



Justyna Aleksandra Wojdyla :: Macromolecular Crystallography Group :: Paul Scherrer Institut

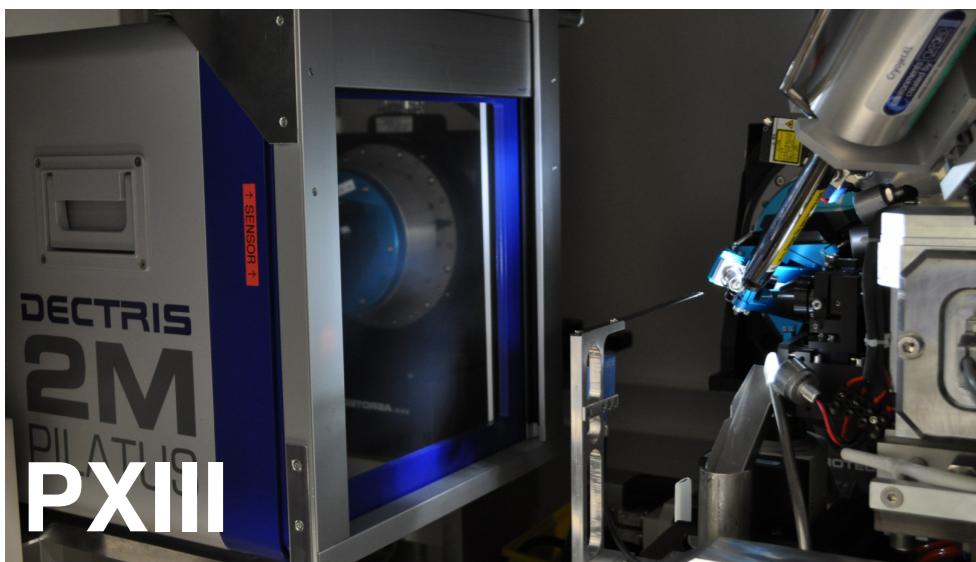
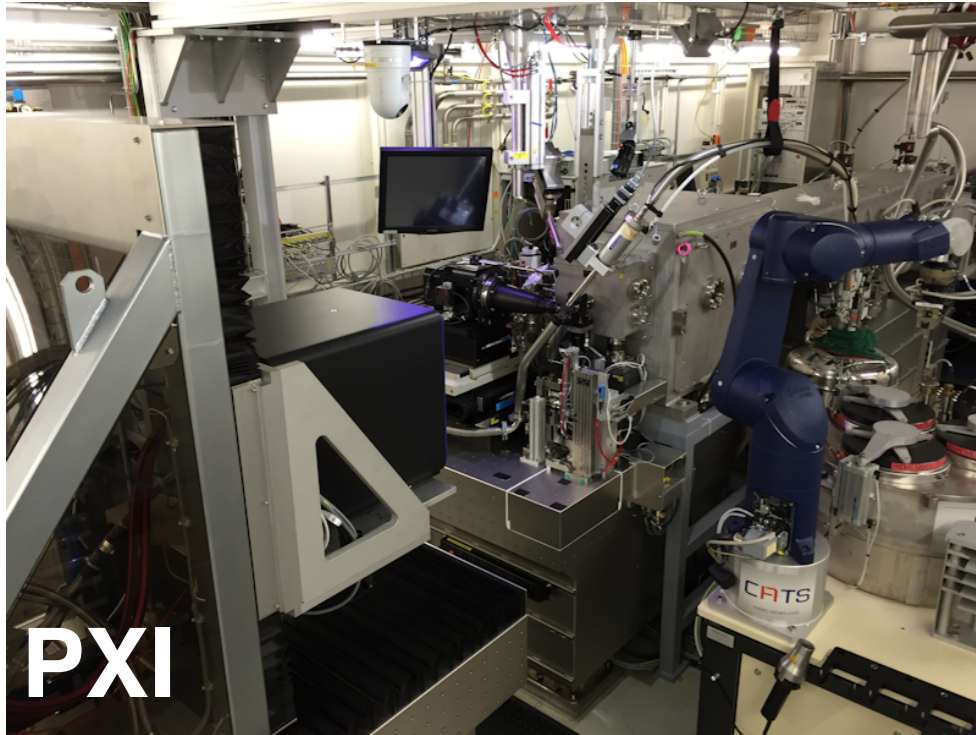
Data acquisition and analysis software at the Swiss Light Source macromolecular crystallography beamlines

NOBUGS, 18-10-2016

Outline

- Introduction to SLS MX beamlines
- DA+ data acquisition software
- Automatic data processing
- Database & tracker
- Eiger 16M

SLS MX beamlines



	PXI (X06SA)	PXII (X10SA)	PXIII (X06DA)
Wavelength range [Å]	0.7 - 2.2	0.62 - 2.07	0.71 - 2.07
Source	in-vacuum undulator	in-vacuum undulator	bending magnet
Spectral range	5.7 - 17.5	6.0 - 20.0	6.0 - 17.5
Flux at 12.4 keV [ph/s]	$> 2 \times 10^{12}$	$> 2 \times 10^{12}$	5×10^{11}
Focused spot size h x v [μm^2]	5 x 5 (2 x 1) -> 100 x 100	73 x 16	80 x 45
Detector	EIGER 16M	PILATUS 6M	PILATUS 2M-F
Frame rates [Hz]	133 750 (4M ROI)	25	25
Goniometer	single-axis	single-axis	multi-axis PRIGo

