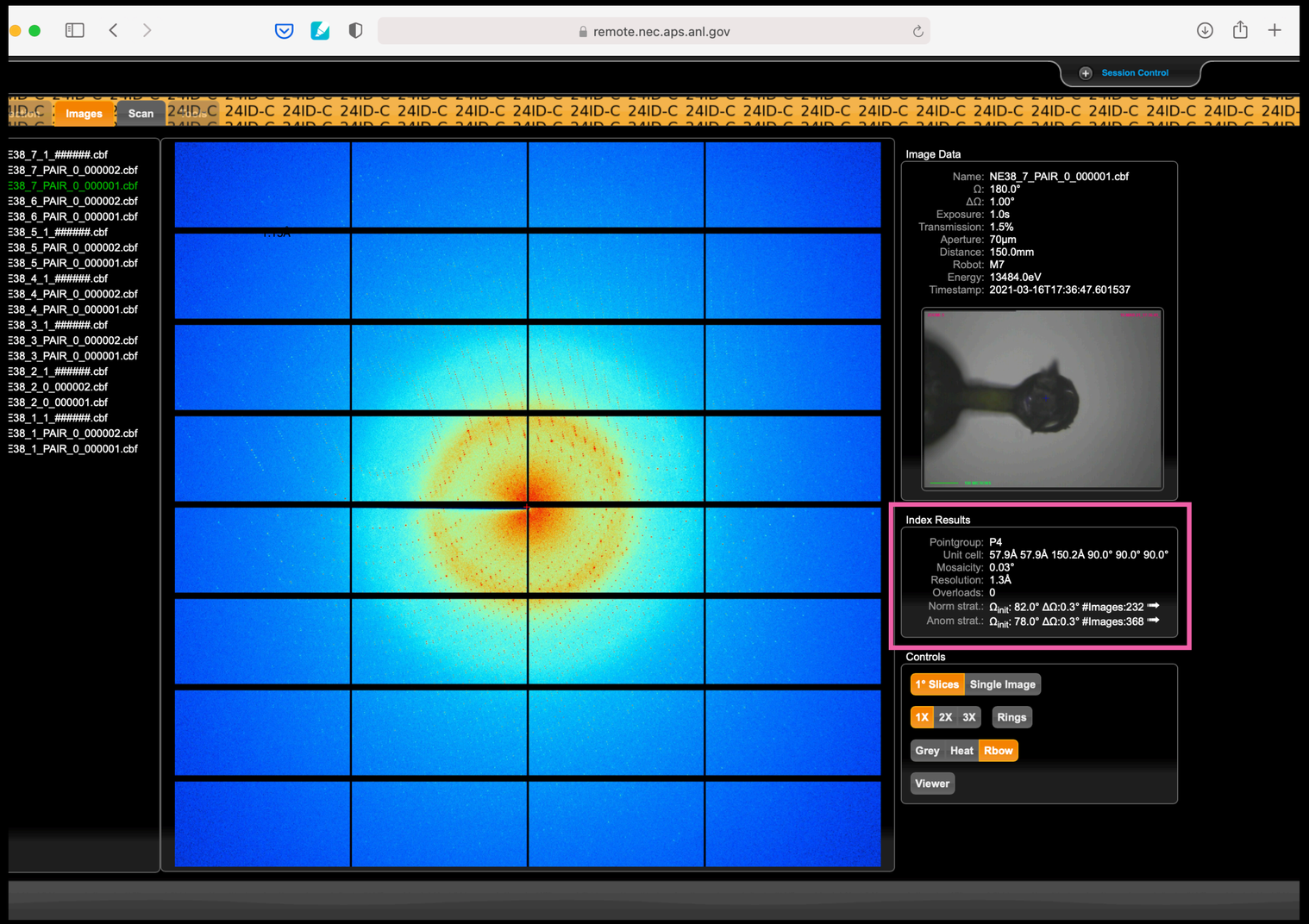
That’s All Great But…

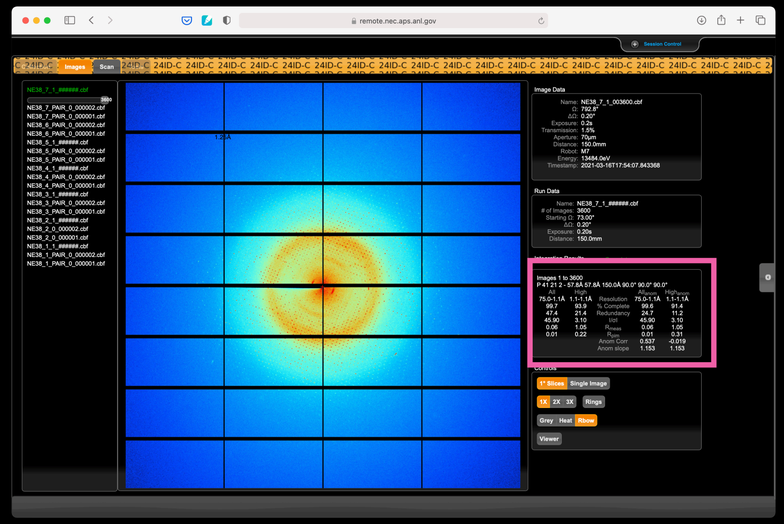
- Very difficult to add new features
- NE-CAT specific code throughout
- PHP UI was outdated

