Data Acquisition, Processing and Monitoring at the NSLS-II Life Science Beamlines - an Adaptable Approach for Long-Term Development

Jun Aishima
Data Science and Systems Integration Program
National Synchrotron Light Source II
Outline

● What opportunities do we have?
● How are we seizing the opportunities?
● How are we developing the systems to meet the challenges of the future?
Outline

• What opportunities do we have?

• How are we seizing the opportunities?

• How are we developing the systems to meet the challenges of the future?

• New facility with small, intense beams
• Many beamlines already built with many more to come
• Biology beamlines
  – High degree of automation
  – Heterogenous beamline hardware and software ecosystems
  – Large user community
• LSDC

National Synchrotron Light Source II
● 29 beamlines in operation, 4 under development

● Storage ring: 792m circumference, 3 GeV, 400 mA operating current designed to deliver an electron beam with very small emittance (0.5 nm-rad H, 8 pm-rad V)

● One of the newest synchrotrons in the world, started operating in 2015 and has a large user community around the US and the world
<table>
<thead>
<tr>
<th>BL</th>
<th>Min. beam size</th>
<th>Flux</th>
<th>Detector</th>
<th>Max fr. rate</th>
<th>Robotic sample changer</th>
<th># samples</th>
<th>Data Acquisition GUI</th>
<th>LIMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMX</td>
<td>7x5 µm²</td>
<td>4e12</td>
<td>Eiger 9M</td>
<td>200+</td>
<td>Y</td>
<td>384</td>
<td>LSDC</td>
<td>SynchWeb /ISPyB</td>
</tr>
<tr>
<td>FMX</td>
<td>1x1.5 µm²</td>
<td>3.5e12</td>
<td>Eiger 16M</td>
<td>100+</td>
<td>Y</td>
<td>384</td>
<td>LSDC</td>
<td>SynchWeb /ISPyB</td>
</tr>
<tr>
<td>NYX</td>
<td>10x10 µm²</td>
<td>1e12</td>
<td>ADSC HF-4M</td>
<td>25</td>
<td>Y</td>
<td>240</td>
<td>Blulce</td>
<td>None</td>
</tr>
<tr>
<td>LiX</td>
<td>0.4 mm² (scattering)</td>
<td>3x</td>
<td>Pilatus</td>
<td></td>
<td>Y</td>
<td>54</td>
<td>Custom</td>
<td>None</td>
</tr>
<tr>
<td>XFP</td>
<td>0.1 x 0.4 mm²</td>
<td>1.6e16</td>
<td>N/A</td>
<td></td>
<td>N</td>
<td>96</td>
<td>Custom</td>
<td>None</td>
</tr>
<tr>
<td>Cryo-EM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Vendor</td>
<td>None</td>
</tr>
</tbody>
</table>
This talk will concentrate on LSDC (Life Sciences Data Collection)
Outline

• What opportunities do we have?
  • New facility with small, intense beams
  • Many beamlines already built with many more to come
  • Biology beamlines
    - High degree of automation
    - Heterogenous beamline hardware and software ecosystems
    - Large user community
  • LSDC

• How are we seizing the opportunities?

• How are we developing the systems to meet the challenges of the future?
Outline

● What opportunities do we have?

● How are we seizing the opportunities?

● How are we developing the systems to meet the challenges of the future?

● Experiments have been implemented
● Sample exchange (as we have seen in Edwin Lazo’s talk) is reliable
● Remote experiments are running – including automated data collection
● Processing systems effectively providing feedback
● LIMS system integrated
Everything you need to collect remotely

- Layout inspired by MXCube
- Sample viewing/centering/rastering feedback area
- Parameters area
- Sample handling and data collection history area
LSDC and related components