PXI

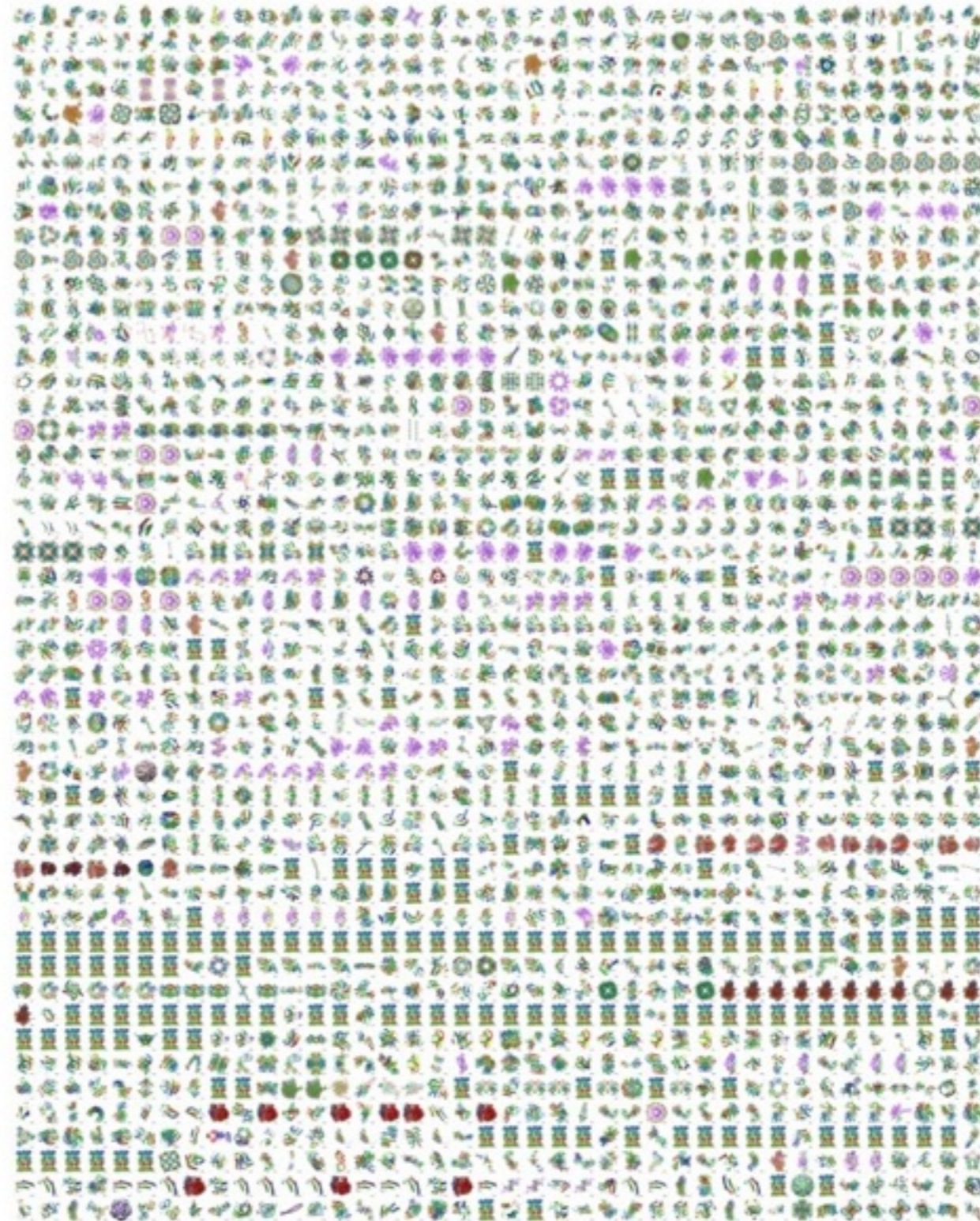


31 August 2016

2000 Structures at Xo6SA

<http://www.psi.ch/sls>

#1 in Europe

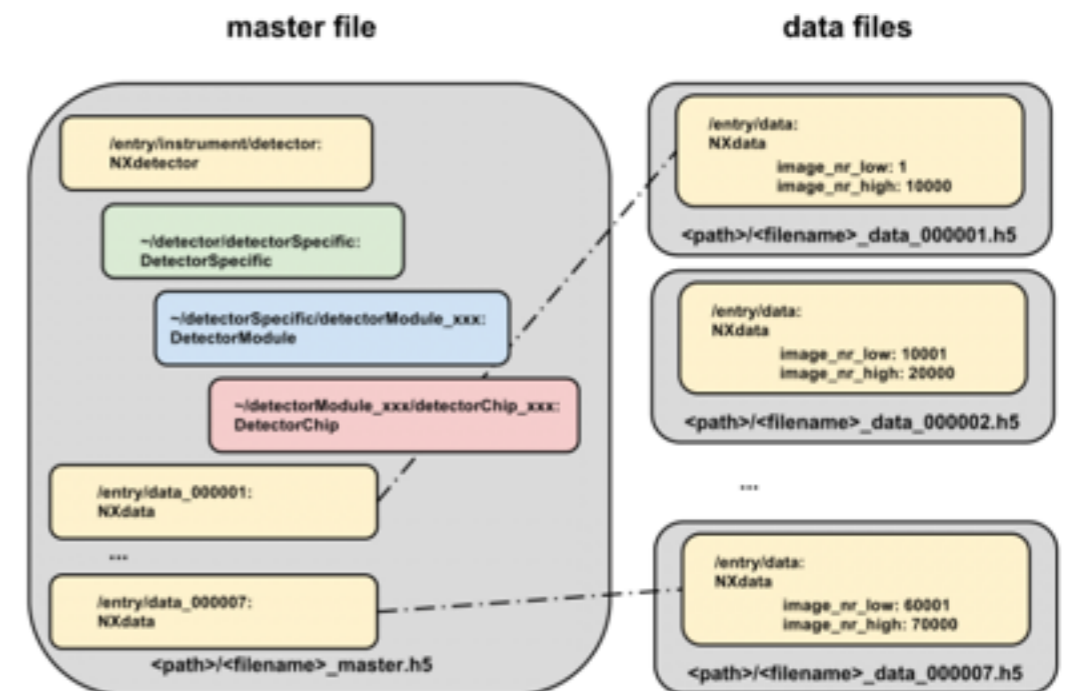


Data collection at PX beamline

- 8 hours shift
- ~75% of time screening for well diffracting crystals (up to 170 crystals)
- up to 10 datasets per hour

```
##PCRF: VERSION 1.5, CBFLib v0.7.8 - PILATUS detectors
data_dataset_1_00001
_array_data.header_convention "PILATUS_1.2"
_array_data.header_contents
{
# Detector: PILATUS 2MP, S/N 24-0109-F
# 2013-05-19T11:29:35.073
# Pixel_size 172e-6 m x 172e-6 m
# Silicon sensor, thickness 0.000450 m
# Exposure_time 0.0977200 s
# Exposure_period 0.1000000 s
# Tau = 383.8e-09 s
# Count_cutoff 127302 counts
# Threshold_setting: 3085 eV
# Gain_setting: high gain (vrf = -0.150)
# N_excluded_pixels = 289
# Excluded_pixels: badpix_mask.tif
# Flat_field: FP_p2m0109_E6170_T3085_vrf_m0p15.tif
# Trim_file: p2m0109_E6170_T3085_vrf_m0p15.bin
# Image_path: /ramdisk/el0007/20130519/6D/LXE1/2/omegal/
# Wavelength 2.06439 A
# Detector_distance 0.12000 m
# Detector_Voffset 0.00000 m
# Beam_xy (492.00, 853.00) pixels
# Flux 13898727302.3
# Filter_transmission 1.0000
# Start_angle 0.0000 deg.
# Angle_increment 0.1000 deg.
# Phi 0.0000 deg.
# Chi 0.0000 deg.
# Oscillation_axis OMEGA
;
_array_data.data
{
--CIF-BINARY-FORMAT-SECTION--
Content-Type: application/octet-stream;
  conversions="x-CBF_BYTE_OFFSET"
Content-Transfer-Encoding: BINARY
X-Binary-Size: 2474855
X-Binary-ID: 1
X-Binary-Element-Type: "signed 32-bit integer"
X-Binary-Element-Byte-Order: LITTLE_ENDIAN
Content-MD5: 486naj8InVEyXCooFRtVNO==
X-Binary-Number-of-Elements: 2474525
X-Binary-Size-Fastest-Dimension: 1475
X-Binary-Size-Second-Dimension: 1679
X-Binary-Size-Padding: 4995
}
```

- data format
 - cbf (Pilatus 6M and 2M-F) - 1 image per frame (few MB)
 - hdf5 (Eiger 16M)
 - 350MB master file per dataset
 - 300-800MB per one data file (100 frames)



https://www.dectris.com/EIGER_X_Features.html#main_head_navigation

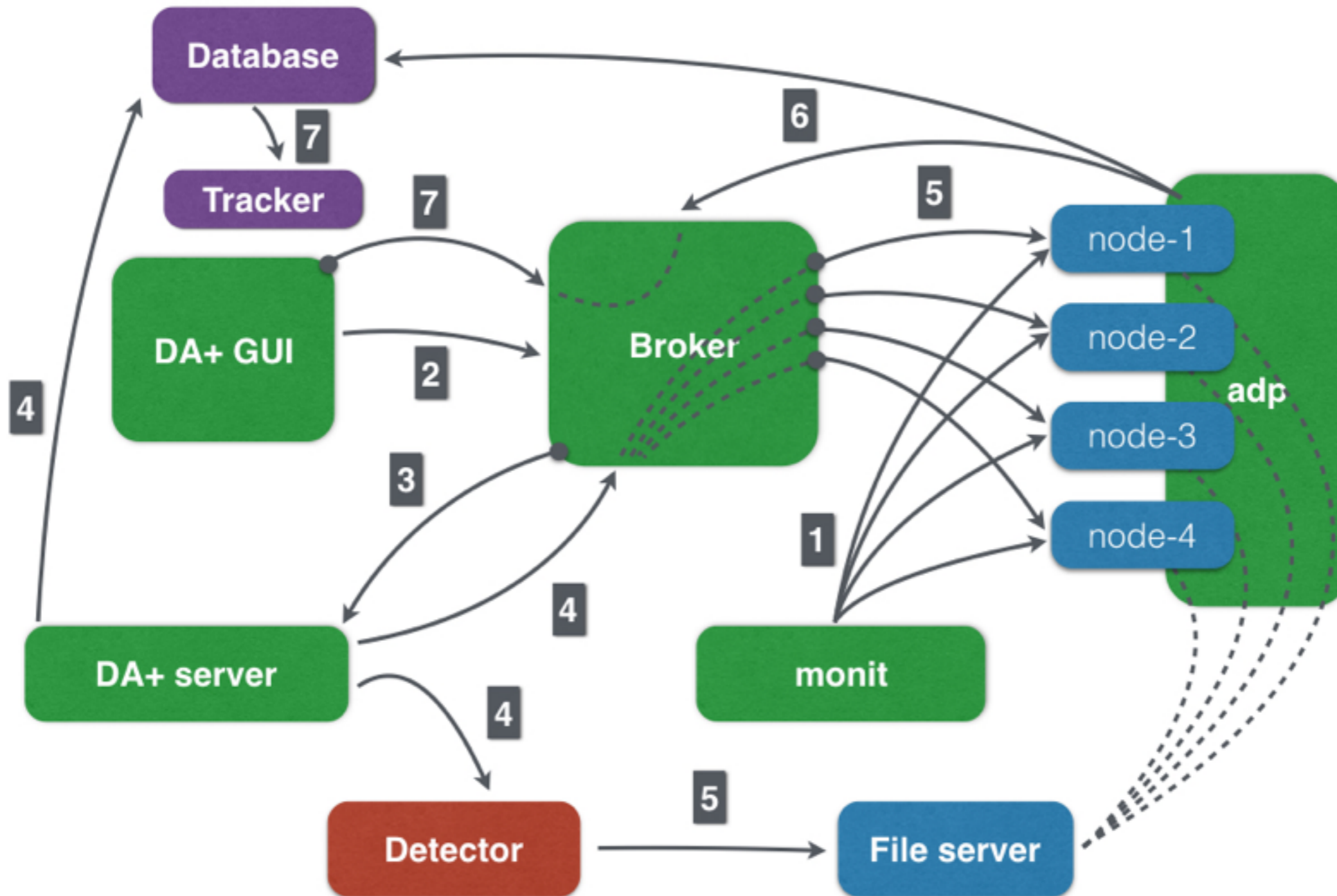
DAQ from users' perspective

The image displays a complex DAQ (Data Acquisition) interface for a synchrotron experiment, viewed from a user's perspective. The interface is divided into several main sections:

- Terminal (Top Left):** A green terminal window showing system logs and status messages.
- Camera View (Middle Left):** A live video feed of a sample mounted on a goniometer. The sample is a small, dark, irregularly shaped object. A scale bar indicates 50 μm . The camera is labeled "DA+".
- Control Panel (Middle Left):** A sidebar with various controls including "Sample Exchange", "Sample Alignment", "Data Collection", "Screening", and "Collection". It includes fields for folder names, run numbers, and exposure times.
- Diffraction Pattern (Top Right):** A large, circular diffraction pattern showing a grid of red spots on a dark background. The pattern is centered on a bright spot, indicating a well-aligned sample.
- Data Analysis Dashboard (Bottom Right):** A web-based interface showing experimental parameters, crystal parameters, and a list of collected data points. The dashboard includes sections for "Experiment parameters", "Crystal parameters", "Twining", and "Pointless: Space group uniquely determined: P 4 2 1 2 0".