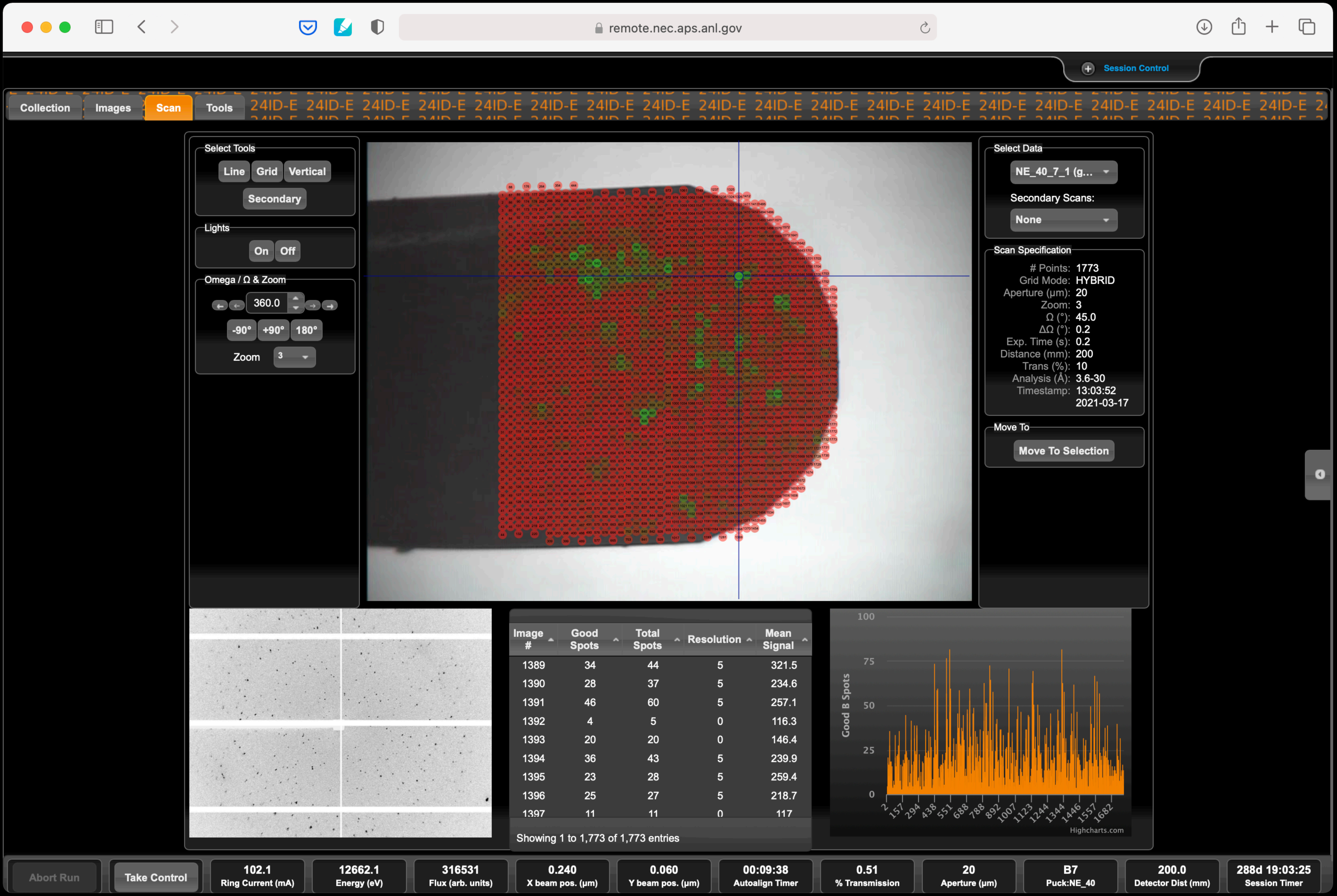
An Interruption to Our Regularly Scheduled Program


remote.nec.aps.anl.gov
Session Control
Collection
Images
Scan
Tools
24ID-E
Change Sample
Prev
Other
Next
Manual Centering
Zoom
3
3-Click Centering & Saving
3-Click
Abort
Save
Clear
Image Tools
Circle
Ruler
Point
Lights
Main
Aux
Data Collection
Snap
Run
Omega / Ω
360.0
-90°
+90°
180°
Abort Run
Take Control
102.0
Ring Current (mA)
12662.1
Energy (eV)
316566
Flux (arb. units)
0.690
X beam pos. (µm)
-0.610
Y beam pos. (µm)
00:09:25
Autoalign Timer
0.51
% Transmission
20
Aperture (µm)
B7
Sample
200.0
Detector Dist (mm)
288d 19:03:12
Session Timer