GUI

- Viewing + feedback
- Data collection setup
- Samples + data collection queue

Server

- Soft IOC
- Queue
- Collection

Sample, Collection, Processing services

- ADEiger IOC
- Eiger Detector

- Det lib
- Robot lib

MongoDB

Robot IOC

- Processing cluster
- EMBL Library

ISPyB/MariaDB

Staubli Sample Robot

Albula diffraction viewer

SynchWeb monitoring webapp
LSDC: current state at a glance

- Currently used at AMX and FMX
- Python (PyQt) GUI client
  - Configure all typical experiments (standard oscillation, diffraction and fluorescence-based rastering, vector/helical, energy scan)
  - GUI-driven sample exchange
  - Rastering heatmaps
  - Albula (Dectris) for diffraction image visualization
- Python server
  - Perform the queued experiments configured by the user
  - Using the libraries below
  - Communication with databases
  - Hardware configuration management
  - Initiating processing
    - FastDP, Xia2
    - Dimple
    - Chooch, Raddose, Dozor, DIALS Spotfinder, etc...
  - Libraries modularizing PyEpics calls to interact with PVs of the EPICS Control System
    - Detector, robot, general PVs
- On both client and server, many custom components including sample centering/rastering definition/raster feedback view, queuing system
LSDC GUI and Albula (Dectris) on NX
LSDC GUI + Albula, Synchweb
What challenges do data acquisition systems face?

- Evolving infrastructure
- Faster hardware
- Software improvements
- New techniques
- New beamlines
- Knowledge transfer among developers
- Integration within own facility
- Integration with outside beamlines
LSDC Core Modules

- State machine for safe experiment state transitions – the Governor
- X-ray fluorescence spectrum using standard libraries (Ophyd, Bluesky, Matplotlib, and PyMca)
- Ophyd/Bluesky for MX experiments (in progress)
- Sample, data collection, data processing information storage via web services
- ISPyB-API library (Diamond Light Source) for ISPyB interaction
- Interaction with Albula (Dectris) for diffraction visualization
- EPICS PV interaction code modularized into libraries
The Governor manages Experiment States

- Define **experiment states** with default positions
- Define **transitions** between states
- Define multiple **positions** per positioner
  - Motors (Near, Far, Park, ...)
  - Valves (Open, Close)
- Define allowed **ranges** within states
- Test transitions thoroughly once
- From now on, just call transitions to cycle through experiment states
Working with Bluesky

- Developed at NSLS-II and used at most beamlines
- Used at 5 US DOE Light Sources and the Australian Synchrotron
- Modularized data acquisition system
  - Control system abstraction layer
  - Procedure running code
  - Access layer to data
- http://blueskyproject.io
Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

- Experiments have been implemented
- Sample exchange (as we have seen in Edwin Lazo's talk) is reliable
- Remote experiments are running – including automated data collection
- Processing systems effectively providing feedback
- LIMS system integrated
- Modularization well under way
Outline

• What opportunities do we have?

• How are we seizing the opportunities?

• How are we developing the systems to meet the challenges of the future?

• Hardware abstraction to enable new beamlines (NYX in progress, more in the future?)
• New experiments
• Minimize disruptions due to infrastructure changes
• Rapid development of processing systems
• Continued modularization opportunities to incorporate/supply to other beamlines and facilities
• Improved infrastructure for better user and staff experience
Proposed future LSDC system - changes

Microscope Module
Parameter entry/API connection
Sample Handling Module

GUI

Authorization/Proposal

Bluesky-Queueserver

Queueserver API

Protocols:
1 – standard
2 – raster
3 – vector
4 – energy scan
……

Channel

Run Engine

Message broker

Ophyd

Collection done

IOC 1

IOC 2

Processing pipeline
Rastering pipeline
Other pipelines

Ophyd

Ophyd

Ophyd

Ophyd

Collection done

Run Engine

Message broker

Ophyd

Ophyd

Ophyd

Ophyd

Collection done
How LSDC is being prepared for long-term development

• Improving the core scanning code with Ophyd/Bluesky
  • Hardware abstraction layer
  • Offline testing
  • Better modularization
• Replace custom collection queueing code with Bluesky-Queueserver
• Modularization of GUI code
  • Specify a new "microscope" module that can be used at many beamlines to view samples and control experiments
• Software engineering improvements
• Improved processing systems triggered via Kafka message broker
  • More flexible, less hard-coded processing – adding multi-crystal, multi-dataset (from Dale’s talk)
  • Reprocessing
LSDC is on Github!
https://github.com/NSLS-II/lsdc