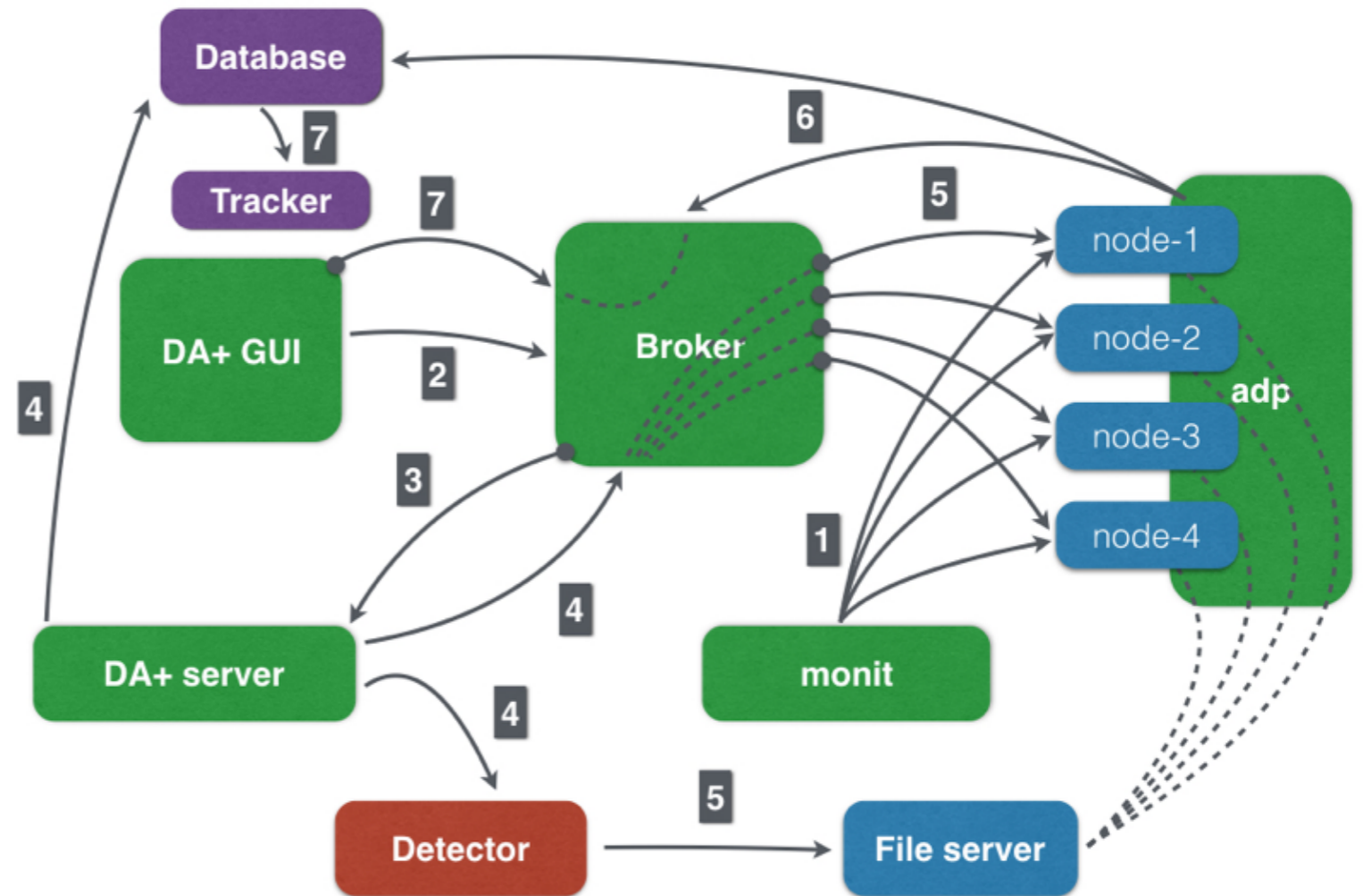
At the bottom of the screen, there is a status bar with the text "Adapted for GFA".

Overall infrastructure



Software infrastructure

- DA+ data acquisition software
- distributed services and components
 - DA+ GUI
 - DA+ server
 - Broker (Apache ActiveMQ)
 - Automatic data processing (adp)
 - Database
 - Tracker
- Python and Java
- messaging and streaming technologies



Hardware infrastructure

- **Computing**

- online clusters - 4 or 2 nodes Dual Xeon E5-2697v2 (2.70 GHz), 24 cores, 256GB ram, Scientific Linux 6.4
 - data processing
 - spot finding (raster)
- raster cluster - 3 nodes Dual Xeon E5-2697v2, 24 cores, 256GB ram, Scientific Linux 6.4
 - spot finding (raster)
- offline cluster - 16 nodes Dual Xeon E5-2690v3 (2.60 GHz), 256GB ram, Scientific Linux 7.0
 - MX software (available via nomachine)

node-1

node-2

node-3

node-4

- **Storage**

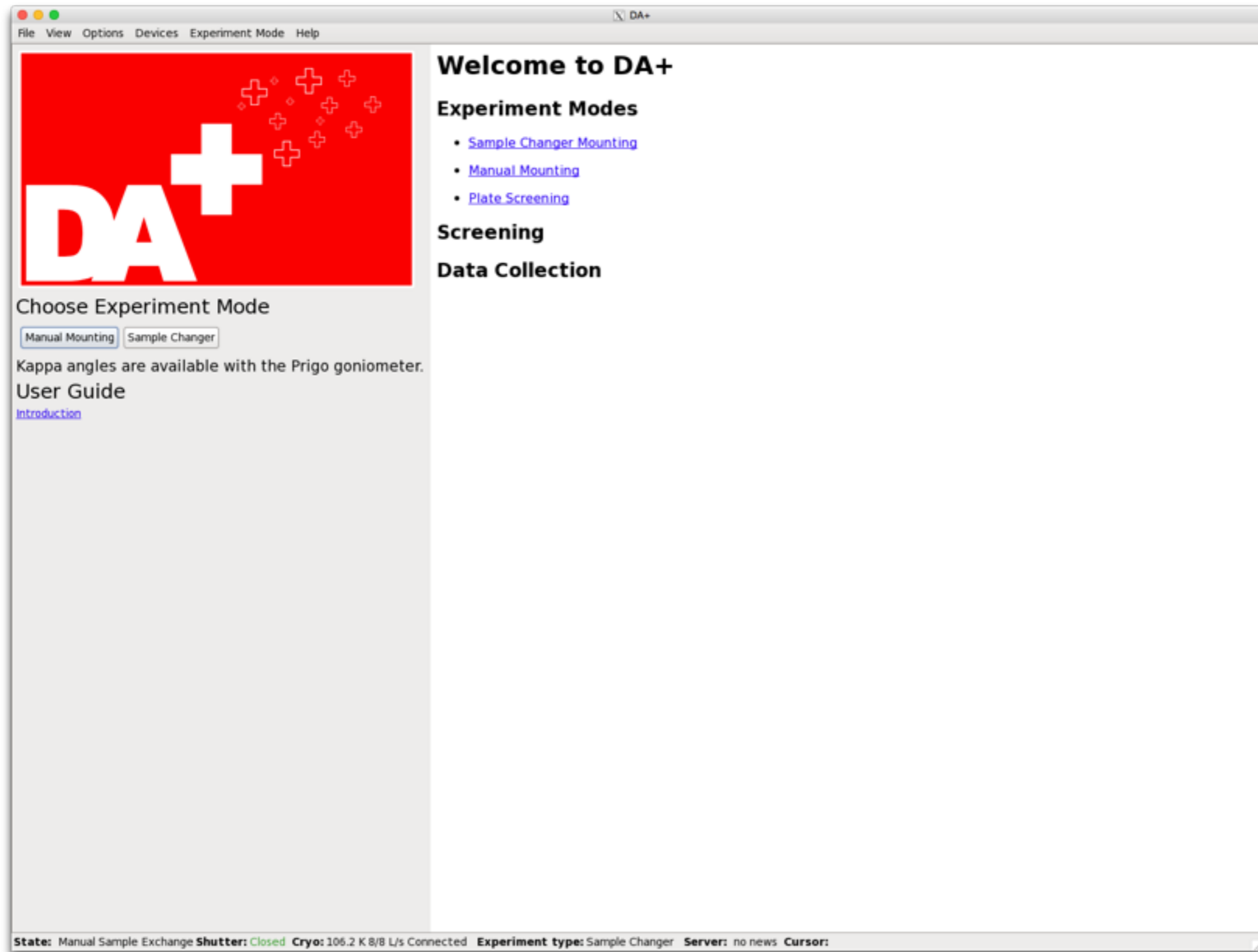
- IBM GPFS version 4.1
- 1.2 PB total
- 175 TB for all MX beamlines