RAPD in Remote





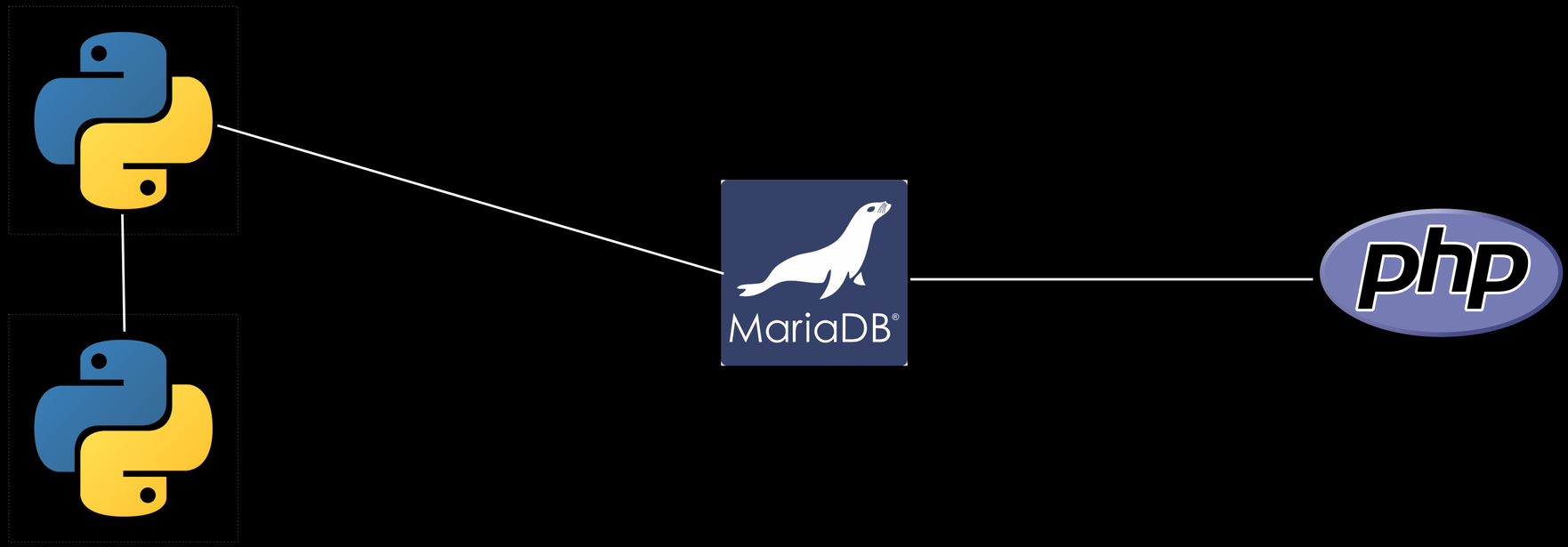

remote.nec.aps.anl.gov
Session Control
Collection
Images
Scan
Tools
24ID-E
Select Tools
Line
Grid
Vertical
Secondary
Lights
On
Off
Omega / Ω & Zoom
360.0
-90°
+90°
180°
Zoom
3
Select Data
NE_40_7_1 (g...
Secondary Scans:
None
Scan Specification
Points: 1773
Grid Mode: HYBRID
Aperture (μm): 20
Zoom: 3
Ω (°): 45.0
ΔΩ (°): 0.2
Exp. Time (s): 0.2
Distance (mm): 200
Trans (%): 10
Analysis (Å): 3.6-30
Timestamp: 13:03:52 2021-03-17
Move To
Move To Selection
Image # Good Spots Total Spots Resolution Mean Signal
1389 34 44 5 321.5
1390 28 37 5 234.6
1391 46 60 5 257.1
1392 4 5 0 116.3
1393 20 20 0 146.4
1394 36 43 5 239.9
1395 23 28 5 259.4
1396 25 27 5 218.7
1397 11 11 0 117
Showing 1 to 1,773 of 1,773 entries
Good B Spots
100
75
50
25
0
2
157
294
438
551
688
788
892
1007
1123
1244
1344
1446
1557
1682
Highcharts.com
Abort Run
Take Control
102.1
Ring Current (mA)
12662.1
Energy (eV)
316531
Flux (arb. units)
0.240
X beam pos. (μm)
0.060
Y beam pos. (μm)
00:09:38
Autoalign Timer
0.51
% Transmission
20
Aperture (μm)
B7
Puck:NE_40
200.0
Detector Dist (mm)
288d 19:03:25
Session Timer

RAPDv2 Development

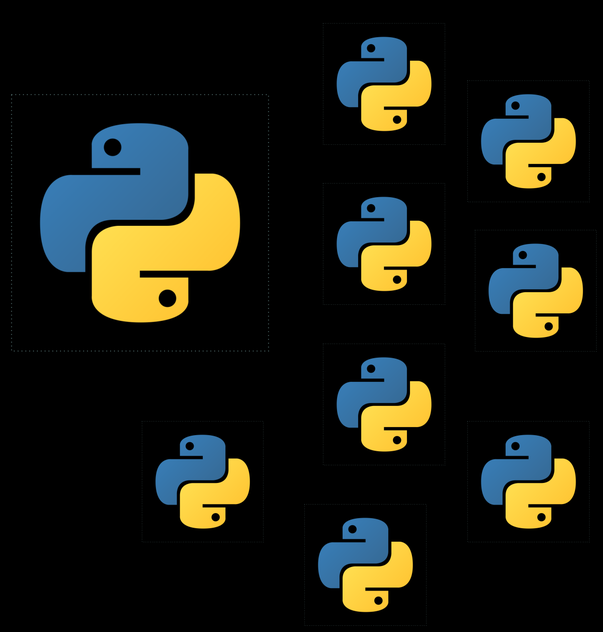
Our Goals For RAPDv2

- Take the best of RAPDv1 and throw out the rest
- Borrow knowledge from remote access program
- Flexible design
- Multiple modes of use
- Modern UI