- Future
  - Continuous Integration
  - Unit testing
  - pip installation
  - Docker

LSDC - Life Sciences Data Collection

Author: John Skinner, formerly of National Synchrotron Light Source II, Brookhaven National Laboratory, NY

Currently used at the NSLS-II MX beamlines - AMX, FMX

Offline fork of https://github.com/jskinner53/lsdc

Needs more documentation as only a module diagram currently exists.
Common systems being developed

• NSLS-II as a whole
  – Guacamole to complement NX
  – Single sign-on
  – Direct access to web applications once signed in
  – Better integration of proposal system

• LSDC modules
  – Bluesky queueing system
  – Bluesky GUI experiment configuration module
  – Microscope module
  – Sample management module
  – Externalize processing with Kafka message broker system (Dale's talk from yesterday)
PyMDA: microcrystal data assembly using Python – Multi-crystal crystallography and Sulfur-SAD at 5 keV

- G. Guo, P. Zhu, Q. Liu (BNL Biology Dept)

- Data collection at 5 keV (λ = 2.48 Å) from 5-10 μm thaumatin crystals
- Strategy of data analysis:
  - Initial data assembly based on CC1/2
  - Crystal and frame rejections based on SSmRmerge

Ultrafast raster-scanning serial crystallography data processing:
- ~200 partial datasets for structure solution
- Equally high data quality for detector frame rates of 200, 500 and 750 Hz

Guo et al., IUCrJ 6, 532 (2019)
Guo et al., IUCrJ 5, 238 (2018)

Protein K structure refined to 1.5 Å resolution: Rcryst = 16.7% and Rfree = 21.4% (500Hz dataset)

Gao, Y. et al., J. Synchrotron Rad. (2018), 25, 1362-1370
Outline

- What opportunities do we have?
- How are we seizing the opportunities?
- How are we developing the systems to meet the challenges of the future?

- Hardware abstraction to enable new beamlines (NYX in progress, more in the future?)
- New experiments
- Minimize disruptions due to infrastructure changes
- Rapid development of processing systems
- Continued modularization opportunities to incorporate/supply to other beamlines and facilities
- Improved infrastructure for better user and staff experience
What challenges do data acquisition systems face?

- Evolving infrastructure
- Faster hardware
- Software improvements
- New techniques
- New beamlines
- Knowledge transfer among developers
- Integration within own facility
- Integration with outside beamlines
The future is bright for the NSLS-II beamlines!

- More common infrastructure systems at NSLS-II
  - Good for users
  - Good for staff!
- Development of new cross-beamline software projects
- Bluesky as both a cross-beamline and cross-facility project
- LSDC as a framework for GUI-based data acquisition
- All while keeping the existing users happy!
Thanks!

- John Skinner and Bob Sweet for their long-term work on LSDC and its predecessors
- Beamline staff – Stu Myers, Martin Fuchs, Jean Jakoncic, Dale Kreitler, Edwin Lazo in particular
- Data Science and Systems Integration group
  - Tom Caswell, Dan Allan – Bluesky originators
  - Marcus Hanwell – microscope module
  - Maksim Rakitin – CI, packaging, and Ophyd/Bluesky integration
- SynchWeb, ispyb-api developers (Diamond Light Source)
- Albula library (Dectris)
Current MX beamline situation – what the user sees

User

VPN
BNL password

Support Staff

Google
Hangouts

NX client
Controls password

LSDC
GUI

Albula

SynchWeb