File server

- Python-based
- initialises all necessary software and hardware (EPICS)
- communicates with GUI and adp (via broker)
- communicates with database (via REST API)

```
2016-10-12 10:07:21.458 - mx - INFO - ... init worker manager
http://mx-app-2.psi.ch:80/worker/analysis/x06da/e14965
2016-10-12 10:07:21.522 - mx - INFO - ... init data transfer manager
2016-10-12 10:07:21.685 - mx - INFO - Initializing pilatus streamer
2016-10-12 10:07:22.050 - mx - INFO - ... init delay generator
2016-10-12 10:07:23.691 - mx - INFO - ... init shutter
2016-10-12 10:07:23.749 - mx - INFO - ... init hub
2016-10-12 10:07:23.749 - mx - INFO - ... init pilatus
2016-10-12 10:07:23.886 - mx - INFO - ... init abr
2016-10-12 10:07:23.911 - mx - INFO - ... init albula
2016-10-12 10:07:23.912 - mx - INFO - ... init idl
2016-10-12 10:07:24.072 - mx - INFO - ... init transmission
2016-10-12 10:07:24.154 - mx - INFO - ... init collimator
2016-10-12 10:07:24.154 - mx - IEBUG - Initializing motor: X06DA-ES-MS:TRZ1
2016-10-12 10:07:24.182 - mx - INFO - ... init optics helper
2016-10-12 10:07:24.682 - mx - INFO - ... init positioner Prigo
2016-10-12 10:07:26.287 - mx - INFO - ... init motors: dtz, dty, bsz
2016-10-12 10:07:26.287 - mx - IEBUG - Initializing motor: X06DA-ES-DET:TRZ1
2016-10-12 10:07:26.320 - mx - IEBUG - Initializing motor: X06DA-ES-DET:TRY1
2016-10-12 10:07:26.374 - mx - IEBUG - Initializing motor: X06DA-ES-BS:TRZ1
2016-10-12 10:07:26.428 - mx - INFO - ... init channels: current, flux
2016-10-12 10:07:26.483 - mx - INFO - ... init configuration
2016-10-12 10:07:26.492 - mx - INFO - ... attaching to message queue: mx-broker.psi.ch:61613/queue/daqserver/x06da/queue
2016-10-12 10:07:26.993 - mx - INFO - Ready for next scan request.

:-*
```

DA+ GUI

DA+ GUI

The screenshot displays the DA+ GUI interface, which is used for controlling a synchrotron beamline. The main window is titled "Data Collection MAD" and features a central camera view showing a sample mounted on a goniometer. The sample is a small, light-colored, irregularly shaped object, possibly a crystal, mounted on a dark, reflective holder. A yellow box highlights a specific area on the sample, with dimensions of 90 μm and 50 μm indicated.

The interface is divided into several panels:

- Alignment Control:** Located on the right side, it includes controls for "Wash Sample", "Camera Zoom" (1x, 2x, 5x, 8x), "Illumination", "Back Reflector" (checked), "Front Light", "Omega" (rotation) with buttons for -90.000, 0°, -10°, +10°, -45°, +45°, -90°, +90°, and a 180° rotation button. It also has "Sample Movement" controls and "Prigo Angles" (CHI and PHI) set to 0.
- Data Collection:** Located on the left side, it includes fields for "Folder" (20161012/crystal1/), "Prefix" (crystal), "Run" (1), and "First Index" (1). It also has "High Resolution" (2.644 Å), "Detector Distance" (300 mm), "Start Angle" (0°), and "Beam Transmission" (1.000). There are tabs for "Screening", "Collection", and "Advanced", with "Advanced" selected. It includes "Oscillation Angle" (0.5°), "Exposure Time" (0.5 s), and a "Schedule" dropdown set to "2 every 90°".
- Overall Status / Log Viewer:** Located at the bottom, it shows the system's status. It indicates "Ready to Go: yes" and provides details about the machine/beamline (Ring Current: 401.0 mA, Energy: 12398.4 eV, Wavelength: 1.000 Å, Flux: 3.323e+11 ph/s, Cryojet Status: 103.7 K Flows: 9/9 L/s, Status: Connected) and a radiation damage estimate (average, conservative). It also shows a "Beamline Token" and a log of events, including "12.10.16 07:00 Scheduled start of Beamline devel." and "11.10.16 13:00 Beamline Development, 400mA Top-up".

At the bottom left, there are buttons for "Start Data Acquisition" and "Abort Data Acquisition". The status bar at the very bottom reads: "State: Sample Alignment Shutter: Closed Cryo: 103.7 K 9/9 L/s Connected Experiment type: Manual Exchange Server: no news Cursor: 1165, 1147".

DA+ GUI MAD tab

Choose element



Change energy



Transmission search



Take spectrum



Scan edge



The screenshot displays the DA+ GUI MAD tab interface. On the left, there are five control panels corresponding to the numbered arrows:

- Select element & edge:** Element: Ta, ROI: 7814.5 to 8468.2
- Set energy:** Energy: 9921.0 eV
- Fluorescence detector:** Current position: Park
- Set beam transmission:** Transmission search: Success, Transmission: 0.05071
- X-ray fluorescence measurement:** Take spectrum: Success, ROI counts: 1472 counts, Deadtime: 0.00%
- Element edge scan:** Scan Ta(L-III) edge: Success

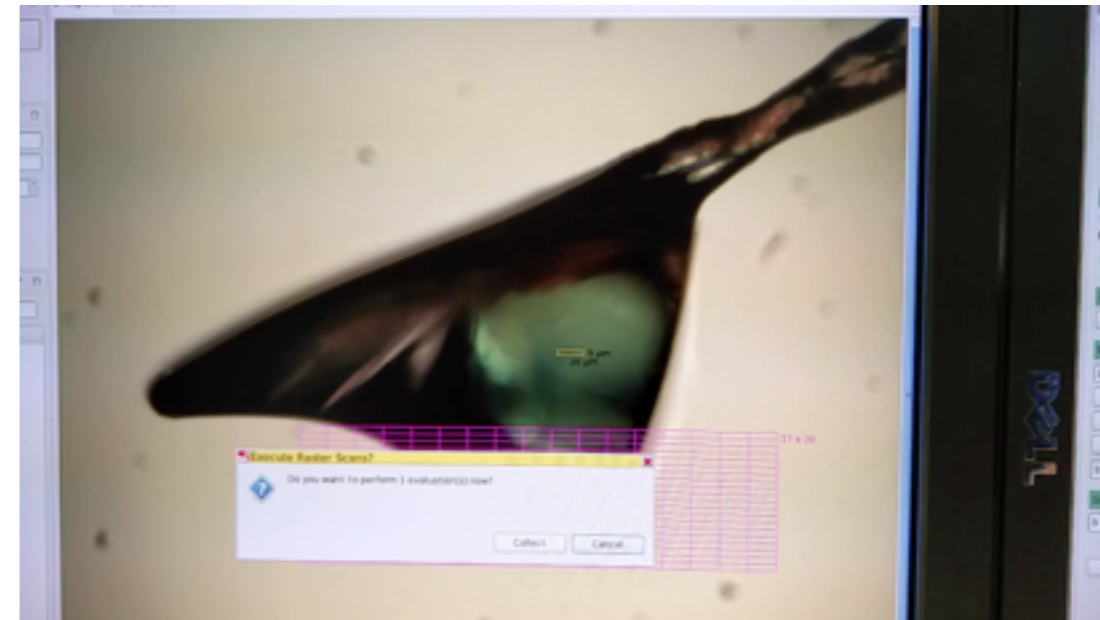
On the right, there are two plots:

- MAD Spectrum:** X-ray fluorescence measurement for Ta(L-III) at Mono=9921.0eV. The plot shows Counts vs Energy [eV] with a prominent peak at approximately 9881 eV.
- MAD Scan:** Element edge scan for Ta(L-III). The plot shows Counts [ROI] vs Energy [eV] with an inflection at 9877.8 eV and a peak at 9881.1 eV. The plot includes data points for counts [ROI], f', and f''.

At the bottom, there is a **MAD Messages** window showing log entries for writing edge scan data.

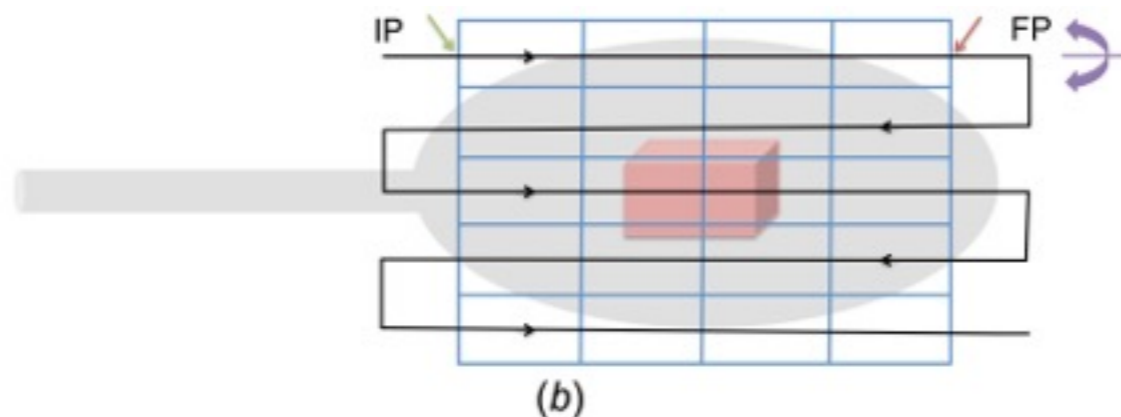
State: Xrf Collection Shutter: Closed Cryo: 105.3 K 8/8 L/s Connected Experiment type: Manual Exchange Server: no news Cursor:

- D3 diffractometer with 3 linear translation stages (GMX, GMY and GMZ), air-bearing spindle for the ω rotation & sample head with two linear stages (STY & STZ)
- GMX, GMY, GMZ and ω are driven by Aerotech 3200 controller, while STY & STZ *via* standard motor drivers (EPICS)
- Continuous grid scan without oscillation
 - single row, multiple row and single column supported
- Continuous grid scan with oscillation
 - single row and multiple row supported