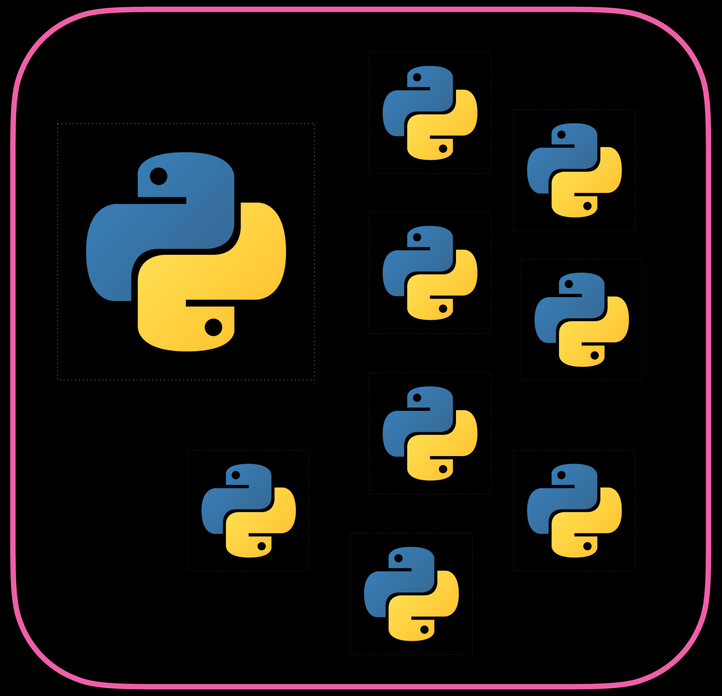
RAPDv1 Design



RAPDv2 Design

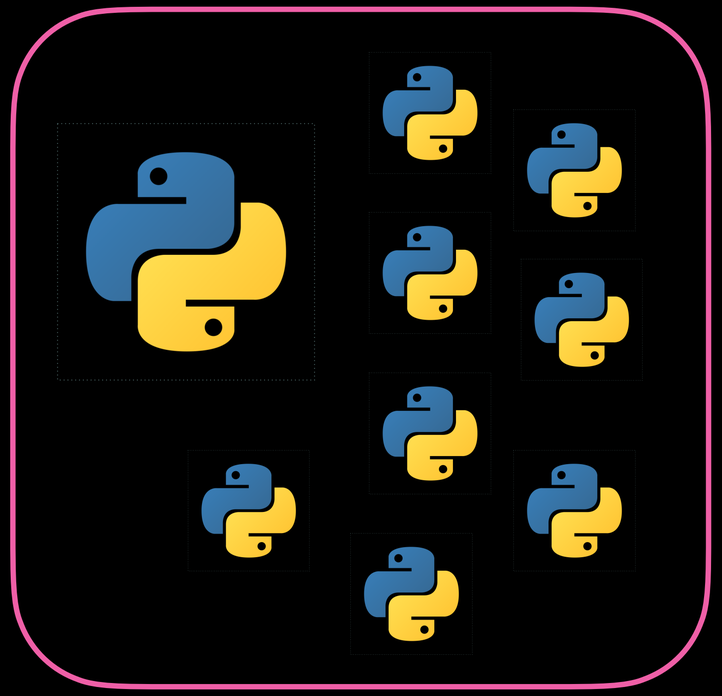


RAPDv2 Design



Plugin-based architecture
Python 2.7

RAPDv2 Design



Plugin-based architecture
Python 2.7

Core Python code is beamline-agnostic

Where does specificity come from?

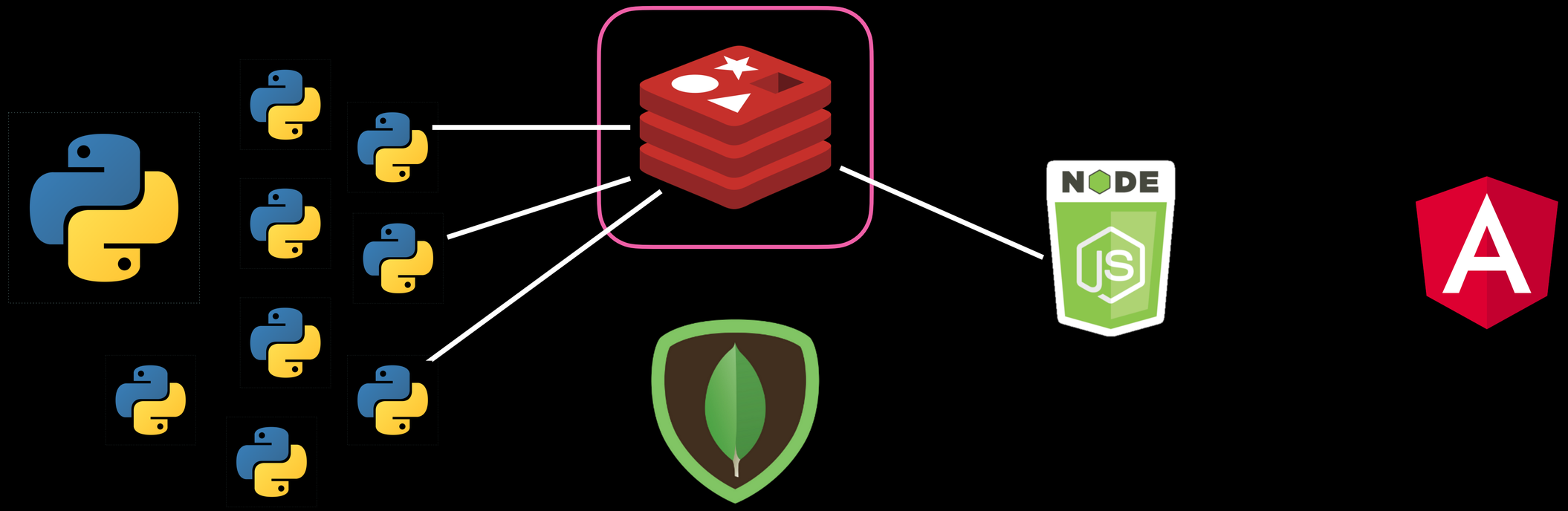
- Configuration file
- Site-specific plugins - detector, gatherers

Data processing plugins now run in CLI

What's a plugin?

- Connection to databases
- Data gatherers
- Connection to cluster
- Processing pipelines

RAPDv2 Design



Redis-based inter-process communication

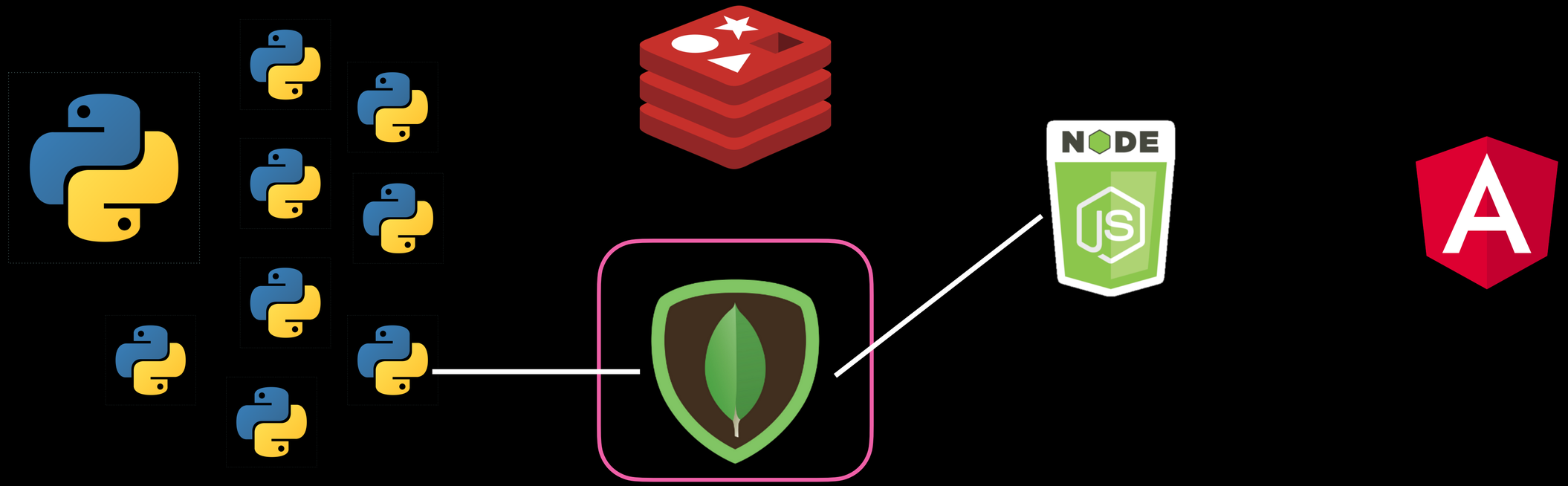
RAPDv2 Design

Redis-based inter-process communication

Redis is an indispensable tool for us.

- key-value store
- publication-subscribe interprocess communication
- redis.io

RAPDv2 Design



Document-based database
MongoDB

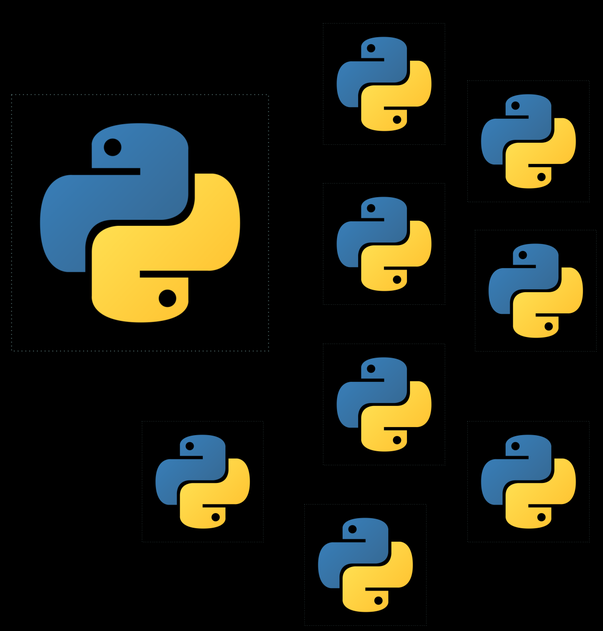
RAPDv2 Design

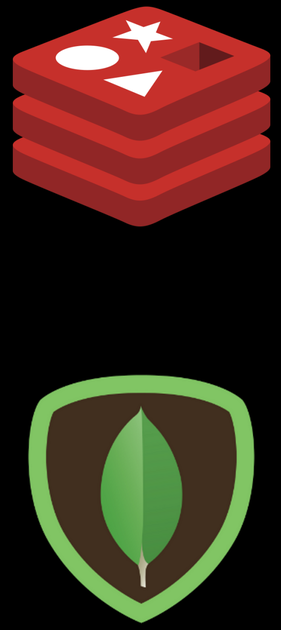
Document-based database
MongoDB

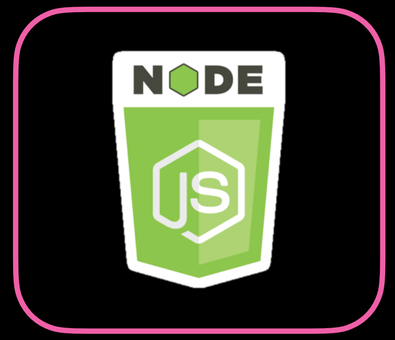
MongoDB much more flexible than SQL-based solutions

- Stores documents & files
- More flexibility
- Meshes well with modern UI design
- mongodb.com