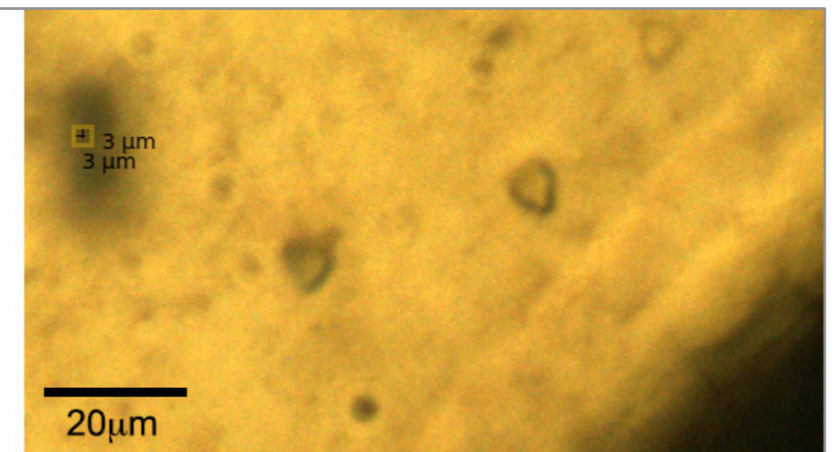
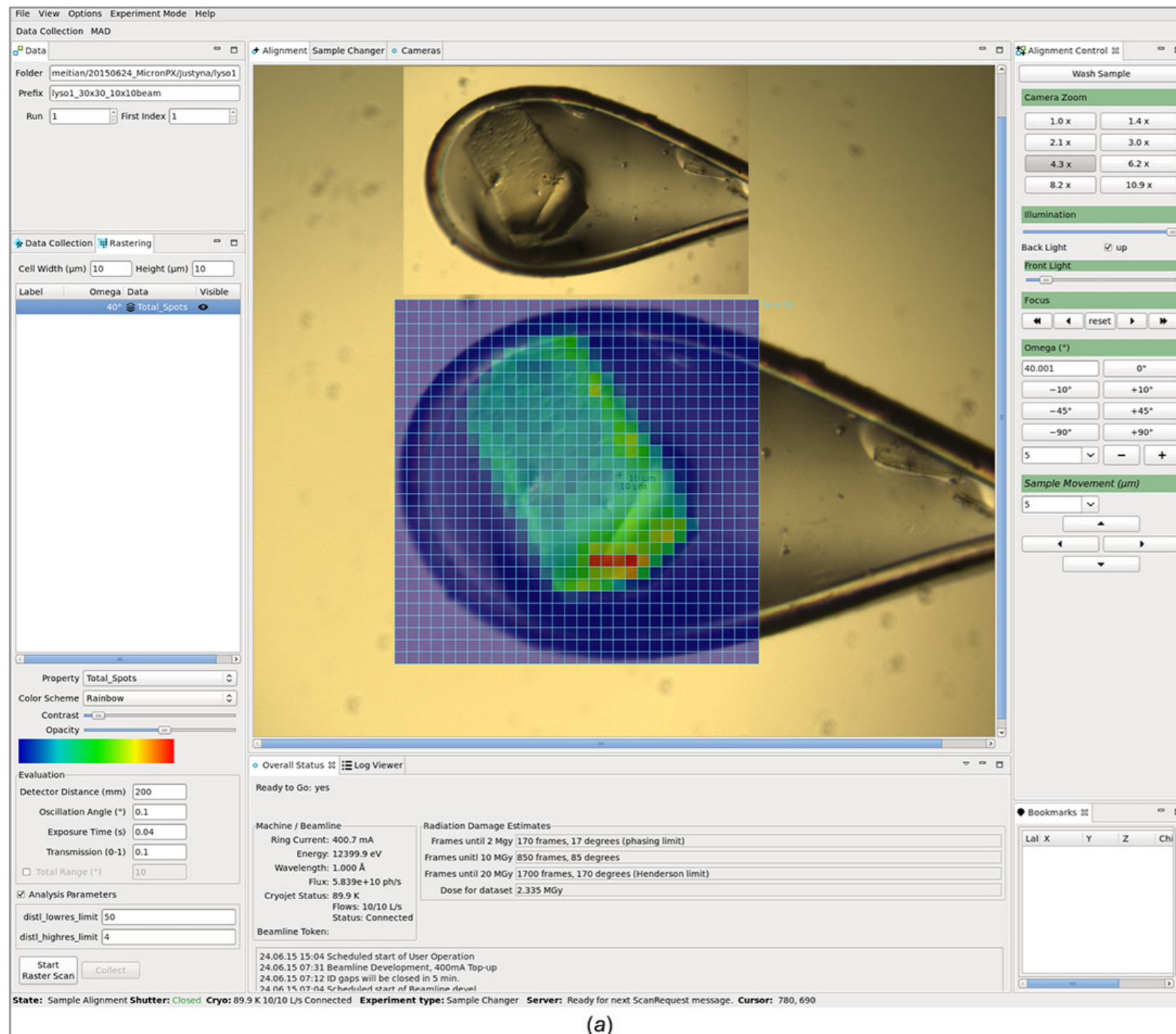
(a)

Wojdyla et al, Journal of Applied Crystallography, 2016

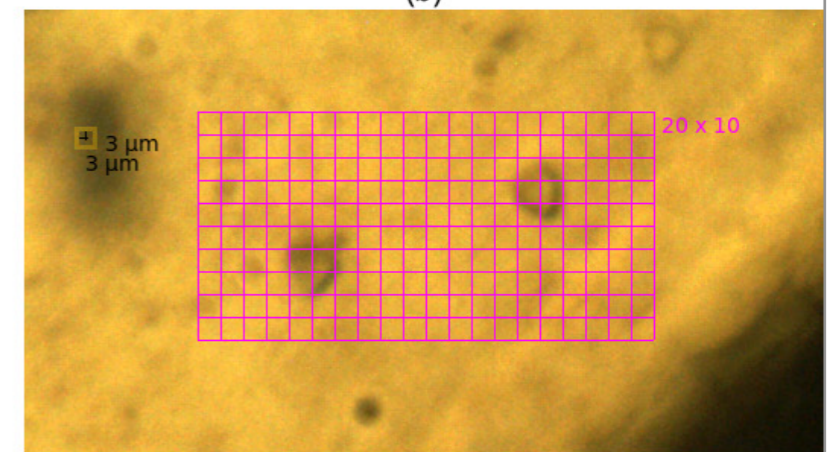


(b)

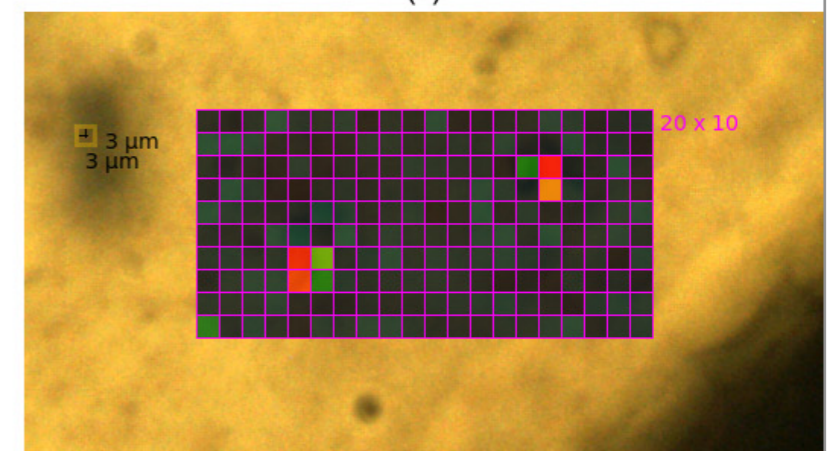
- Results of analysis with *labelit.distl* or Cheetah reported back and displayed in the GUI



(b)



(c)



(d)

- 200x100µm insulin crystal in nylon loop measured at 100K with PILATUS 6M-F detector, 10x10µm beam and 0.04s exposure -> 900 images collected in 36s

- localisation of membrane protein PepT_{St} microcrystals in the LCP between silicon nitride windows with 100K with PILATUS 6M-F detector (3x3µm beam and 0.1s exposure)

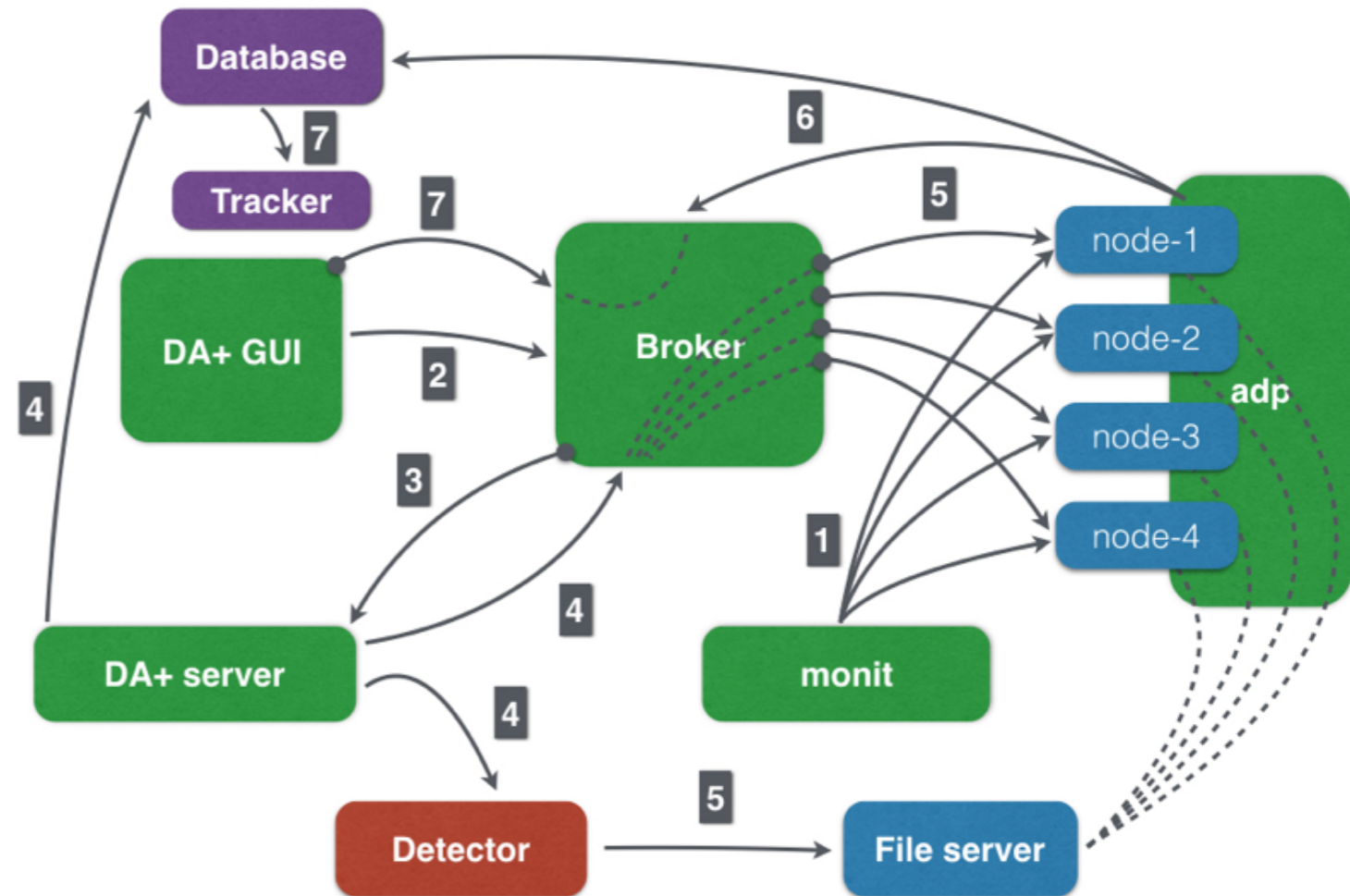
- Grid scan
 - streamed images are directly analysed in the memory (separate from adp)
- Screening - few diffraction images to evaluate quality of crystals
 - convert h5 data to cbf format (modification of eiger2cbf script) -> indexing & strategy with mosflm

- Dataset

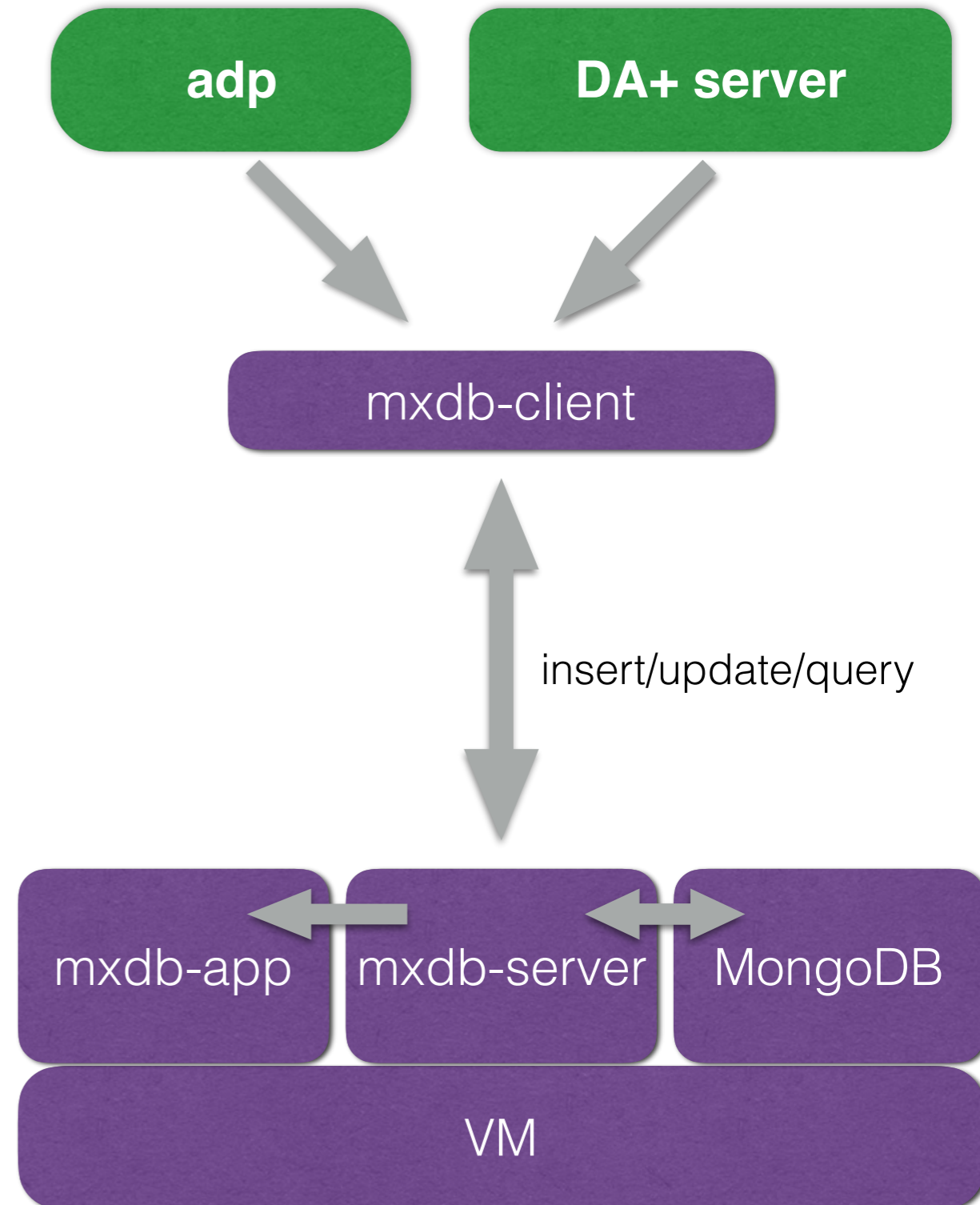
- fast_xds - split into 3 steps
- manual_xds
- go.com/goeiger.com

totalRange [°]	< 60	> 60 & < 180	180	> 180
fast_xds_1 XYCORR INIT	30	30	30	30
fast_xds_2 COLSPOT IDXREF	totalRange	60	60	60
fast_xds_3 DEFPIX INTEGRATE CORRECT	totalRange	totalRange	120	180

- Python-based
- monit ensures daemon processes wait for the message (json format) from DA+ server to mx-broker queue /adp_in/ {beamline}
- JobManager evaluates messages content & decides what to do next (using freely available crystallography data processing software)
- Results are sent to mxdb REST client (insert and update query)



- dedicated virtual machine, which serves as Docker daemon with containers
- mxdb REST server
- MongoDB
- mxdb-app (Google Polymer web components)
- adp (processing results) & DA+ server (experiment metadata) communicate with mxdb via mxdb-client
- mxdb REST server & MongoDB database interact using PyMongo



Database & tracker

Tracker

My App

mx-db.psi.ch/x06da/1c4acf2b-1607-4373-8eab-d9018a5f0220

SEARCH

SCREENING: crystal 10:37

Your strategy is being calculated...

PAUL SCHERRER INSTITUT
PSI

e14965@x06da

Database & tracker

My App mx-db.psi.ch/x06da/1c4acf2b-1607-4373-8eab-d9018a5f0220 Search

SCREENING: lysozyme 10:58 SCREENING: crystal 10:38 SCREENING: crystal 10:37

Space Group	P4	a	78.565	alpha	90	Strategy: complete native dataset						
High Resolution	1.068	b	78.565	beta	90	Omega	Phi	Total Range	Chi	Multiplicity	Completeness	Anomalous Completeness
Mosaicity	0.4	c	36.781	gamma	90	-85	0	90	0	3	99	83.9

PSI 20161012/crystal1/adp_lysozyme_1_20161012_105556 e14965@x06da

lysozyme 10:55 👍👎👎👍

crystal 10:39 👍👍👍👍

Experiment parameters

Oscillation Angle	0.1
Exposure Time	0.1
Beam Transmission	0.6
Number of Frames	900
Detector Distance	120
Wavelength	1

Crystal parameters: Go.com

a	78.78	alpha	90
b	78.78	beta	90
c	36.96	gamma	90
Space Group Label	P422	Space Group Number	89
Mosaicity	0.125	ISA	23.54
Wilson B Factor	11.816		

Twinning

Label	Value	Untwinned	Perfect Twin
$\langle I^2 \rangle / \langle I \rangle^2$	1.94	2	1.5
$\langle F \rangle^2 / \langle F^2 \rangle$	0.798	0.785	0.885
$\langle E ^2 - 1 \rangle$	0.721	0.736	0.541

Pointless: Space group is not uniquely determined

Order	Space Group	Space Group Number	Laue Group Confidence	Space Group Confidence
1	P422	89	1	
2	P41212	92		0.533
3	P43212	96		

Go.com (angularRange: 90.0)