RAPDv2 Design

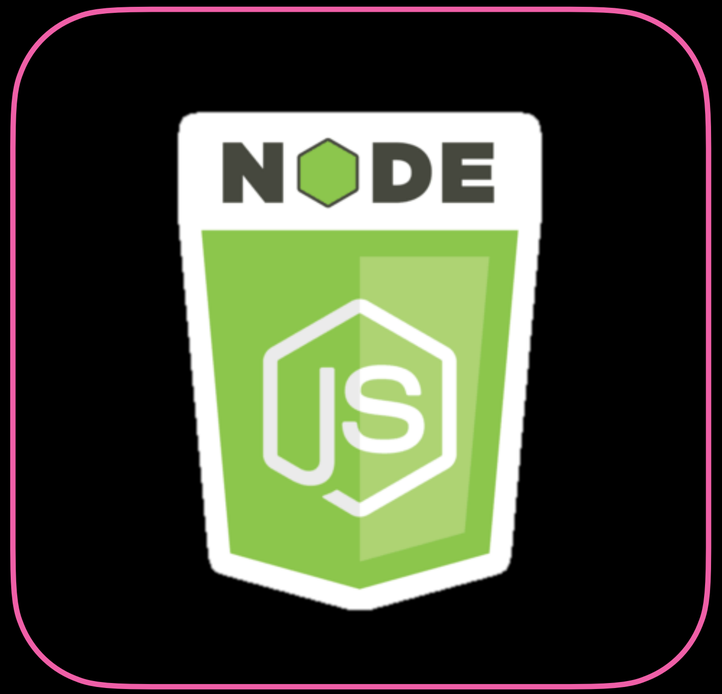






NodeJS
Application server

RAPDv2 Design

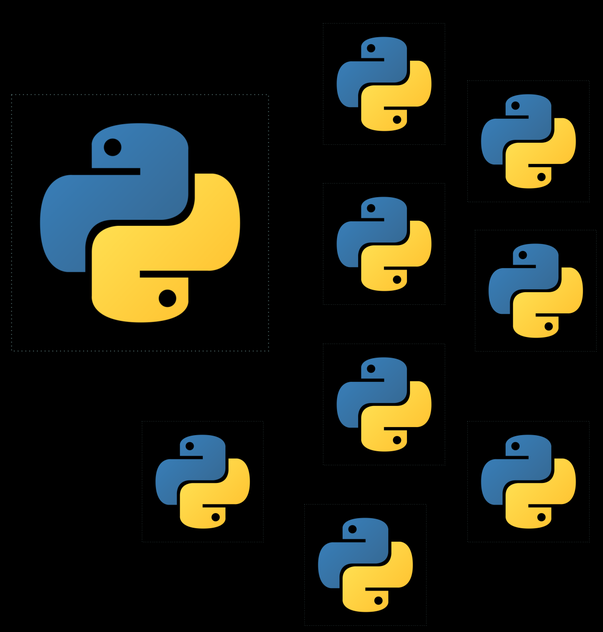


NodeJS
Application server

NodeJS offers a lot of power & flexibility

- Powerful, flexible & fast
- Tight integration with UI
- Provides REST & Websocket access to RAPDv2
- nodejs.org

RAPDv2 Design



Angular UI

RAPDv2 Design

Modern web development platform

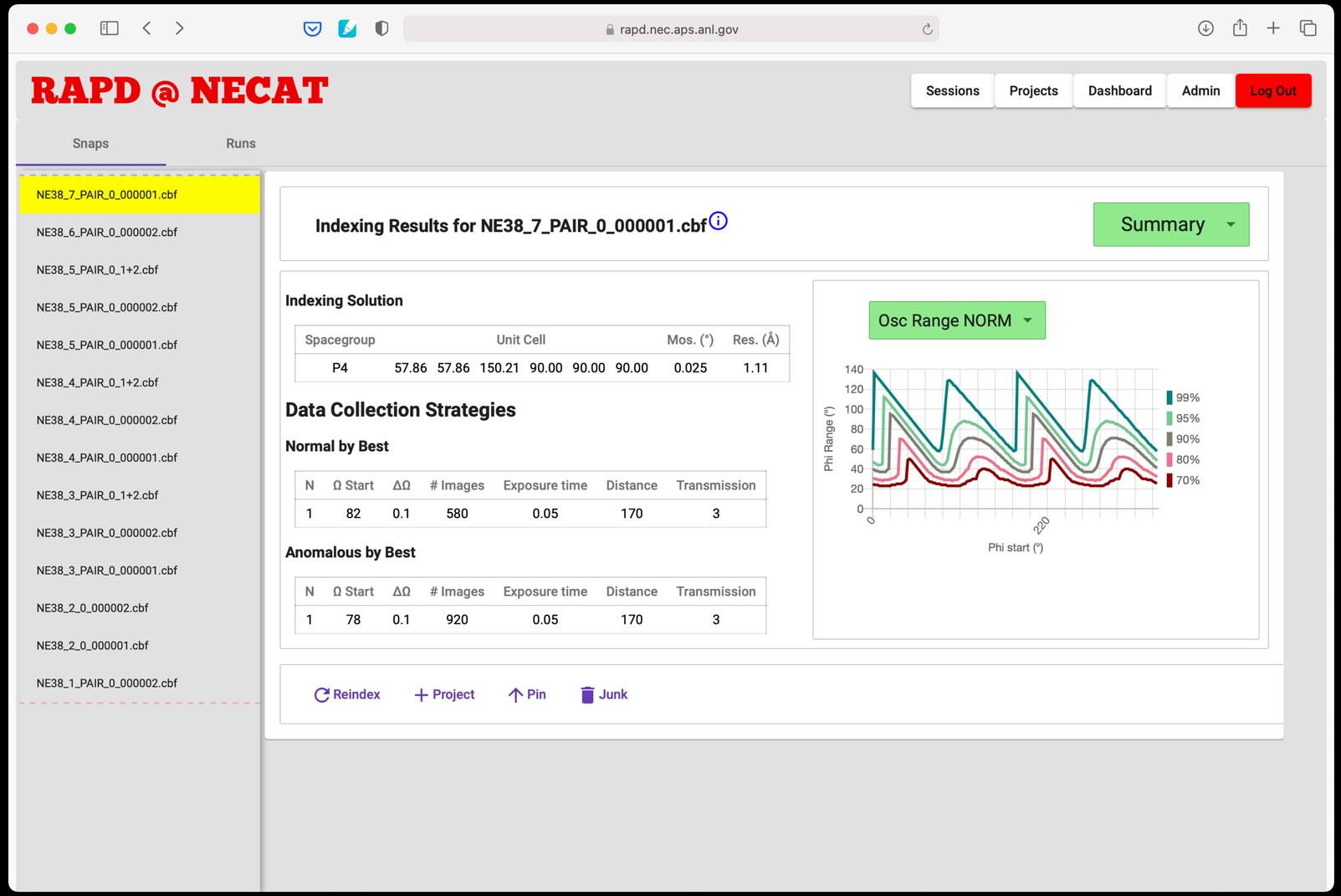
- Very powerful
- Steep learning curve
- Actively developed & large community
- angular.io