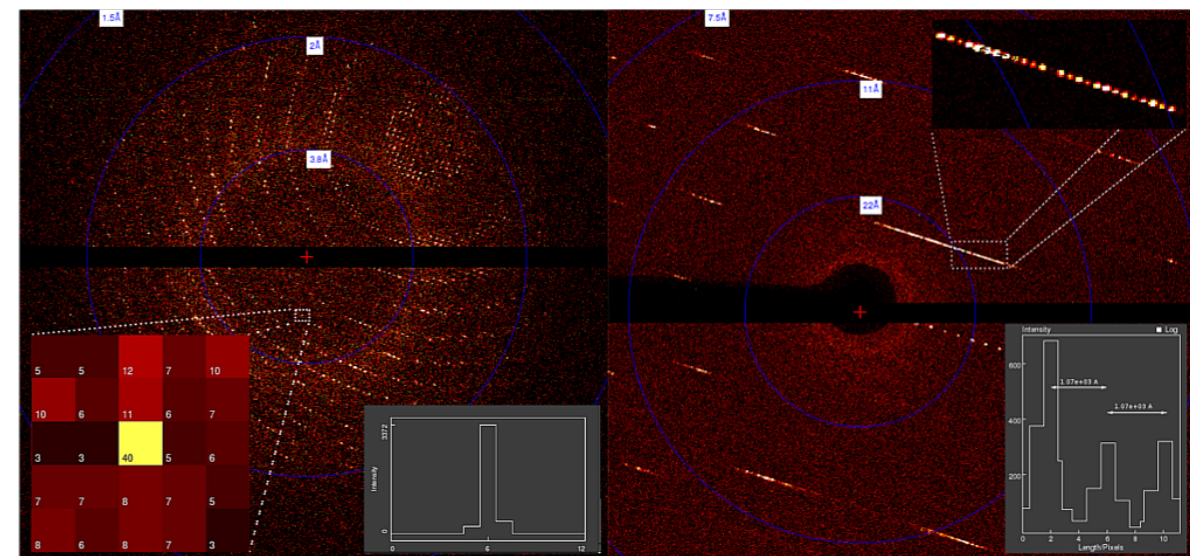
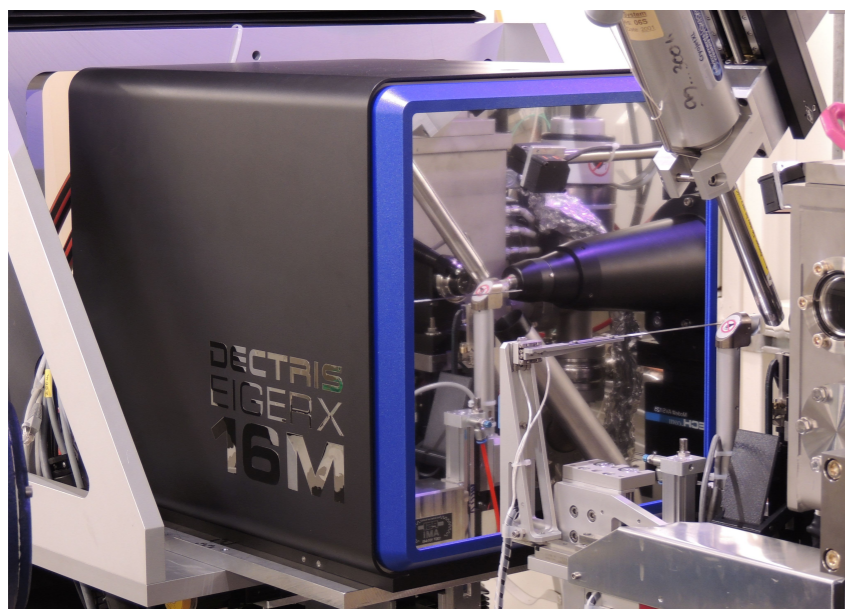
#	Resolution Limit	Observed Reflections	Unique Reflections	Completeness	R-factor observed	R-factor expected	lsigma	Rmeas	cc_half	sigAno
1	3.1	13881	3985	98	2.7	2.9	42.05	3.2	99.8	1.018
2	2.19	24522	7195	97.6	2.6	2.9	39.55	3.1	99.8	1.054
3	1.79	30943	9266	97.2	3	3	35.5	3.5	99.8	1.067
4	1.55	36037	10954	97.5	3.4	3.4	29.55	4.1	99.7	0.984
5	1.39	40894	12554	98.1	4.1	4.1	23.46	4.9	99.7	0.836
6	1.27	40896	13836	98.2	4.9	5.4	17.05	5.9	99.5	0.688
7	1.17	41477	14992	97.8	5.6	6.8	13.14	6.9	99.3	0.619
8	1.1	22508	13160	79.7	6.2	8.3	7.87	8.1	99	0.557
9	1.03	6541	5330	30.4	8.2	11.5	4.53	11.3	98	0.49
10	total	257699	91272	84.1	3	3.2	21.47	3.5	99.9	0.831

Eiger 16M

Detector

- Dectris prototype Eiger 16M integrated at the beamline X06SA (PXI)
- Continuous readout & internal summation
- No readout noise or dark current
- 1 pixel point-spread function
- Small pixel size (better resolution)

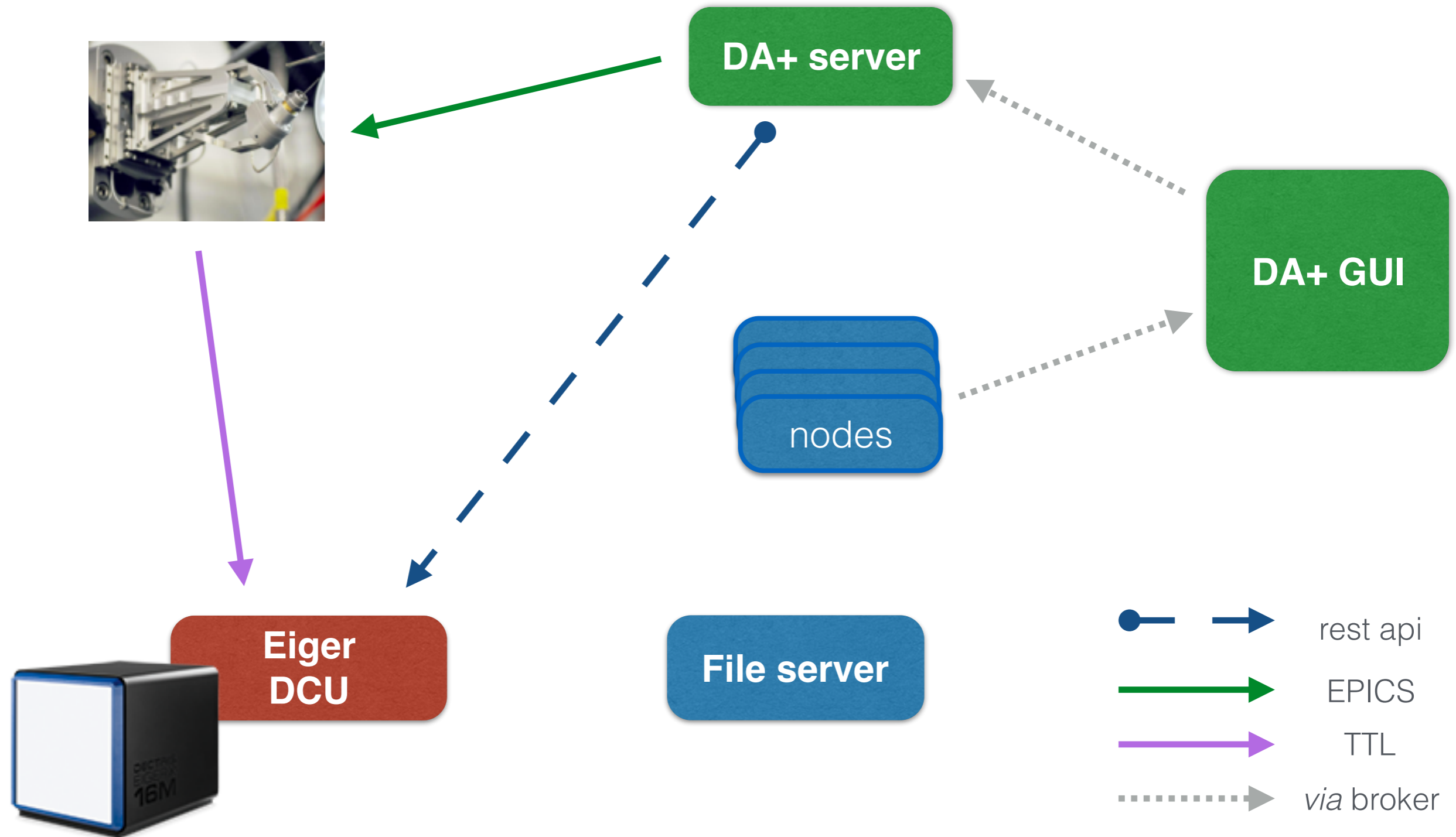
	PILATUS3 (PILATUS2)	EIGER
Pixel size [μm^2]	172 x 172	75 x 75
Pixel density [mm^{-1}]	34	178
Readout time [μs]	950 (2300)	3.8
Maximum frame rate [Hz]	100 - 500 (25 - 60) (6M - 1M)	133 - 3000 (16M - 1M)
Duty cycle [%]	90 - 52	> 99
Counter depth [bits]	20	12
Sensor thickness [μm]	450, 1000	450
Threshold energy [keV]	2.7 - 18 (3.9 - 18)	2.7 - 18



Casanas et al, Acta Crystallogr D Struc Biol, 2016

Eiger 16M

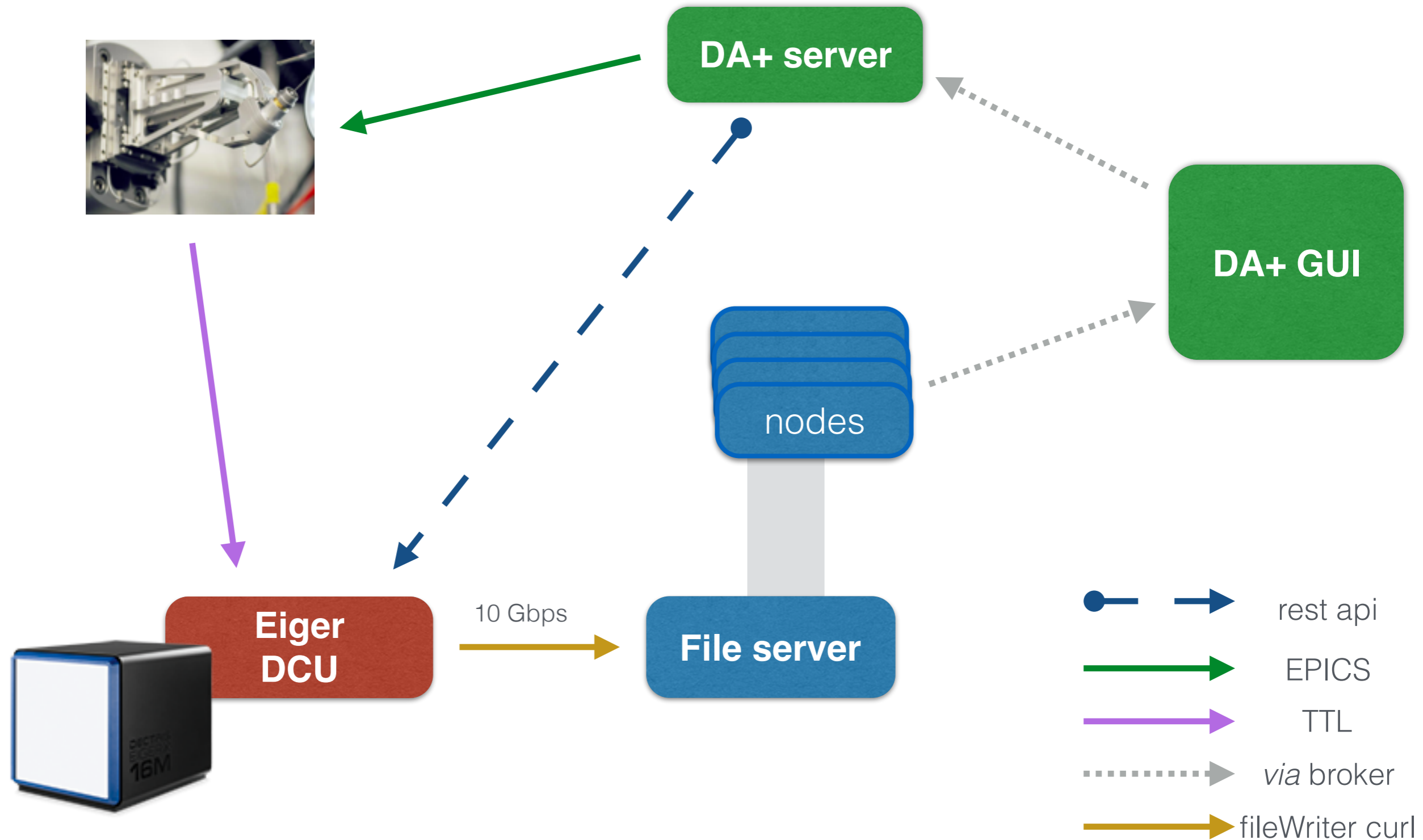
Detector



Eiger 16M fileWriter

Detector

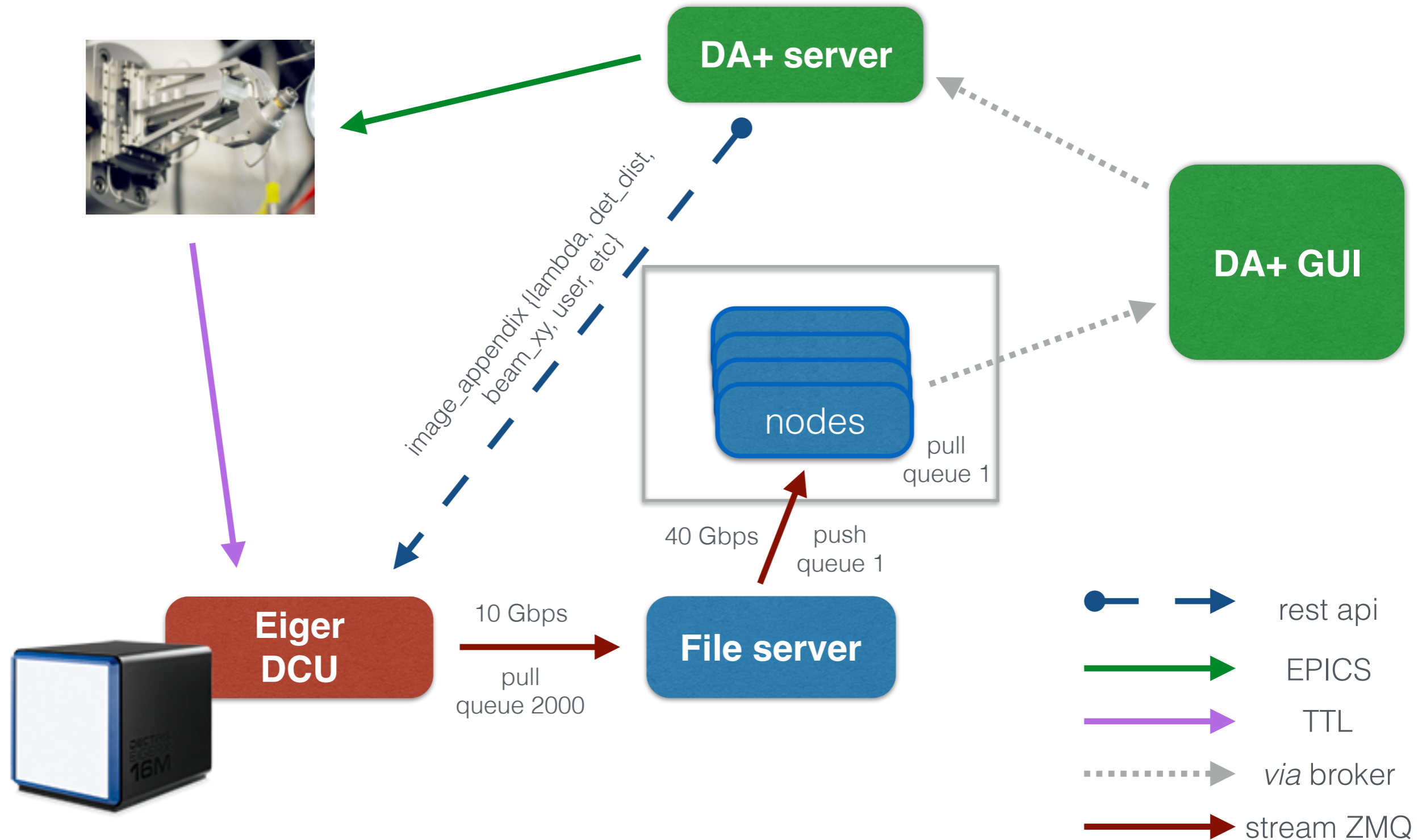
user inspection



Eiger 16M stream

Detector

online analysis










Summary

- developed in-house DAQ software, which supports conventional and advanced data acquisition protocols
- on-the-fly automatic data processing
- MongoDB and Google Polymer based results tracker
- EIGER 16M implementation (grid scan)

Acknowledgements

- Ezequiel Panepucci
- Jakub Kaminski
- Simon Ebner
- Meitian Wang
- MX group

Local contact persons

 5693 Shibom Basu Post Doc MX & serial crystallography, native SAD	 3584 Florian Dworkowski Beamline Scientist X105A Beamline hardware, spectroscopy	 5652 Aaron Finke Post Doc Experimental phasing, small molecules	 3959 Jakub Kaminski Software Engineer Software	 5437 May Marsh Crystallisation Facility Manager MX crystallography, Crystallisation	 3167 Isabelle Martiel Sample Preparation Scientist Sample delivery	 5233 Vincent Olieric Beamline Scientist X06DA MX crystallography, experimental phasing, in situ diffraction, sample changer
---	--	---	--	---	--	---

 5267 Ezequiel Panepucci Software Engineer MX crystallography, DA+ software	 5268 Anuschka Pauluhn Beamline Scientist X105A Beamline hardware	 5129 Takashi Tomizaki Beamline Scientist X06SA MX crystallography	 4175 Meitian Wang MX Group Leader	 5823 Rangana Warshamanage Post Doc Small molecules, data processing	 5455 Tobias Weinert Staff Scientist Serial crystallography, native SAD	 5428 Justyna Wojdyla Beamline Scientist X06DA MX crystallography, data processing
--	--	---	--	---	--	---

Other Staff

 3368 Dominik Buntschu Beamline Technician Beamline hardware, sample changer	 5518 José Gbadinho Software Engineer Controls X06SA + X105A	 5440 James Leuenberger Beamline Engineer Beamline hardware	 3263 Nathalie Meier Beamline Technician Beamline hardware	 3759 Sonia Reber Logistic Assistant Logistics, dryshipper support	 5136 Roman Schneider Beamline Technician Beamline hardware, cryo/vacuum	 5584 Vincent Thominet Optics Engineer Optics
 5827 Laura Vera Lab Technician Crystallisation, in situ diffraction	 3380 Xiaoqiang Wang Software Engineer Controls X06DA					

<https://www.psi.ch/macromolecular-crystallography/>

Thank you

Grid scan

- Position Synchronised Output (PSO) circuit within motion controller issues transistor-transistor logic (TTL) signal within preprogrammed regions (start of row/column) of the GMX or GMY axis
- EIGER 16M can collect n series each consisting of m images with each series triggered by one external TTL
- PILATUS 6M-F is configured to collect $m \times n$ images, each requiring an external trigger