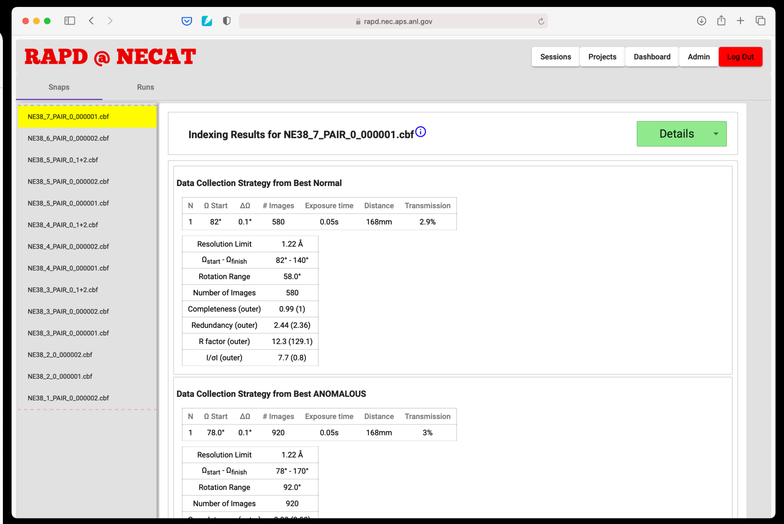
Angular UI

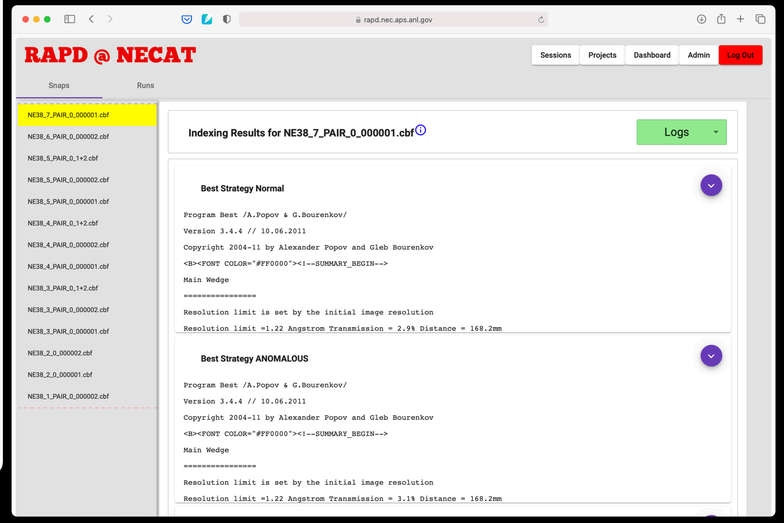
RAPDv2 Current Capabilities

- Indexing & Strategy – DISTL, Labelit, BEST, Mosflm
- Integration, Scaling & Analysis – XDS, Aimless, Pointless, Phaser, Xtriage
- Merging – Aimless, Pointless
- SAD – SHELX, Phenix
- MR – Phaser

RAPDv2 Index







RAPDv2 Integrate

rapd.nec.aps.anl.gov

RAPD @ NECAT

Sessions | Projects | Dashboard | Admin | Log Out

Snaps | Runs

- NE38_7_1_[1-3600].cbf
- NE38_5_1_[1-3600].cbf
- NE38_4_1_[1-3600].cbf
- NE38_3_1_[1-3600].cbf
- NE38_2_1_[1-3600].cbf
- NE38_1_1_[1-1800].cbf

Integration Results for NE38_7_1_[1-3600].cbf

Summary

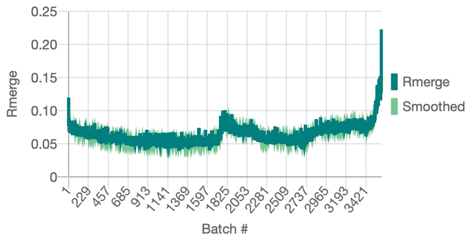
Indexing Solution

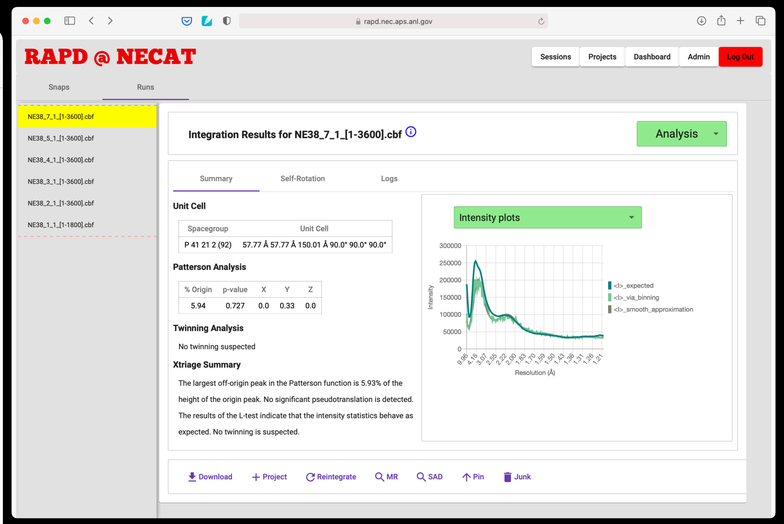
Spacegroup	Unit Cell	Mosaicity	ISA
P 41 21 2	57.77 Å 57.77 Å 150.01 Å 90.0° 90.0° 90.0°	0.06°	28.83

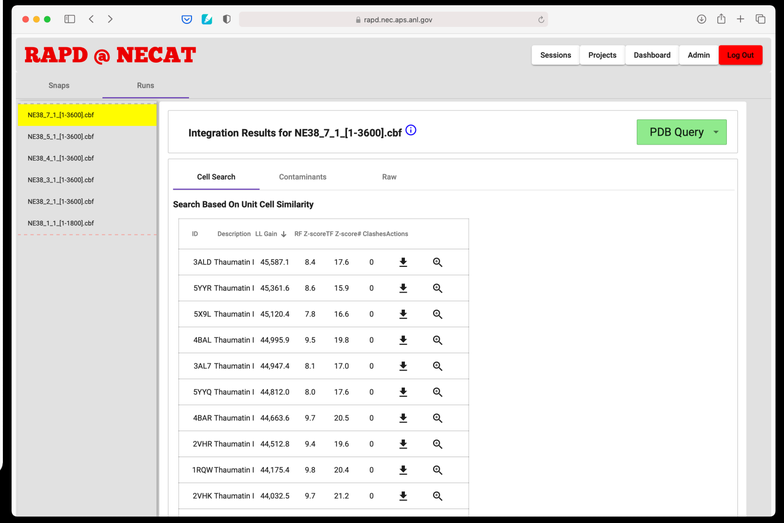
Table 1

	Overall	Inner Shell	Outer Shell
Res Limits (Å)	75.01-1.05	75.01-5.76	1.07-1.05
Completeness (%)	99.3	100	87.6
Multiplicity	45.8	40.4	17.6
I/σ(I)	36.5	149.4	0.4
$CC_{1/2}$	1	1	0.161
R_{merge}	0.064	0.028	5.768
R_{merge} Anomalous	0.063	0.027	5.661
R_{meas}	0.064	0.029	5.936
R_{meas} Anomalous	0.064	0.028	5.986
R_{pim}	0.009	0.004	1.355
R_{pim} Anomalous	0.013	0.005	1.894
Anomalous Completeness	99.1	100	83.9
Anomalous Multiplicity	23.9	26	9.4
Anomalous Correlation	0.356	0.622	0.104
Anomalous Slope	1.084		

Rmerge vs Batch







Our Goals For RAPDv2

Take the best of RAPDv1 and throw out the rest

- Much of the data processing code was adapted
- Some UI elements that we liked were kept

Our Goals For RAPDv2

Borrow knowledge from remote access program

- Use of Redis & MongoDB
- NodeJS as application server

Our Goals For RAPDv2

Flexible design

- Core programs are not site-specific
- Config file determines variables and plugins that define site

Our Goals For RAPDv2

Multiple modes of use

- CLI for data processing & development
- REST interface
- Websocket interface

RAPDv2 Status

Now

- RAPDv2 running as secondary
- Final steps of UI implementation for SAD & MR

Mid 2021

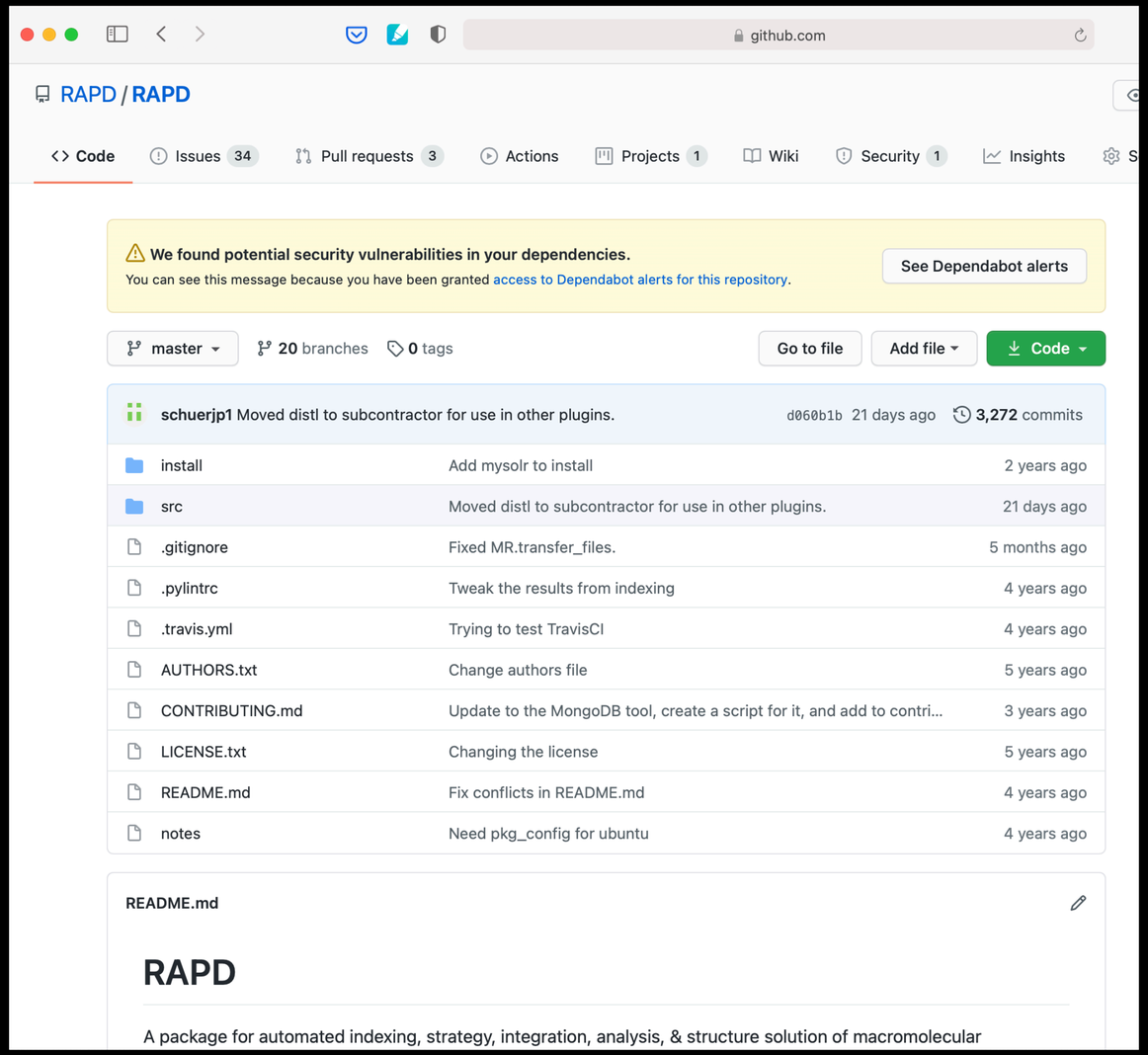
- Replace RAPDv1 as primary
- Built-out operations control

Late 2021

- Merging and Cluster Merging tools

Can I Use RAPDv2?

- If you really like the *bleeding* edge, have at it.
- Waiting until NE-CAT is using as primary (mid-2021) is recommended.



Acknowledgements

Contributors
Jon Schuermann
David Neau
Kay Perry
Jim Withrow

NE-CAT Support Staff
Surajit Banerjee
Ali Kaya
Igor Kourinov
N. Sukumar

Support
NIGMS P30 GM124165